International Journal of Computer Vision, 29, 233–253 (1998)

Generalization to Novel Views: Universal, Class-based, and Model-based Processing

YAEL MOSES AND SHIMON ULLMAN
Department of Applied Mathematics and Computer Science, The Weizmann Institute of Science, Rehovot 76100, Israel

Received November 13, 1996; Accepted January 26, 1998

Abstract. A major problem in object recognition is that a novel image of a given object can be different from all previously seen images. Images can vary considerably due to changes in viewing conditions such as viewing position and illumination. In this paper we distinguish between three types of recognition schemes by the level at which generalization to novel images takes place: universal, class, and model-based. The first is applicable equally to all objects, the second to a class of objects, and the third uses known properties of individual objects. We derive theoretical limitations on each of the three generalization levels. For the universal level, previous results have shown that no invariance can be obtained. Here we show that this limitation holds even when the assumptions made on the objects and the recognition functions are relaxed. We also extend the results to changes of illumination direction. For the class level, previous studies presented specific examples of classes of objects for which functions invariant to viewpoint exist. Here, we distinguish between classes that admit such invariance and classes that do not. We demonstrate that there is a tradeoff between the set of objects that can be discriminated by a given recognition function and the set of images from which the recognition function can recognize these objects. Furthermore, we demonstrate that although functions that are invariant to illumination direction do not exist at the universal level, when the objects are restricted to belong to a given class, an invariant function to illumination direction can be defined. A general conclusion of this study is that class-based processing, that has not been used extensively in the past, is often advantageous for dealing with variations due to viewpoint and illuminant changes.

Keywords: object recognition, invariance

1. Introduction

One of the main problems in recognizing 3D objects is that a 2D image of an object depends not only on its shape but also on the conditions under which the image was taken, e.g., viewpoint and illumination condition. Images of the same object may therefore vary considerably. In daily life we recognize objects in novel images despite the variations between images of the same object. Existing computer systems, on the other hand, are still limited in their ability to perform such an unconstrained object recognition task. Recognition schemes developed in the past addressed the problem of recognizing objects in novel views by suggesting specific techniques for overcoming variations between images of the same object due to changes in viewing conditions. In this paper we study general properties that allow such generalization to take place rather than focus on a specific recognition technique. The main question we address is what are the underlying processes that allow natural or artificial systems to generalize the recognition of an object from familiar views to novel images.

1.1. Levels of generalization

We propose a classification of different generalization processes in recognition based on the specificity of the information used to compensate for variations between images of the same object. We distinguish between three different levels of specificity: *universal*; *class-based*; and *model-based*. We define the three levels and then study the limitations of each of them in overcoming image variations caused by changes of viewpoint and illumination direction. Roughly speaking, the universal level is common to all images independent of the specific set of objects to be recognized. For example, the use of edge extraction to deal with illumination changes is a universal process, applicable to all the incoming images (Canny, 1986; Davis, 1975; Haralick, 1984; Marr and Hildreth, 1980; Torre and Poggio, 1986). At the other extreme lies the model-based level. At this level, the processing applied to compensate for image variations depends on the specific object to be recognized. An example of model-based processing is recognition by 3D alignment (Fischler and Bolles, 1981; Jacobs, 1992; Huttenlocher and Ullman, 1990; Lowe, 1987; Ullman, 1989; Ullman and Basri, 1991; Weinshall, 1993) as we describe in greater detail below. An intermediate level of generalization is the class-based level. At this level, the generalization process uses properties associated with certain classes of objects, for example, the class of faces (e.g. Kanade, 1977), the class of bilaterally symmetric objects (e.g., Fawcett *et al.* , 1994; Moses and Ullman, 1992), or the class of planar objects (e.g., Lamdan *et al.* 1987; Rothwell *et al.* , 1992). Such processes are applicable to any image of an object that belongs to the class in question without a precise model of the individual object.

These different levels of generalization are explained in Section 3. In general, the level at which the system compensates for variations between images of the same object places basic constraints and requirements on the computational aspects of the recognition process. Understanding the theoretical limitations of each of the three levels is important for the development of particular approaches to object recognition.

1.2. Recognition functions

In the analysis that follows we will be interested in three types of recognition functions: consistent, imperfect, and optimal. We next briefly describe these types of recognition functions, a formal definitions are given in Section 2. A recognition function can be regarded as a function from images of objects from a given set, s, to some representation space, $\mathcal{N}$. The representation can be, for instance, an object name, or a canonical view. Clearly, we would like a recognition function to map different images of the same object to the same representation (e.g., the same name) independent of the conditions under which the images were taken (e.g., viewpoint). We call a recognition function *consistent* if it can recognize an object from all of its images. However, consistency is clearly a very strong requirement. For example, if two objects have a single image in common, then they will be entirely unseparated by a consistent recognition function. It is therefore natural to also examine *imperfect* recognition functions that are allowed to misidentify each object from a subset of its images. Of the imperfect recognition functions, we will still be interested in functions that misidentify objects from as few images as possible. We therefore also examine *optimal* recognition functions, that recognize an object from as many images as possible.

1.3. The goal of this paper

It was previously shown (Burns, Weiss and Riseman, 1992; Clemens and Jacobs, 1991; Moses and Ullman, 1992) that the recognition of an object from novel viewpoints *cannot* be performed at the universal level. To make recognition possible we can restrict the requirement of a recognition function in one of two ways: one is to require less than full consistency, the other is to restrict the set of all possible objects and consider class and model-based recognition functions.

In this paper we will investigate the limitations of universal, class-based, and model-based recognition functions. For each of the three generalization levels we will consider three types of recognition functions: consistent, imperfect, and optimal. Furthermore, previous studies that relate to this question focused on image variations due to viewpoint changes. In this paper we extend the viewpoint results to variations due to illumination direction. The main goal of this paper can therefore be summarized in terms of filling all the

Table 1. The goal: to fill in this table.

	Universal	class	model
Consistent			
Optimal			
Imperfect			

entries in the the 3×3 table shown in Table 1. We will refer again to this table in the final discussion.

1.4. Previous results and summary of our results

Previous studies (Burns, Weiss and Riseman, 1992; Clemens and Jacobs, 1991; Moses and Ullman, 1992) established that a universal consistent recognition function must be the constant function (cannot discriminate between any two objects). This result fills in the first entry of the table (universal/consistent) with respect to viewing position. Here we extend this result to objects that consist of 3D contours rather than 3D point set. We extend this result further by also considering the effects of illumination changes. Existing recognition systems often attempt to solve the illumination problem at the universal level, by extracting contour maps (e.g., edges), or special points (e.g., corners) that are illuminant insensitive. It is well known that image representations such as edge map have limitations and can fail on complex images (for example, on face images (Adini, Moses and Ullman, 1997)). Our study shows in fact some of the limitations of an edge-based representation which is quite widely used in practice. We show that for grey-level images, a universal recognition function that is consistent with respect to both viewpoint and illumination direction must still be the constant function. It follows that such a recognition function will fail to discriminate between any two objects.

For universal imperfect recognition functions (i.e., functions that are allowed to missrecognize a subset of the images), it was shown (Burns, Weiss and Riseman, 1992) that a recognition function must still be constant if it is defined on all objects except for a measure zero set of objects, and all their images except for a measure zero set of images for each object. Here we extend this result and show that even if the universal recognition function is allowed to fail to recognize each object from almost half of the set of its images, it must still be the constant function. Finally, we also show that an optimal recognition function does not exist at the universal level.

Regarding the class-level of generalization (second column of Table 1), a number of specific class-based schemes have been proposed for dealing with viewpoint variations by using invariant representations. (A review of invariance for specific classes of objects is given in Zisserman *et al.* 1995.) In this paper we demonstrate that the existence of a recognition scheme that can compensate for changes in viewpoint depends on the class in question. Furthermore, it also depends on the set of object-images for which the scheme is required to recognize the objects correctly. We demonstrate how the set of images for which the recognition function is allowed to misidentify the objects affects the set of objects that can be discriminated by this function. Existing class-based schemes are restricted to image variations resulting from viewpoint changes only. We show that class-based processing can also compensate for variations due to changes of illumination condition, although universal processing is insufficient under similar conditions. We conclude that the class-level of processing is often a useful approach for dealing with the effect of viewing direction and illumination condition.

Finally, regarding the model-based level (third column of Table 1), we show that at this level it is always theoretically possible to overcome image variations due to changes in both illumination and viewpoint conditions by using imperfect or optimal recognition functions.

The rest of the paper is organized as follows. In Section 2 definitions of the functions used in this paper are given. The three generalization levels are defined in Section 3. The theoretical study of the three generalization levels is based on the notion of *reachability partition* which is described in Section 4. The theoretical limitations of the universal, class-based, and model-based levels are presented in Sections 5–7. Finally, Section 8 presents summary and discussion of these results.

2. Recognition functions and consistency

For defining the three levels of generalization, it is convenient to first define the notion of a recognition function. Let s be a finite set of objects taken from a given universe $\mathcal{U}$. Let $\mathcal{I}$ be the set of images of the objects in $\mathcal{U}$. $\mathcal{I}$ depends on the set of objects in the universe $\mathcal{U}$, the projection (e.g., perspective or weak perspective) as well as other imaging parameters such as viewpoint or illumination directions. A recognition function is a function from images of objects from the set $\mathcal{U}$ to some representation space, $\mathcal{N}$ (e.g., the set of names or canonical views of the objects). For example, in a face recognition system the universe $\mathcal{U}$ is the set of human faces. The recognition function f will be defined for all face images: it may compute, for example, a set of relative distances between facial features as in (Kanade, 1977). In practice, the system will then be applied to a finite set of faces.

Clearly, we would like a recognition function to be consistent, i.e., to have the same value on different images of the same object. Formally, we define a recognition function f to be *consistent* if its value in $\mathcal{N}$ is identical for all images of the same object from the set s. That is, if $I_1 = I(o, v_1)$ and $I_2 = I(o, v_2)$ are two images of the same object o from the set s taken from views v_1 and v_2 then $f(I_1) = f(I_2)$. Note that a constant function is in particular also a consistent recognition function, however, it cannot discriminate between any two objects. We are therefore interested also in recognition functions that can discriminate between objects in the set. A recognition function f is called *discriminative*, if its value in $\mathcal{N}$ is different for images of two distinct objects from the set s. That is, if $I_1 = I(o_1, v_1)$ and $I_2 = I(o_2, v_2)$ are two images of objects o_1 and o_2 from the set s, such that $o_1 \neq o_2$, then $f(I_1) \neq f(I_2)$.

Note that when two objects in the set s have a common image, a recognition function cannot be consistent and discriminative at the same time. We therefore combine the properties of consistent and discriminative recognition functions and define an *optimal* recognition function. A recognition function is said to be optimal if it is discriminative, and at the same time it is consistent on images of the same object that are not common to other objects in the set. An optimal recognition function can therefore discriminate between as many objects as possible from a given set while still recognizing each of the objects in the set from as many views as possible. Following is a formal definition of an optimal recognition function.

Definition 1. A recognition function for a finite set of objects s is *optimal* if the following conditions hold:

1. If I_1 and I_2 are two images of the same object $o \in s$, and I_1 and I_2 are not images of any other object in s, then $f(I_1) = f(I_2)$. (This is the consistency property.)
2. If I_1 is an image only of object o_1 and I_2 is an image only of object o_2, then $f(I_1) \neq f(I_2)$. (This is the discriminative property.)
3. If I is a common image of the objects $o_i, o_j \in s$ then $f(I)$ is arbitrary. (In practice, it can be consistent with either o_i or o_j.)

This is a natural definition, it simply means that f performs correct recognition on all the unambiguous images. It follows directly from this definition that for any finite set of objects an optimal recognition function always exits. In Section 7 we will show that at the model-based level (where $s = \mathcal{U}$) any optimal recognition function will fail to recognize each object only from a finite set of its images. The question of the existence of class-based or universal optimal recognition functions is addressed in Section 5 and 6.

In addition to the inherent limitations placed by common images, a recognition system may in practice make errors or misidentify an object from additional images. For example, the human visual system sometimes fails to recognize an unfamiliar view of an object such as a bottle from a top view, that may in principle be recognizable (Biederman, 1985; Warrington and Taylor, 1978). It is therefore natural to examine recognition functions that are not entirely consistent or optimal (see also Section 5.1). We thus consider also *imperfect* recognition functions: functions that fail to be consistent on a subset of images of each object. An imperfect recognition function is consistent only on a subset of images, the *recognizable* images. The recognition function can have arbitrary values on the other images, the *confusable* images. In this case, if I_1 and I_2 are two recognizable images of the same object from the set s, then $f(I_1) = f(I_2)$. However, if one of the images is a confusable image then $f(I_1) \neq f(I_2)$ may hold. Since in certain cases a consistent recognition function does not exist (except the constant function) it is of interest to inquire whether it is possible to recognize at least a subset of the images by an imperfect recognition function. Clearly, in order for an imperfect recognition function to be interesting, the set of recognizable images must be sufficiently large. This question is addressed in Section 5.1 below.

In the analysis we focus on the issue of existence rather than construction. That is, we study the limitations imposed on any recognition function by the ambiguity of common views. The question of constructing an efficient recognition scheme in different domains is of course a major issue and the subject of many studies in recognition.

3. Levels of generalizations

Different recognition schemes attempt to deal with different universes of objects. Some recognition methods attempt to be general and not specific to particular classes of objects. Other schemes attempt to develop methods tailored to a specific class of objects, such as human faces. Finally, some methods are developed to deal with a known pre-determined set of objects, such as a set of machine parts in a specific practical application. Accordingly, we distinguish between the universal, class, and model-based levels.

Model-based level: The universe $\mathcal{U}$ of model-based recognition functions consists of a specific finite set of objects s, that is $\mathcal{U} = s$. In particular, a model-based recognition function can be tailored for the specific set of objects (s). The recognition function f_s in this case, may change when a new object is added to the set s (learning a new object by the system).

The alignment approach is an example of model-based recognition approach. In this case the viewpoint of the image with respect to the model is computed by a function that depends on the candidate model and the image. The model is then transformed to align it with the image. In the absence of sensor errors and occlusions, the transformed model and the image will become identical only if the image contains an instance of the model. The recognition process consists of applying this transformation to all models in its database. This model-based processing allows generalization to new viewing positions, but it is restricted to the set of objects, s, already existing in the database. Another example of model-based processing is provided by some neural-network models. A network may be trained to recognize all the digits from 0 to 9. To recognize a new symbol, the system will have to be trained on the additional symbol object, and a new function will be coded by the net.

Universal level: At the other extreme lies the universal level: the universe $\mathcal{U}$ of universal recognition functions includes all possible 3D objects. The finite set of objects, s, that the system is required to recognize may therefore consist of any subset of 3D objects. The recognition function is defined independently of the set of objects that it will have to recognize. A universal recognition function can be regarded as a fixed bottom-up function that does not change when a new object is added to s. In particular, a universal recognition function that is consistent with respect to a given imaging parameter must be invariant to this parameter. For example, an invariant function to changes of viewpoint of all possible 3D objects, can be regarded as a universal recognition function that is consistent with respect to viewpoint changes.

Recognition schemes usually do not attempt to deal with all possible 3D objects. However, universal processing is still worth considering for two reasons. The first is to understand the limitations on the degree of generality that can be expected from a recognition system. The second motivation is that it is also possible to consider universal processing for dealing with a restricted set of viewing parameters, rather than the entire recognition process. In particular, universal methods have been proposed to deal with changes of illumination. An example of a universal operation widely used in computer vision is the extraction of contours from grey-level images. A major goal of this intermediate representation is to extract image features that are relatively illuminant-insensitive. In biological systems there is evidence suggesting a similar process that emphasizes intensity edges, that is applied in a uniform manner by the primary visual cortex to all incoming images (Hubel, 1962; Hubel and Wiesel, 1968). This stage of processing was modeled as the application of a set of local filters to the incoming image (Daugman, 1985; Marcelja, 1980; Marr and Hildreth, 1980; Pollen and Ronner, 1983). The question still remains whether universal processing of this type is sufficient to produce illumination insensitive representations. This question is taken up in Section 5.3.

Class-based level: An intermediate level between the universal and the model-based levels of generalization is the class-based level. The universe $\mathcal{U}$ of a class-based recognition function consists of all possible objects within a given class of objects. This class may be, for example, the class of faces, cars, symmetric objects, or planar objects. The finite set of objects, s, that the system is required to recognize can consist of any subset of the objects in $\mathcal{U}$. In this case the recognition function depends on the class to which the object is assumed to belong, and can use constraints imposed by the class to compensate for changes in viewing conditions. However, it is independent of the specific set of objects, s, that can be selected from the class. When a new object from the class is learned by the system, it will not affect the recognition function. An invariant function to viewing position of all objects that belongs to a given class (e.g., Zisserman *et al.* 1995), can be regarded as a class-based recognition function that is consistent with respect to changes of viewpoint.

We refer to a class as a large (possibly infinite) collection of objects (see also Section 6). The class-based recognition function, f_C, is fixed for the class C. Such functions may be constructed after learning several examples of objects from the class. However, they should then be applicable to any finite set of objects that belongs to the class C. For example, if the object in the image is assumed to be a face, the class-based recognition function can be based on extracting facial features such as the location of the eyes, mouth and nose (Brunelli and Poggio, 1991; Craw, Ellis and Lishman, 1987; Kanade, 1977; Kaya and Kobayashi, 1972; Nixon, 1985; Yuille, Cohen and Hallinan, 1992; Wong, Law and Tsang, 1989). Such a process can then be applied to recognize any finite set of faces, but is not applicable to other objects.

Note that different systems can compensate for image variations due to a given imaging parameter at different generalization levels. Consider for example the task of recognizing a specific triangle despite position, orientation, and scale changes in the image plane. A recognition system can apply a similarity transformations to the image to align it with a candidate model. Such a system generalizes to novel views at the model-based level and requires a different model for each triangle to be recognized. A different system can compute the list of the triangle's angles as a new representation of the image. In this case, the scheme applies to all possible triangles, and it overcomes the variability between images of the triangle at the class-based level of processing.

The three levels of generalization were defined for recognition functions. Within this framework one can also analyze methods for compensating for a particular viewing parameter as a part of a more general recognition process. Consider a given scheme that is supposed to filter out illumination effects without compensating for other imaging parameters, such as viewing position. In this case the set of images $\mathcal{I}$ depends on the set of objects in $\mathcal{U}$, the projection model, and the illumination. Variations in viewing position, for example, are not considered. The universe $\mathcal{U}$ can in particular consist of several poses of the same object. The consistency condition on f is required to hold only for images of the same object taken with different illumination condition. The output of this function can later be used as an intermediate representation for a more complete recognition function that compensates also for viewpoint changes. The complete recognition system may thus compensate for different viewing parameters at different levels. The alignment approach mentioned above compensates for different parameters at different levels. In this approach the first stage often consists of representing the grey-level image by its edge map. This stage is a universal process that results in an image representation that is often insensitive to illumination changes. The next stage can be either class-based or model-based. For example, for the class of planar objects, the effects of viewing position can be handle by a class-based recognition function that computes an invariant representation for each object (Rothwell et al, 1992; Lamdan and Wolfson, 1988). Alternatively, the second stage may deal with arbitrary 3D objects using a model-based approach, by using 3D models of the objects in the database. In this case each candidate model from s (the database) is projected to align best with the target image.

4. Reachability partition

To study the limitations of each of the three generalization levels we determine what sets of objects from a given universe cannot be discriminated by a recognition function.

We first consider consistent recognition functions. Since a consistent recognition function yields the same value for all the images of a given object, it will produce the same value for any two objects that share an image. This motivates the following definition of a *reachability* sequence.

Definition 2. A reachability sequence is a sequence of objects such that every two successive objects share an image.

Note that reachability depends on the choice of projection model (e.g., parallel, or perspective), since two different objects may share an image under one projection model but not another. Clearly, a consistent function must have the same value for all the images of the objects in a reachability sequence. The following proposition therefore follows directly.

Proposition 1. *Any consistent recognition function cannot discriminate between two reachable objects (objects that can be connected by a reachability sequence).*

Note that although two reachable objects do not necessarily share an image, the recognition function cannot discriminate between them. The existence of a consistent recognition function that can discriminate between objects is determined by the reachability partition of a given universe. It is therefore sufficient to study the reachability partition of the three generalization levels in order to fill in the first row in Table 1.

The reachability relation determines also the existence of optimal recognition functions. By definition, an optimal recognition function always exists for a finite set of objects, and therefore a model-based optimal recognition function always exists. For class-based recognition functions the reachability partition determines the set of objects for which such functions can be optimal. A recognition function can be optimal for every set of objects $s \subset \mathcal{U}$ only if every object in s belongs to a distinct reachability partition of $\mathcal{U}$. To see that, consider two objects that share an image. Given a recognition function, it is straightforward to choose a set of objects such that the recognition function is not optimal. This set should consists of either the two objects that share an image or only one of them, depending on the values of the recognition function on their images. In particular, if the function has the same value on all images of the two objects, than it is not optimal on s that consists of the two objects, otherwise it is not optimal on the set s that consists of only one of the object for which the function has different value on the common image. It also follows that no optimal universal recognition function exists since, as we will see below, the reachability partition of a universal function is trivial. It follows that to obtain optimal recognition, it is often necessary to tailor the recognition function to the set of objects under consideration. (Although, in practice, it is common to first define a recognition function in general, and then apply it without modification to different sets of objects.) It is therefore sufficient to study the reachability partition of the three generalization levels in order to fill in the second row in Table 1.

We next turn to consider imperfect recognition functions (the last row in Table 1). The notion of reachability can be extended in a natural manner to an imperfect recognition function, namely, a function that is consistent only on a subset of images for each object in the universe. Given an imperfect recognition function f, let $\mathcal{I}_{rec}$ be the set of images that f is consistent on. We define $\mathcal{I}_{rec}$-reachability sequence to be a sequence of objects such that every pair of successive objects share a recognizable image of both objects. As in the consistent case, the value of the imperfect recognition function must be identical for all the images of the objects in an $\mathcal{I}_{rec}$-reachability sequence. Therefore, two $\mathcal{I}_{rec}$-reachable objects cannot be discriminated by any imperfect recognition function that is consistent on the same images as the imperfect function f.

Reachability is an equivalence relation that does not depend on the specific recognition function used. An $\mathcal{I}_{rec}$-reachability is also an equivalence relation, and it depends only on the set of recognizable images of an imperfect recognition function, f. Therefore, any universe of objects can be divided by a reachability (or $\mathcal{I}_{rec}$-reachability) partition such that two objects are within the same reachability partition if and only if they are reachable (or $\mathcal{I}_{rec}$-reachable) from one another. The reachability partition defines the subsets of objects that can be discriminated by a consistent (or imperfect) recognition function for a given universe.

It is therefore useful to study the reachability and $\mathcal{I}_{rec}$-reachability partition of the universe in question. Note that when the set of excluded views is changed, then also the $\mathcal{I}_{rec}$-reachability partition is changed. As a result, the existence of imperfect recognition function strongly depends on the set of excluded images.

In the following sections we study the limitations of the three generalization levels by using the reachability partition of different universe. We assume in the rest of this paper a weak perspective projection model.

5. Universal recognition functions

The universe of a universal recognition function consists of all possible 3D objects. To show that a universal recognition function must be a constant function, two strong assumptions were made in previous studies (Burns, Weiss and Riseman, 1992; Clemens, 1990; Moses and Ullman, 1992). The objects were assumed to consist of 3D point sets and the recognition function was assumed to be entirely consistent on all but a measure zero of objects, and for all but a measure zero of the set of images of each object. Furthermore, only consistency with respect to viewpoint changes were previously considered in (Burns, Weiss and Riseman, 1992; Clemens, 1990). In this section we extend the previous results and show that a universal recognition function must still be the constant function when the conditions are relaxed in one of the following manners.

- The recognition function is imperfect. That is, for each object the recognition function is allowed to misidentify the object from almost half of its images.
- The objects consist of contours rather than 3D points.
- The objects consist of n Lambertian surface patches rather than n points. In this case a grey-level value (which depends on the surface normal, the illumination direction, and the point albedo) is associated with each point on the image. Furthermore, the recognition function is required to be consistent with respect to changes of both viewpoint and illumination direction.

To prove these results we show that the reachability (or $\mathcal{I}_{rec}$-reachability) partition in the universal case is the trivial partition (any two objects are reachable). It also follows that for each of the above cases an optimal recognition function cannot be universal.

5.1. *Imperfect recognition functions*

We first consider imperfect recognition functions, namely, functions that are not necessarily consistent on all possible images of the object. In this case, the recognition function is allowed to misidentify each object from a subset of its images. For example, the function may identify only one object from an image that is common to two objects, (that is, the image is recognizable for one object and confusable for the other), or it can misidentify both objects (that is, the image is confusable for both objects). This case is more realistic than the assumptions regarding an ideal errorless recognition function (Clemens and Jacobs, 1991), or an ideal recognition function that remains undefined for a measure zero of images for each object (Burns, Weiss and Riseman, 1992). The main conclusion from this section is that universal processing is more severely limited than previously analyzed: the limitations of a universal consistent recognition function hold even if the function is allowed to misidentify about half of the images. To establish this claim, it is sufficient to prove (see Section 4) that in the universal case, any two objects are $\mathcal{I}_{rec}$-reachable if the set of images for which f is consistent satisfies certain assumptions.

Let us define more precisely the claim and the conditions under which it is established. As in previous studies, we assume here that an object consists of a set of 3D points in space. An image of such an object under orthographic projection is uniquely determined by the object shape (the points 3D location) and the viewing parameters. The viewing parameters are the viewing direction $\vec{v}$, rotation and reflection in the image plane R (2×2 rotation and reflection matrix), translation vector $\vec{t} \in \mathcal{R}^2$, and a scaling factor $s \in \mathcal{R}$. We assume here that if f is consistent on a given image, then it will also be consistent on the same image scaled, rotated, reflected, and translated (except the trivial scaling in which an object vanishes to a point, which will make the proof trivial and therefore uninteresting). In other words, if an object is recognizable from a given image, then it is also recognizable from its transformed images in the image plane.

A recognition function may be consistent on a different set of viewing directions for each object in the universe. For example, consider two objects: a bottle and a plate. A recognition system may misidentify the bottle from its top view and correctly identify it from its side view, but may misidentify the plate from its side view and recognize it from its top view. In this case the recognition function is inconsistent on the set of views that are close to the top view of the bottle and similarly on a set of views that are close to the side view of the plate. Note that such a recognition function is consistent on the side view of the bottle and the top view of the plate. We assume here that the set of confusable views of all objects in a small neighborhood of an objects is bounded. This assumption follows from the following two natural assumptions on the confusable viewing directions (viewpoints corresponding to confusable images). The first is that the set of confusable viewing directions for a given object is bounded. This is a natural assumption, because we clearly would like a recognition function to recognize an object in a large number of its images. The second of our assumptions above is that two similar objects have similar sets of confusable viewing direction. That is, if we consider again the bottle and plate example, we assume that objects that are similar to the bottle will have similar (but not necessarily identical) sets of confusable viewing directions (close to the top view). These assumptions can be regarded as a smoothness assumption on the sets of confusable views, as we explain below. Note that we do not assume that the recognition function is smooth, the only smoothness assumed here is on the sets of viewing directions for which the recognition function is inconsistent.

We next formally define these assumptions. Given a function f, for every object O let $E_f(O)$ denote the set of viewing directions for which f is not guaranteed to be consistent (Figure 1 graphically presents $E_f(O)$). That is, $E_f(O)$ is a set of points on the unit sphere which consists of the viewing directions that correspond to the confusable images of O. Our second assumption is that two similar objects have similar sets of confusable viewing direction. The similarity of two objects is taken to be the Euclidian distance between the two objects in R^{3n}, where an object with n points is regarded as a point in R^{3n}. Our smoothness assumption is on the function $E_f(O)$. That is, if the distance between two objects O and O' is small, then the difference in $E_f(O)$ and $E_f(O')$ is small as well.

We next consider our assumption that the set of confusable viewing directions for a small neighborhood of objects is bounded. Let us define, for an object O, the set of confusable viewing directions $\hat{E}_f(O,\epsilon)$, for all objects in the ϵ-neighborhood of O. Formally,

$$\hat{E}_f(O,\epsilon) \quad = \bigcup_{|O-O'|<\epsilon} E_f(O')$$

That is, $\hat{E}_f(O,\epsilon)$ contains all the viewing directions for which an image of at least one object in the neighborhood of radius ϵ around O is confusable from that viewpoint (Figure 1). In particular, if O and O' are two objects such that $|O - O'| < \epsilon$, and $E_f(O)$, $E_f(O')$, are the sets of views in which their images are confusable, then $E_f(O) \bigcup E_f(O') \subseteq \hat{E}_f(O,\epsilon)$. If we take, for example, ϵ to be infinity, then $\hat{E}_f(O,\epsilon)$ will be the set of viewing directions for which f is inconsistent on at least one of the objects in $\mathcal{U}$. In this case, we limit the confusable views of all objects simultaneously. On the other hand, if we take $\epsilon = 0$, then we independently limit the set of confusable viewing direction of each object. In this case the smoothness assumption is dropped. It can be shown that when the smoothness assumption is dropped, then our results do not hold anymore. In particular, if the smoothness assumption is dropped, it is possible to construct an example such that for each object there exists a single confusable direction, but the $\mathcal{I}_{rec}$-reachability partition is non-trivial.

We define $\Phi(O,\epsilon)$ to be the measure (on the unit sphere) of $\hat{E}_f(O,\epsilon)$. We next establish the proposition that in the universal case, even if f makes errors on almost half of the sphere of viewing directions, then the $\mathcal{I}_{rec}$-reachability partition consists of the entire universe. It follows that any universal recognition function must be constant even if $\Phi(O,\epsilon)$ is substantial. Similarly, every optimal recognition function for which $\Phi(O,\epsilon)$ is substantial cannot be universal.

Proposition 2. *Let $\mathcal{I}_{rec}$ be the set of recognizable images of a recognition function f defined on weak perspective projection of all possible 3D point objects. Assume that for every object O there exists a neighborhood ϵ_O such that $\Phi(O,\epsilon_O) < D$. D is fixed for all objects and taken in the proof to be half of the unit sphere. Then any two objects are $\mathcal{I}_{rec}$-reachable and cannot be discriminated by a universal recognition function.*

Proof: We assumed above that if f is consistent on a given image then it is also consistent on the image scaled, rotated, reflected, and translated by any factor. We can therefore consider only objects that are points inside the unit sphere, B_0^{3n}, of R^{3n}.

Let O_a and O_b be two objects in B_0^{3n}. We have to show that O_a and O_b are $\mathcal{I}_{rec}$-reachable. We first construct a reachability sequence by ignoring the confusable images of f. (This sequence can be used to prove that in the consistent case every two objects are reachable as in Burns *et al.* , 1992, and Moses and Ullman, 1992). We then show how the reachability sequence can be modified to become $\mathcal{I}_{rec}$-reachability sequence.

Let the first object in the sequence be O_a, and the last object be O_b. Each object in the sequence consists of the same points as the previous one, except for one point of object O_a that is replaced by a new point from the object O_b. Formally, the i-th object in the sequence ($1 \leq i \leq n$) is given by

$$O_i = (\mathbf{p}_1^b, \mathbf{p}_2^b, ..., \mathbf{p}_{i-1}^b, \mathbf{p}_i^a, ..., \mathbf{p}_n^a)$$

where $\mathbf{p}_i^a$ and $\mathbf{p}_i^b$ are the i-th points of O_a and O_b, respectively. (Vectors are denoted here and in subsequent sections by boldface characters.) By the sequence construction, every two successive objects differ by a single point. The direction in which the two objects project to the same image is the vector defined by the two non-identical points of the successive objects.

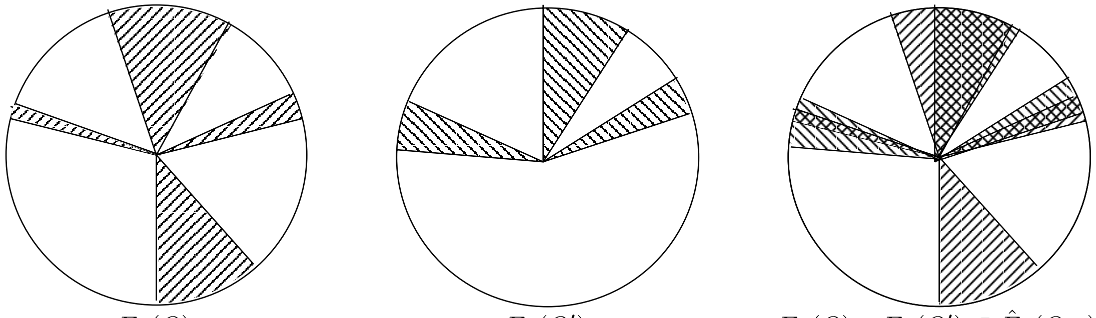

Fig. 1. $E_f(O)$ and $E_f(O')$ are the set of confusable viewing direction for two similar objects O and O', respectively. $E_f(O) \cup E_f(O')$ is the union of the two sets $E_f(O)$ and $E_f(O')$.

If the sequence constructed above is also an $\mathcal{I}_{rec}$-reachability sequence then the construction is terminated. Otherwise, using the three claims below we show that it is always possible to modify the sequence to become an $\mathcal{I}_{rec}$-reachability sequence. This is done by adding sub-sequences which are $\mathcal{I}_{rec}$-reachable connecting pairs of objects that share a confusable image in the original sequence. We next list the three claims. The proofs of these claims are given in the Appendix.

Claim 1. *There exists a fixed $\epsilon > 0$ and a fixed $\hat{D} < D$ such that for every object $O \in B_0^{3n}$, $\Phi_f(O, \epsilon) \leq \hat{D}$. That is, instead of having spheres of different radii, we now have at each point a sphere with fixed radius, ϵ, such that $\Phi_f(O, \epsilon) \leq \hat{D}$.*

In the following two claims, let (O_i, O_{i+1}) be a pair of successive objects from the original sequence that do not share a recognizable image. Note that by construction successive objects in this sequence differ by a single point. Let d be the distance between O_i and O_{i+1} as measured in R^{3n}. Let $0 < \delta < \epsilon$ be a constant that is a function of ϵ and $\hat{D}$ (in the proof of Claim 2 the value of δ is explicitly defined).

Claim 2. *If $d < \delta$ then there exists an object O_c such that the two pairs (O_i, O_c) and (O_c, O_{i+1}) share a recognizable image.*

Claim 3. *If $d \geq \delta$, then there exists a sequence of objects, $O_{i,1}, ..., O_{i,n}$ (where $O_i = O_{i,1}$ and $O_{i,n} = O_{i+1}$) such that each pair of successive objects in this sequence differ in a single point, and the distance between a pair of successive objects in this sequence is less than δ.*

We now show that these claims suffice. Given the initial reachability sequence, replace each pair of successive objects that do not share a recognizable image (O_i, O_{i+1}) such that $d \geq \delta$ by the subsequence $(O_i = O_{i,1}, O_{i,2}, ..., O_{i,n} = O_{i+1})$, using Claim 3. In the new sequence, the distance between all successive objects is less than δ. It is therefore possible to replace each such pair of objects that do not share a recognizable image (O_k, O_{k+1}) by the subsequence (O_k, O_c, O_{k+1}) using Claim 2. As a result, an $\mathcal{I}_{rec}$-reachability sequence consisting of a finite number of objects is obtained. □

This result is an extension of previous results (Burns, Weiss and Riseman, 1992; Clemens, 1990; Moses and Ullman, 1992), showing that at the universal level a consistent recognition function is necessarily a constant function. We conclude that the same limitations hold even if the recognition function is allowed to miss-recognize half of the object's images.

5.2. *Contour images*

In this section we consider objects that consist of 3D contours, rather than a set of discrete 3D points. The images in this case are binary contours which depend only on the camera position and the 3D shape of the object's contour. We prove below that any pair of general 3D contour objects are reachable. It follows (see Section 4) that every universal recognition function that is invariant to viewing position of 3D contour objects is the constant function. Similarly, every recognition function defined on general 3D contour objects cannot be optimal.

Proposition 3. *The reachability partition of the universe consists of 3D contour objects is trivial (any 3D contour objects are reachable).*

Proof: Let O_a and O_b be two general contour objects. In general an object may consists of several contours. We next show how each of two non-identical object contours are reachable. Here we cannot replace one point at a time as we did earlier in Section 5.1, but we can nevertheless construct a simple reachability sequence connecting the two objects. Let the parametric form of the two different contours of O_a and O_b be $O_a = \{x_a(t), y_a(t), z_a(t)\}$ and $O_b = \{x_b(t), y_b(t), z_b(t)\}$ respectively ($0 \leq t \leq 1$). The reachability sequence between O_a and O_b consists of the following four objects:

$$\begin{array}{ccccc} O_a & = & O_1 & = & \{x_a(t), y_a(t), z_a(t)\} \\ & & O_2 & = & \{x_a(t), y_a(t), z_b(t)\} \\ & & O_3 & = & \{x_a(t), y_b(t), z_b(t)\} \\ O_b & = & O_4 & = & \{x_b(t), y_b(t), z_b(t)\} \end{array} \tag{1}$$

It can be easily verified that O_1 and O_2 share the view $\mathbf{v}_{12} = (0, 0, 1)^T$. Similarly the common view of the pair of objects (O_2, O_3) and (O_3, O_4) are given by $\mathbf{v}_{23} = (0, 1, 0)^T$ and $\mathbf{v}_{34} = (1, 0, 0)^T$, respectively. If the two contours are perpendicular to each other, it is possible to avoid degenerated contours that project to a point, by adding an intermediate contour that is neither perpendicular to O_1 contour nor to O_b contour. (Note that a similar construction can be used for objects that consist of 3D point sets.) It follows that the sequence given in Eq. 1 is a reachability sequence, and therefore any universal recognition function that is consistent with respect to viewpoint of contour objects must be the constant function. □

5.3. Consistency with respect to illumination

So far, only binary images of point objects were considered, however, real objects consist of surfaces and their images contain grey level values. In this section we relax our assumption on the objects, and proceed one step toward real objects. With each point of the object, $\mathbf{p}$, we associate not only its 3D location but also a unit surface normal $\hat{\mathbf{N}}(\mathbf{p})$, and a reflectance value ρ_p. An object in this case consists therefore of n surface patches in space rather than n points in space. We further assume that the surface reflectance is Lambertian. An image of a given object now depends on the points locations, the normal and reflectance at each point, the camera position, and also on the illumination condition, that is, the intensity and positions of the light sources. An image now contains more information than before: in addition to the location of the n points, we have the grey levels at each point. Since the images contain now more information than binary images, the question arises as to whether this information can be used in the generalization process. Clearly real objects consist of surfaces rather than patches, but patch objects are of interest as a step toward real objects. In particular, when a recognition function is applied to a set of points, the grey-level values of these points are also given. The question, in this case, is whether these grey-level values can change the limitations of the recognition function.

In this section we will show that any two objects composed of n Lambertian surface patches are reachable. Here a pair of objects have a common image if there exists a viewing direction and illumination condition for which the two images (the points locations in the image as well as their grey levels) are identical. It will follow that any universal recognition function that is consistent with respect to illumination and viewpoint directions must be a constant function. Similarly, every optimal recognition function that is consistent with respect to both illumination and viewpoint conditions of such 3D objects cannot be universal.

Proposition 4. *The reachability partition of the universe of objects consisting of 3D small Lambertian surface patches is trivial (any two such objects are reachable).*

Proof: Let O_a and O_b be two objects. The reachability sequence between the two objects O_a and O_b is obtained by the concatenation of two sequences: a sequence between the objects O_a and O_c, and a sequence between the objects O_c and O_b. The object O_c consists of the same patches location as the O_b-object, but the normal direction, and the albedo at each patch is identical to the normal direction and albedo of the corresponding O_a-patch. The sequence between O_a and O_c can be constructed in the same manner as the consistent reachability sequence in Proposition 2. For each two successive objects in the sequence, the grey-level values of two points of the two successive objects that project to the same location are identical. This is true because we assume here Lambertian reflectance and therefore the grey-level value of a point does not depend on the viewpoint but only on the illumination location, the albedo, and the normal to the points. The normals and the albedo of the points are identical by construction, and the illumination can be taken to be the same for all the objects in this sequence. Hence, O_a and O_c are reachable.

The O_c and O_b objects consist of the same patches location but with different normal direction and albedo at each corresponding patch. Let us construct a sequence such that each successive pair in the sequence will differ in only one patch. The first and the last objects in the sequence are O_c and O_b, respectively. Let $\mathbf{p}_c$ and $\mathbf{p}_b$ be the two non-identical patches in a successive pair. It is left to show that for every two non-identical patches located in the same position, there exists an illumination location such that the grey-level values of $\mathbf{p}_c$ and $\mathbf{p}_b$ are identical (the same illumination for both patches). Note that for any such illumination, which in particular is identical to both objects, the images of the two objects will be identical. This is because all the patches except two have remained the same, and therefore the grey level in the image will only be affected at these two points. (The viewpoint does not come into play here since the patches location are the same and we assume Lambertian reflectance function of the objects.)

Let $\hat{\mathbf{N}}(\mathbf{p}_c)$ and $\hat{\mathbf{N}}(\mathbf{p}_b)$ be the unit vectors in the normal directions, ρ_c and ρ_b be the albedo of the points $\mathbf{p}_c$ and $\mathbf{p}_b$ respectively. The intensity values at the points are given by

$$I_c = \rho_c \mathbf{l} \cdot \hat{\mathbf{N}}(\mathbf{p}_c)$$

$$I_b = \rho_b \mathbf{l} \cdot \hat{\mathbf{N}}(\mathbf{p}_b),$$

where the direction of $\mathbf{l}$ is pointing to the light source location and the magnitude of $\mathbf{l}$ is the light source intensity (for details regarding the images of Lambertian surfaces see Horn 1977). For $\mathbf{p}_c$ and $\mathbf{p}_b$ to have identical grey-level values in an image, it must be shown that there exists a vector $\mathbf{l}$ satisfying the following equation:

$$\rho_b \, \mathbf{l} \cdot \hat{\mathbf{N}}(\mathbf{p}_b) = \rho_c \, \mathbf{l} \cdot \hat{\mathbf{N}}(\mathbf{p}_c)$$

Such $\mathbf{l}$ clearly exists, because it is defined by one linear equation in three variables. The vector $\mathbf{l}$ should also satisfy $\mathbf{l} \cdot \hat{\mathbf{N}}(\mathbf{p}_c) > 0$ and $\mathbf{l} \cdot \hat{\mathbf{N}}(\mathbf{p}_b) > 0$. This is again possible since if $\mathbf{l} \cdot \hat{\mathbf{N}}(\mathbf{p}_b) < 0$, then $\mathbf{l} \cdot \hat{\mathbf{N}}(\mathbf{p}_c) < 0$ as well, and $-\mathbf{l}$ can then be selected for the solution. If $\hat{\mathbf{N}}(\mathbf{p}_c) = \hat{\mathbf{N}}(\mathbf{p}_b)$ but $\rho_c \neq \rho_b$, the solution for $\mathbf{l}$ is such that $\mathbf{l} \cdot \hat{\mathbf{N}}(\mathbf{p}_c) = 0$. We can add one intermediate object to the sequence, with albedo ρ_b and normal $\hat{\mathbf{N}}(\mathbf{p}_c)' \neq \hat{\mathbf{N}}(\mathbf{p}_c)$. □

Note that the same proof holds for two objects that have a uniform color (albedo). If the object is not Lambertian but has a specular component, the intensity at each point depends on the viewpoint, the surface normal, the light source position, and some other surface specular parameters (Phong 1975). It remains an open question whether it is also possible to construct a non-lambertian reachability sequence between non-lambertian objects.

6. Class-based recognition functions

The analysis of universal recognition yielded mainly negative results: under a wide range of conditions, universal recognition function do not exist. In this section we turn to the class level, and show that for many classes of interest useful class-based recognition functions can be defined. The universe of a class-based recognition function is limited to a set of objects, usually large or even infinite such as the set of planar objects, or bilaterally symmetric objects. Here we show that the existence of a non-trivial reachability partition for a class of objects depends on the class in question. It follows that the existence of a consistent class-based recognition function, and the existence of a class-based optimal recognition function for a given class of objects, depends on the class in question. Furthermore we show that for non-consistent recognition function the existence of $\mathcal{I}_{rec}$-reachability partition depends also on the set of images excluded. In this section we also consider classes that are defined not only by constraining the set of objects, but also by excluding a relatively small set of views from the viewing sphere. For such classes, the existence of optimal recognition function, depends not only on the class but also on the restricted set of allowed views.

We will first give as a natural example a class of objects for which the reachability partition is trivial. It follows that, by analogy with the universal case, every consistent recognition function for this class must be constant. Similarly, class-based recognition functions for this class cannot be optimal. For some other classes of objects, e.g. the linear combination of prototypical objects, it can be shown that non-trivial reachability partition exists (Basri and Moses, 1998). Other examples of classes with excluded views that has non-trivial reachability partition were also considered in studies of invariances. For a given class of objects, there exists a non-trivial reachability partition if and only if there exists a class-based recognition function that is invariant to viewing condition of objects in the class. It follows that known examples of invariance schemes that are specific to classes of objects can be used to demonstrate the existence of class-based recognition functions (e.g., invariance for the classes of planar objects, bilaterally symmetrical objects, and polyhedral objects). The class-based invariant schemes suggested in the past were not necessarily consistent on all possible views of the objects. Furthermore, they were also not necessarily optimal for each subset of objects from the class. In this section we will demonstrate, using the class of bilaterally symmetric objects, that by excluding few of its views and studying its reachability partition it is possible to define for the new class (the class with excluded views) a class-based recognition function that is optimal.

Existing class-based recognition systems consider the problem of invariant representations only under changes in viewpoint. In the previous section we have shown that a consistent universal recognition function with respect to both illumination condition and viewpoint does not exist. Here we will show that by restricting the universe of a recognition function to a class of objects, a consistent class-based recognition function with respect to both illumination and viewpoint does exist. We will show this by analyzing a class of 3D Lambertian patches that are bilaterally symmetric.

6.1. *The class of a prototypical object*

To demonstrate that the reachability partition of a class of objects can be trivial, we consider here the simple class of a prototypical object. The class of a prototypical object is defined as the set of all objects that are sufficiently close to a given generic object. For example, one can consider all faces that lie within a certain distance from some prototypical face. For objects composed of n points in space, such a class can be thought of as a sphere in R^{3n} around the prototypical object. For such classes of objects, all the results established for the universal case hold. The point to note is that the entire reachability sequence will lie within the boundaries of the class in question. It follows that for such classes a class-based consistent recognition function must be constant and cannot discriminate between any two objects. Similarly, any class-based recognition function for such classes cannot be optimal.

6.2. *The class of bilaterally symmetric objects*

In this section we first demonstrate the existence of a nontrivial $\mathcal{I}_{rec}$-reachability partition for a class of objects, where the consistency of the recognition function is defined with respect to both viewpoint and illumination changes. We consider the class of bilaterally symmetric objects, where the correspondence between pairs of symmetric points in the image is given. If we consider all images of this class, taken from any viewpoint, then it can be shown that the reachability partition of this class is trivial, as in the universal case. However, we will show that by eliminating a single viewing direction from the viewing sphere, the perpendicular view to the symmetry plane, the $\mathcal{I}_{rec}$-reachability partition of this class becomes non-trivial. In this case, we consider all images of bilaterally symmetric objects that contain the symmetry or skew symmetry of the objects. It follows that a non-trivial recognition function that is consistent on all views but one with respect to viewpoint and illumination direction exists for the class of bilaterally symmetric objects. However, this recognition function will not be optimal for any subset of objects, since the I_{rec}-reachability partition consists of more than a single object. We therefore show that by further restricting the set of images, it is possible to define a recognition function for the class of bilaterally symmetric objects that is optimal for all possible subset of objects of this class. To this end, we will combine the invariance for the class of bilaterally symmetric objects presented here with an invariance suggested in the past by Rothwell *et al.* (1993).

6.2.1. *Consistency with respect to viewing direction*

Consider the reachability partition of the class of bilaterally symmetric objects, where the consistency of the recognition function is with respect to viewpoint changes. The images considered here are assumed to be the weak perspective projection of bilaterally symmetric objects, consisting of 3D points in space. For every point in the image, its symmetric point is assumed to appear as well. Without loss of generality, let a symmetric object be $O = (\mathbf{l}_1, \mathbf{r}_1, ..., \mathbf{l}_n, \mathbf{r}_n)$, where $\mathbf{l}_i = (x_i, y_i, z_i)^T$ and $\mathbf{r}_i = (-x_i, y_i, z_i)^T$ for $1 \leq i \leq n$. That is, $\mathbf{l}_i$ and $\mathbf{r}_i$ are a pair of symmetric points about the y-z plane. Let the new coordinates of the points $\mathbf{l}_i$ and $\mathbf{r}_i$ following a rotation, R, scaling s, projection to the x-y plane, and translation, $\mathbf{t} \in \mathcal{R}^2$ be given by

$$\begin{aligned} \mathbf{l}'_i &= Proj(sR\,\mathbf{l}) + \mathbf{t} \\ &= s\begin{pmatrix} r_{11}x_i + r_{12}y_i + r_{13}z_i \\ r_{21}x_i + r_{22}y_i + r_{23}z_i \end{pmatrix} \quad + \ \mathbf{t} \end{aligned}$$

$$\begin{aligned} \mathbf{r}'_i &= Proj(sR\,\mathbf{r}) + \mathbf{t} \\ &= s\begin{pmatrix} -r_{11}x_i + r_{12}y_i + r_{13}z_i \\ -r_{21}x_i + r_{22}y_i + r_{23}z_i \end{pmatrix} \ + \ \mathbf{t} \end{aligned}$$

We obtain that all the image distances between symmetric paired are scaled by the same factor:

$$d(\mathbf{l}'_i, \mathbf{r}'_i) = ||\mathbf{l}'_i - \mathbf{r}'_i|| = 2s||(r_{11}, r_{21})^T||\, x_i$$

In particular the ratios between the image distances of two pairs of symmetric points is fixed under changes of viewpoint, and is given by

$$\frac{d(\mathbf{l}'_i, \mathbf{r}'_i)}{d(\mathbf{l}'_1, \mathbf{r}'_1)} = \frac{x_i}{x_1}.$$

These ratios define a partition of the class of symmetric objects to equivalence subclasses of non-reachable objects. Let $d_i = d(\mathbf{l}_i, \mathbf{r}_i)$ be the distance between a pair of symmetric points, $\mathbf{l}_i$ and $\mathbf{r}_i$. Define the function h by

$$h(\mathbf{l}_1, \mathbf{r}_1, ..., \mathbf{l}_n, \mathbf{r}_n) = \{\frac{d_2}{d_1}, \frac{d_3}{d_1}, ...\frac{d_n}{d_1}\}$$

Proposition 5. *Two symmetric objects O_a and O_b are reachable if and only if $h(O_a) = h(O_b)$.*

Proof: Let $h(O_a) = h(O_b)$. It must be shown that O_a and O_b are reachable by a sequence of symmetric objects. That is, there exists a sequence of symmetric objects starting with O_a and ending with O_b such that any two successive objects have a projection in common.

Let the two symmetric objects be:

$$\begin{aligned} O_a &= (\mathbf{l}^a_1, \mathbf{r}^a_1, ..., \mathbf{l}^a_n, \mathbf{r}^a_n) \\ O_b &= (\mathbf{l}^b_1, \mathbf{r}^b_1, ..., \mathbf{l}^b_n, \mathbf{r}^b_n). \end{aligned}$$

The first object in the sequence will be O_a, and denote the second object in the sequence by O_c. To construct the sequence, we choose the second object in the sequence to be the object O_a scaled by

$$s = \frac{d^b_1}{d^a_1}$$

where $d^a_1 = d(\mathbf{l}^a_1, \mathbf{r}^a_1)$ and $d^b_1 = d(\mathbf{l}^b_1, \mathbf{r}^b_1)$. That is,

$$O_c = (\mathbf{l}^c_1, \mathbf{r}^c_1, ..., \mathbf{l}^c_n, \mathbf{r}^c_n) = (s\,\mathbf{l}^a_1, s\,\mathbf{r}^a_1, ..., s\,\mathbf{l}^a_n, s\,\mathbf{r}^a_n)$$

By our assumption $h(O_a) = h(O_b)$, that is:

$$\frac{d^a_i}{d^a_1} = \frac{d^b_i}{d^b_1}.$$

It follows

$$d^c_i = sd^a_i = \frac{d^b_1 d^a_i}{d^a_1} = d^b_1\frac{d^b_i}{d^b_1} = d^b_i$$

That is, $d^c_i = d^b_i$. In particular, the symmetric images of O_c and O_b (taken from the frontal view) satisfy $x^c_i = x^b_i$ for every i. However, y^c_i is not necessarily equal to y^b_i. The rest of the reachability sequence, between O_c and O_b is constructed as follows. Each object in the sequence consists of the same points as its preceding object, except for a pair of symmetric points of the object O_c which are replaced by a new pair of symmetric points of the object O_b. The direction for which the two objects project to identical image is the vector connecting the corresponding non-identical points of O_c and of O_b-point. Note that this vector is parallel to the y-z plane, hence the view is frontal and the symmetry of the image is maintained. In this manner we obtain a reachability sequence connecting any two objects, O_a and O_b, for which the relative distances between symmetric points are identical ($h(O_a) = h(O_b)$).

Let $h(O_a) \neq h(O_b)$. It must be shown that O_a and O_b are not reachable by a sequence of symmetric objects. Assume that there exists a sequence of symmetric objects starting with O_a and ending with O_b such that every two successive objects have a projection in common. For every two successive objects, O_i and O_{i+1}, $h(O_i) = h(O_{i+1})$ because O_i and O_{i+1} have a common orthographic projection, and h is independent of the viewing position. It follows that for every two objects, O_i and O_j, in the sequence connecting the objects O_a and O_b, $h(O_i) = h(O_j)$. This contradicts the assumption that $h(O_{a_1}) = h(O_a) \neq h(O_b) = h(O_{a_n})$.

□

6.2.2. Optimal recognition function The consistent class-based recognition function with respect to viewpoint defined above can be used to discriminate only between objects that differ in the relative distance of symmetric points. In particular, objects that consist of pairs of points that differ only in their height (y-component) and depth (z-component) are reachable and cannot be discriminated by a consistent recognition function. Note that we consider here all the images in which the symmetry or the skew-symmetry is presented. We only excluded views taken perpendicular to the symmetry plane.

A different invariance for the class of bilaterally symmetric objects was presented by Rothwell *et al.* (1993). They used the observation that all the midpoints of pairs of symmetric points are located on the same plane. It is therefore possible to use affine coordinates of the midpoints to define the y and z coordinates of each pair of symmetric points (up to an affine transformation). Such an invariant representation cannot discriminate between objects that differ only by the relative distance of pairs of symmetric points from the symmetry plane, which is the invariance we suggested in the previous section. In particular, the invariance proposed by Rothwell *et al.* (1993) is inconsistent on frontal views for such objects. However, it can discriminate between objects that differ in the y and z coordinates of their points which the invariant function suggested in Section 6.2.1 fail to discriminate. Furthermore, the invariance suggested by Rothwell *et al.* will be consistent on a side view, for which our invariance is not defined. By combining our invariance with that of Rothwell *et al.* (1993), we obtain a recognition function that is consistent on all the images (except the frontal views and the side view as mentioned above), and can discriminate between all possible bilaterally symmetric objects. This function is optimal on all possible subsets of the class of bilaterally symmetric objects. This result demonstrates the existence of a tradeoff between consistency, the set of objects that can be discriminated by a given recognition function, and the set of images that the recognition function is defined on. By giving up a restricted set of views, a powerful recognition function can be defined for the class of all 3D bilaterally symmetric objects.

6.2.3. Consistency with respect to illumination We next turn to consider the reachability partition of the class of bilaterally symmetric objects where the consistency of the recognition function is with respect to illumination. The images considered are the weak perspective projection of bilaterally symmetric objects consisting of 3D small Lambertian surface patches. In Section 5 we proved that the reachability partition for similar objects that are not constrained to be bilaterally symmetric must be trivial. The class we consider here is an infinite one and it demonstrates that class constraints may induce a non-trivial reachability partition. It will follow that for such a class a consistent recognition function that can discriminate between subsets of objects from the class does exist. It is left for future research to study a more realistic cases where the objects are not necessarily Lambertian, and the image contains attached and cast shadows, and occlusions.

For every point in the image, its symmetric point from the object is assumed to appear in the image as well. Each object point $\mathbf{p}$ has a surface normal $\hat{\mathbf{N}}(\mathbf{p})$, and a reflectance value ρ_p associate with it. Two symmetric points, $\mathbf{l}_i$ and $\mathbf{r}_i$, have the same value of ρ, and their normals are symmetric about the y-z plane. That is, if $\hat{\mathbf{N}}(\mathbf{l}_i) = (n_x^i, n_y^i, n_z^i)^T$ then $\hat{\mathbf{N}}(\mathbf{r}_i) = (-n_x^i, n_y^i, n_z^i)^T$.

We next show that the $\rho\hat{\mathbf{N}}$ at each object point can be computed up to a global scale factor of the x component and up to a global affine transformation of the y and z components. These two invariances are similar to those computed for the viewpoint case, where the actual point position was computed (rather than the normal) up to a scale factor of the x component and up to an affine transformation of the y and z components.

Let $\mathbf{l}_i$ and $\mathbf{r}_i$ be a pair of symmetric points. The grey-level at a point $\mathbf{l}_i$ and at its symmetric point $\mathbf{r}_i$ are given by $I(\mathbf{l}_i) = \rho_{l_i}\mathbf{l}\cdot\hat{\mathbf{N}}(\mathbf{l}_i)$ and $I(\mathbf{r}_i) = \rho_{r_i}\mathbf{l}\cdot\hat{\mathbf{N}}(\mathbf{r}_i)$. The difference and the average of the grey-level of the two symmetric points are given by

$$\begin{aligned} I(\mathbf{l}_i) - I(\mathbf{r}_i) &= \rho_{l_i}\mathbf{l}\cdot(\hat{\mathbf{N}}(\mathbf{l}_i) - \hat{\mathbf{N}}(\mathbf{r}_i)) \\ \tfrac{1}{2}(I(\mathbf{l}_i) + I(\mathbf{r}_i)) &= \tfrac{1}{2}\rho_{l_i}\mathbf{l}\cdot(\hat{\mathbf{N}}(\mathbf{l}_i) + \hat{\mathbf{N}}(\mathbf{r}_i)) \end{aligned}$$

If $\mathbf{l} = (l_x, l_y, l_z)$ then

$$\begin{aligned} I(\mathbf{l}_i) - I(\mathbf{r}_i) &= 2\rho_{p_i}\mathbf{l}\cdot(n_x, 0, 0)^T \\ &= 2\rho_{p_i}l_x n_x^i \\ I(\mathbf{l}_i) + I(\mathbf{r}_i) &= 2\rho_{p_i}\mathbf{l}\cdot(0, n_y, n_z)^T \\ &= 2\rho_{p_i}(l_y n_y^i + l_z n_z^i) \end{aligned}$$

The relative difference of the grey-level values of two pairs of symmetric points is given by

$$\frac{I(\mathbf{l}_i) - I(\mathbf{r}_i)}{I(\mathbf{l}_1) - I(\mathbf{l}_1)} = \frac{\rho_{p_i} n_x^i}{\rho_{p_1} n_x^1}$$

This ratio is clearly independent of the illumination direction. It can be shown (in a similar manner to proposition 5) that it defines a nontrivial partition of the class of symmetric objects to equivalence subclasses of reachable objects. The second invariance is given by

$$\frac{1}{2}(I(\mathbf{l}_i) + I(\mathbf{r}_i)) = \rho_{p_i}(l_y n_y^i + l_z n_z^i).$$

This invariance is the projection of the normals at all points to the symmetry plane scaled by the point albedo. The $\rho_{p_i} n_y$ and $\rho_{p_i} n_z$ of each of these points can be computed up to a global affine transformation. Note that since the invariance is a function of the points normal and albedo up to affine transformation, it is independent of the illumination.

By combining the results of viewpoint reachability partition and illuminant reachability partition, a recognition function for the class of bilaterally symmetric objects can be constructed that will be consistent with respect to changes of viewpoint as well as illumination condition.

We conclude that although a consistent recognition function with respect to changes in viewpoint and illumination direction does not exist at the universal level, by restricting the universe to a class of objects, a consistent recognition function can be found. It is of interest therefore to attempt to identify large and useful classes of objects for which, similar to the class of bilaterally symmetric objects, class-based consistent recognition functions are possible.

7. Model-based recognition functions

At the model-based level the recognition function is constructed for a given finite set of objects, s. In this case the recognition function is specifically tailored to the set of objects that it is required to recognize. An optimal recognition function always exits at the model-based level since $s = \mathcal{U}$. In the following proposition we prove that a model-based optimal recognition function recognize each object from all but finite set of its images.

Proposition 6. *For a given finite set of objects, s, an optimal recognition function fails to recognize each object from at most a finite set of images.*

Proof: Define the value of an optimal recognition function on images of an object $o_i \in s$ to be i on all images of o_i that are not common to other objects in the set s. For images that are common to at least one other object in the set s, define the value to be arbitrary (or equal to the value of one of the objects that project to this image). It is sufficient to show that for a finite set of objects the number of images that are common to two or more objects is finite.

The number of images that are common to two different objects is at most two. This holds because the number of images required to reconstruct the 3D shape of a rigid object is at most three. The number of images that are common to two different objects depends on the camera (the projection model): three images are required for weak perspective projection (Ullman, 1979), two images are sufficient for affine, (Koenderink and Van Doorn, 1991; Ullman and Basri, 1991) perspective (Longuet-Higgins, 1981; Tsai and Huang, 1984), and projective (Faugeras, 1992) projections. □

We conclude that a fixed finite set of objects, has a finite set of confusable images, and an optimal scheme will recognize correctly all the remaining non-confusable images. Imperfect recognition functions also exist, but they may misidentify objects from additional images. Finally, a non-constant consistent recognition function exists if the partition of the set of objects is non-trivial.

There are several examples of model-based recognition systems that compensate for image variations due to changes of viewpoint. Recently, systems that compensate at the model-based level to variations due to changes in illumination conditions were also suggested (Hallinan, 1994; Belhumeur, Hespanha and Kriegman, 1997; Moses, 1993; Shashua, 1992; Viola and Wells, 1995). In general, a model-based system can identify an object in a given image by comparing the image to the models in the system library. The model, in this case, must explicitly or implicitly contain information of the object shape, and reflectance properties (when grey-level images are considered).

8. Summary and Discussion

In this paper we distinguished between three levels at which a recognition system can compensate for variations between images of the same object. We studied the inherent limitations placed on the level at which a recognition system can compensate for image variations due to viewpoint and illumination conditions. We will first briefly summarize the results and then discuss their implications.

Three types of recognition functions were considered for each of the levels, consistent, optimal, and imperfect. Table 2 summarizes whether a consistent, optimal, or imperfect recognition functions can compensate for image variation due to changes of viewing direction and illumination condition at each of the three generalization levels.

Previous studies proved that a universal recognition function cannot discriminate between any two 3D objects that consist of a set of points. Our study extends this result and proves that even when the constraints on the objects and the recognition function are relaxed substantially, the recognition function must still be the constant function. In particular, we showed that at the universal level a recognition function that is consistent with respect to viewing position, illumination condition or both, for all possible point objects, is a constant function. It follows that such a function cannot discriminate between any two objects. Furthermore, we showed that even when the recognition function is allowed to make errors on a substantial fraction of the viewing directions (almost half of the viewing sphere), it must still be the constant function. Finally, a universal recognition function which is defined on images of objects that consist of 3D contours rather than 3D point sets was shown to be a constant function. Due to these inherent limitations, the universal level is usually too broad to be useful in recognition. For the class-based level, it was shown that the existence of non-trivial recognition functions depends on the class in question. Several recognition systems were suggested in the past for specific classes of objects by using functions that are invariant to changes in viewpoint. Our study shows that such class-based recognition functions exist even if the images contain grey-level values at each point. In this case, the recognition function must be invariant to changes of both illumination condition and viewpoint. Furthermore, we demonstrate, using the class of bilaterally symmetric objects, the tradeoff between the set of objects that can be discriminated by a class-based recognition function, and the set of images on which the function is consistent. It is sometimes possible to restrict the set of views and obtain highly discriminative class-based recognition function.

Finally, at the model-based level it is theoretically always possible to define an optimal recognition function that is consistent on most images of the objects and at the same time can discriminate between all the objects in the set.

In the current study objects consisting of points, contours, or surface patches in space were considered. Real objects are more complex. However, many recognition systems proceed by first finding special contours or points in the image, and then applying the recognition process to them. The points or contours found in the first stage are usually projections of stable object features. If such points (or contours) are used, our results of the universal case apply to these systems directly. The extension of these results to surfaces is beyond the scope of this paper and is left for future study.

A general conclusion from this study is that the class-based level of processing can be advantageous in generalizing to novel viewpoint and illumination conditions, since it is more specific than the universal level and more general than model-based schemes. The results established in this paper indicate that universal recognition schemes cannot overcome the variation between images of the same object due to changes in illumination condition and viewpoint. This is particularly noteworthy with respect to illumination, since it is often assumed that illumination can be compensated for by universal operations such as low-pass filtering and edge detection. It follows that a recognition scheme should attempt to compensate for illumination and viewpoint variations at a more specific level of processing, i.e., class-based or model-based. As shown in this study, the use of class-based scheme is possible for some, but not all, classes of objects. Under the conditions examined in this study, a model-based recognition scheme is always sufficient for overcoming image variations due to viewpoint and illumination conditions. However, class-based recognition schemes have an advantage over model-based schemes since knowledge concerning a known class of objects can be used for recognizing novel objects from the class in question, without changing the recognition process. Model-based schemes typically require multiple 2D views (or a detailed 3D model) to recognize a novel object under different illumination and viewpoint conditions. By using a class-based scheme, it becomes possible to generalize for illumination and viewpoint changes based on a

Table 2. Summary of our results. (a) This was shown for viewpoint in (Burns, Weiss and Riseman, 1992; Clemens and Jacobs, 1991; Moses and Ullman, 1992) and for illumination in Proposition 4; (b) Since the reachability partition was shown to be trivial. (c) This was shown for viewpoint in Proposition 2; (d-f) Examples of two classes were given: the class of bilaterally symmetric objects (Section 6.2) for which the reachability is non-trivial with respect to both illumination and viewpoint, and the class of prototypical object (Section 6.1) for which the reachability partition is trivial; (g) A set that contains only two objects that have a common image is an example of a set for which a consistent recognition function does not exists; (h) follows directly from Definition 1 and (i) from Proposition 6.

	Universal	Class	Model
Consistent	(a) Constant	(d) Depends on $\mathcal{U}$	(g) Depends on $\mathcal{U}$
Optimal	(b) Does not exist	(e) Depends on $\mathcal{U}$ and excluded views	(h) Always exists
Imperfect	(c) Does not exist	(f) Depends on $\mathcal{U}$ and excluded views	(i) Always exists

single 2D view. To recognize for instance a face in a novel image, such a scheme will use general properties of the class of faces to compensate for illumination and viewpoint changes of a specific individual. Note that in using such class-based recognition schemes, the system must classify objects in the image before identifying them. For example, in order to identify a particular face, the system must first determine the object class (e.g., a face, a symmetrical object, etc.), then use class-specific process to identify the face. Object classification, a useful process in its own right, is also used here as a first stage for more specific identification.

A recent psychophysical study of the level at which generalization takes place in the human visual system suggests that the class-based level indeed plays an important role in recognizing faces in novel images (Moses, Edelman and Ullman, 1995). The study compared generalization capacity for upright and inverted faces. Inverted face images are known to be more difficult to recognize. The study did not focus on this difficulty however, but on humans' ability to generalize from trained familiar views to novel ones. A considerable difference was found in subjects' ability to generalize from highly familiar to novel images between the upright and inverted conditions. For upright faces, subjects could recognize novel images of a face taken under different illumination and viewpoint after learning a single image of the face in question. This ability was significantly impaired when inverted faces were used in the training as well as in the testing set. The difference in subjects' ability to generalize to novel views in upright and inverted faces indicates that the process involved in overcoming image variations due to changes in viewpoint and illumination direction is not operating at the universal level. At the same time, the ability to generalize to novel face images of upright faces across very large variations in viewpoint and illumination conditions (up to 54^o of camera direction and left vs. right illumination) based on a single 2D view suggests a capacity to use class-based information in the compensation process. Taken together, the computational and psychophysical results suggest that class-based processing is a promising direction in object recognition for dealing with variations due to viewpoint and illumination changes.

Acknowledgment

This paper is based in part on our paper (Moses and Ullman, 1992). The work was supported in part by the Israel Science Foundation, in part by Grant 85041-1-95 from the Ministry of Science and the Arts, and in part by the Koret Foundation. We would like to thank David Jacobs, Ronen Basri, and Shimon Edelman for their comments on an earlier draft of this paper.

Appendix Inconsistent recognition function

In this appendix, the proofs of claims 1, 2, 3, of Proposition 2 are given.

Proof of Claim 1: Let $\bar{B}_0^{3n}$ be the close unit sphere in R^{3n}. By our assumption, for every $O \in \bar{B}_0^{3n}$ there exists an ϵ_O such that $\Phi_f(O, \epsilon_O) < D$. Consider the family of open sets $B^{3n}(O, \epsilon_O/2)$ for every $O \in \bar{B}_0^{3n}$. This is an infinite cover of the unit sphere $\bar{B}_0^{3n}$. Since the sphere $\bar{B}_0^{3n}$ is a compact set, there exists a finite subset, $\{B^{3n}(O_i, \epsilon_i/2)\}_{i=0}^{m}$ that covers $\bar{B}_0^{3n}$. Let ϵ be the minimum radius in this finite cover ($\epsilon = \min(\epsilon_i/2)$, for $0 \leq i \leq m$). Since for each object in this cover, $\Phi_f(O_i, \epsilon_i) < D$, it follows that for each object in the cover, there exists $D_i < D$ such that $\Phi_f(O_i, \epsilon_i) \leq D_i$ for $0 \leq i \leq m$. Let $\hat{D}$ be the maximum of D_i in this finite cover. It follows that for every object in this cover, $\Phi_f(O_i,\epsilon_i) \leq \hat{D} < D$.

Every object $O \in B_0^{3n}$ satisfies $O \in B^{3n}(O_i,\epsilon_i/2)$ for some $0 \leq i \leq m$, since $\{B^{3n}(O_i,\epsilon_i/2)\}_{i=0}^{m}$ is a cover of $\bar{B}_0^{3n}$. In particular, since $\epsilon \leq \epsilon_i/2$ for $0 \leq i \leq m$ it follows that every $B^{3n}(O,\epsilon) \subseteq B^{3n}(O_i,\epsilon_i)$. We thus have:

$$\Phi_f(O,\epsilon) \leq \Phi_f(O_i,\epsilon_i) \leq \hat{D} < D$$

Hence, $\Phi_f(O,\epsilon) \leq \hat{D}$ for every $x \in B_0^{3n}$.

Proof of Claim 2: By the sequence construction, the objects O_i and O_{i+1} differ by only one point. Let O be the object that consists of the $n-1$ identical points of O_i and O_{i+1}. Let $\mathbf{p}_i$ and $\mathbf{p}_{i+1}$ be the non-identical points of O_i and O_{i+1}, respectively. We define the object $O \cup \mathbf{p}$ to be the object that consists of the point $\mathbf{p}$ and the points of O. For example, $O_i = O \cup \mathbf{p}_i$ and $O_{i+1} = O \cup \mathbf{p}_{i+1}$.

By our assumption, the distance between O_i and O_{i+1} is less than δ. Since O_i and O_{i+1} differ only in the points $\mathbf{p}_i$ and $\mathbf{p}_{i+1}$, it follows that the distance between $\mathbf{p}_i$ and $\mathbf{p}_{i+1}$ is less than δ. Let $\mathbf{p}$ be the point $\frac{\mathbf{p}_i+\mathbf{p}_{i+1}}{2}$. The distance between $O\cup\mathbf{p}$ and $O_i = O\cup\mathbf{p_i}$ and the distance between $O \cup \mathbf{p}$ and $O_{i+1} = O \cup \mathbf{p_i}$ are both less than ϵ (since we assume that $\delta < \epsilon$.) It follows that $O \cup \mathbf{p}_i$, $O \cup \mathbf{p}_{i+1} \in B(O \cup \mathbf{p}, \epsilon)$ (a ball of radius ϵ centered at $O \cup \mathbf{p}$).

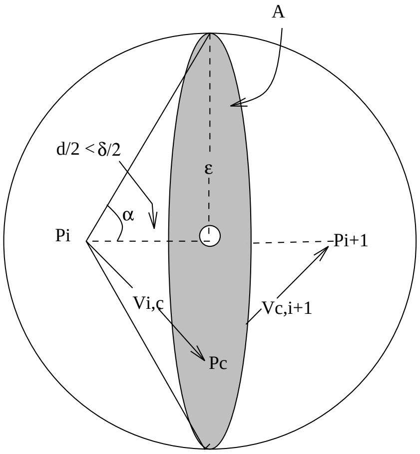

Fig. Inconsistent recognition function.1. The two non-identical points, $\mathbf{p}_i$ and $\mathbf{p}_{i+1}$ in the $B(p,\epsilon)$. A is the equidistance plane between $\mathbf{p_i}$ and $\mathbf{p_{i+1}}$. $\mathbf{v_{i,c}}$ and $\mathbf{v_{c,i+1}}$ are the directions of the common image of the object pairs $(O \cup \mathbf{p_i}, O \cup \mathbf{p}_c)$, and $(O \cup \mathbf{p_c}, O \cup \mathbf{p}_{i+1})$, respectively. (see proof of claim 2).

Consider the plane A of equidistant points from $\mathbf{p}_i$ and $\mathbf{p}_{i+1}$ in the sphere $B(p,\epsilon)$. We claim that there exists a point $\mathbf{p}_c$ on A such that both object pairs $(O \cup \mathbf{p}_i, O \cup \mathbf{p}_c)$ and $(O \cup \mathbf{p}_{i+1}, O \cup \mathbf{p}_c)$ share recognizable images. We will next prove that if such a point does not exist, then it contradicts Claim 1. Assume that for every point $\mathbf{p_c} \in A$, one of the pairs $(O\cup\mathbf{p}_c, O\cup\mathbf{p}_i)$ or $(O\cup\mathbf{p}_c, O\cup\mathbf{p}_{i+1})$ does not share a recognizable image. Let $\mathbf{v_{i,c}}$ and $\mathbf{v_{c,i+1}}$ be the directions of the common image of the object pairs $(O \cup \mathbf{p_i}, O \cup \mathbf{p}_c)$, and $(O \cup \mathbf{p_c}, O \cup \mathbf{p}_{i+1})$, respectively (see Figure 2). It follows that at least one of these directions are confusable, that is $\mathbf{v_{i,c}} \in \hat{E}(O \cup \mathbf{p}, \epsilon)$ or $\mathbf{v_{c,i+1}} \in \hat{E}(O \cup \mathbf{p}, \epsilon)$. What is left to be shown is that if for every point $\mathbf{p_c} \in A$ either the object pair $(O \cup \mathbf{p}_i, O \cup \mathbf{p}_c)$ or the object pair $(O \cup \mathbf{p}_{i+1}, O \cup \mathbf{p}_c)$ does not share a recognizable image, then the measure $\Phi_f(O \cup \mathbf{p}, \epsilon) > \hat{D}$, which contradict Claim 1.

Consider the cone of directions, C with apex α^o from the point $\mathbf{p}_i$ to the plane A, where $\alpha = \arctan(\epsilon/(d/2)) > \arctan(\epsilon/(\delta/2))$. This cone of directions consists of all possible directions that result in a common image of a pair of objects $(O\cup\mathbf{p}_i, O\cup\mathbf{p}_c)$, where $\mathbf{p}_c \in A$. If an object $O \cup \mathbf{p_c}$ ($\mathbf{p_c} \in A$) does not share a recognizable image with $O \cup \mathbf{p_i}$ it follows that the viewing directions: $\mathbf{v}_{i,c} \in C$ is confusable; In particular it follows that this view $\mathbf{v}_{i,c} \in \hat{E}(O,\epsilon)$. Let B_i be the set of views that correspond to a confusable image of the object $O \cup \mathbf{p_i}$ and $O \cup \mathbf{p_c}$ where $\mathbf{p_c} \in A$. It follows that $B_i \subseteq C \cap \hat{E}(O,\epsilon)$. In a similar manner we can consider $B_{i+1} \subseteq C \cap \hat{E}(O,\epsilon)$ to be the set of directions that corresponds to the confusable images of the objects $O \cup \mathbf{p_{i+1}}$ and any of the objects $O\cup\mathbf{p_c} \in A$. To deal with these directions we define the direction $Corr(\mathbf{v}_{c,i+1}) = \mathbf{v_{i,c}}$ to be the corresponding direction to the direction $\mathbf{v_{c,i+1}}$, in the sense that they both relate to the same object $O \cup \mathbf{p_c}$. We consider the corresponding directions of the confusable directions associated with $O \cup \mathbf{p}_{i+1}$ and $O \cup \mathbf{p_c}$ where $\mathbf{p_c} \in A$. In this case $Corr(B_{i+1}) \subseteq C$.

Let $B = B_i \cup Corr(B_{i+1}) \subseteq C$. That is B consists of all viewing directions that are associated with all objects $O \cup \mathbf{p_c}$ where $\mathbf{p_c} \in A$, such that either $O \cup \mathbf{p_i}$ or $O \cup \mathbf{p_{i+1}}$ share confusable images. We now show that there exists a $\mathbf{v} \in C$ such that $\mathbf{v} \notin B$. If such $\vec{v}$ exists, it follows that the objects $O \cup (\mathbf{p_i} + \mathbf{v})$ share a recognizable image with O_i and O_{i+1} (where $(\mathbf{p_i} + \mathbf{v}) \in A$). Since $B \subseteq C$, it is sufficient to show that B is contained in a set that its measure is less than the measure of C. The measure of $\hat{E}(O,\epsilon)$ is less than $\hat{D}$ by assumption. Furthermore, by our assumption, if a given viewing direction $\mathbf{v} \in \hat{E}(O,\epsilon)$ then also $-\mathbf{v} \in \hat{E}(O, \epsilon)$ (where $-\mathbf{v}$ is the vector in the opposite direction to $\mathbf{v}$). It follows that the measure of $C \cap \hat{E}(O, \epsilon) \leq \hat{D}/2$. Similarly it can be shown that $Corr(B_{i+1})$ is contained in a set with a measure less than $\hat{D}/2$. It follows that B is contained in a set with a measure less than $\hat{D}/2 + \hat{D}/2 = \hat{D}$. We can choose δ such that the measure of C is larger than $\hat{D}$ (since when δ tend to zero the measure of C is tend to half the unit sphere and $\hat{D} < D$). Since $d \leq \delta$, it follows that B is contained in a set with a measure strictly smaller then the measure of C.

Proof of Claim 3: Let O_i and O_{i+1} be two successive objects that share a confusable image. By the sequence construction we can assume that $O_i = O \cup \mathbf{p_i}$ and $O_{i+1} = O \cup \mathbf{p_{i+1}}$. The distance between O_i and O_{i+1} is $> 2\epsilon/\tan(\alpha)$. Let $n = |\mathbf{p_{i+1}} - \mathbf{p_i}|/\delta$. We can construct the sequence by choosing close enough objects in the following manner:

$$O_{i,j} = O \cup (\mathbf{p_i} + \frac{j}{n}(\mathbf{p_{i+1}} - \mathbf{p_i}))$$

References

Adini, Y., Moses, Y. and Ullman, S. 1997. Face recognition: the problem of compensating for illumination changes. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 19:721-732.

Basri, R. and Moses, Y. 1998. When is it possible to identify 3D objects from single images using class constraints? In *International Conference on Computer Vision,* pp. 541-548.

Belhumeur, P.N., Hespanha, J.P. and Kriegman, D.J. 1997. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7): 711-720.

Biederman, I. 1985. Human image understanding: recent research and a theory. *Computer, Graphics, and Image Processing,* 32:29-73.

Brunelli, R. and Poggio, T. 1991. HyperBF networks for real object recognition. In *IJCAI,* Australia, pp. 1278-1284.

Burns, J.B. Weiss, R.S. and Riseman, E.M. 1992. The non-existence of general-case view-invariants. In J. L. Mundy and A. Zisserman, Eds., *Geometrical Invariance in Computer Vision,* M.I.T. Press.

Canny, J. F. 1986. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence,* 8:679-698.

Clemens, D.J. and Jacobs, D.W. 1990. Model-group indexing for recognition. In *Proc. Image Understanding Workshop,* pp. 604-613.

Clemens, D.J. and Jacobs, D.W. 1991. Space and time bounds on indexing 3D models from 2D images. *Pattern Analysis and Machine Intelligence,* 13(10):1007-1017.

Craw, I., Ellis, H. and Lishman, J.R. 1987. Automatic extraction of face-features. *Pattern Recognition Letters,* 5:183-187.

Daugman, J. G. 1985. Uncertainty relation for resolution in space, spatial frequency and orientation, optimized by two dimensional cortical filters. *Journal of Optical Society of America,* 2:1160-1169.

Davis, L. S. 1975. A survey of edge detection techniques. *Computer Graphics and Image Processing,* 4:248-270.

Faugeras, O.D. 1992. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proc. European Conference on Computer Vision,* pp. 563-564.

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.

Fischler, M. A., and Bolles, R. C. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM,* 24:381-395.

Hallinan, P.W. A low-dimensional representation of human faces for arbitrary lighting conditions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* pp. 995-999.

Haralick, R. M. 1984. Digital step edges from zero crossings of second directional derivatives. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 6:58-68.

Hubel, D.G. and Wiesel, T.N. 1962. Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology,* 160:106-154.

Hubel, D.G. and Wiesel, T.N. 1968. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology,* 195:215-243.

Huttenlocher, D. P., and Ullman, S. 1990. Recognizing solid objects by alignment with an image. *International Journal of Computer Vision*, 5(2): 195-212.

Jacobs, D. 1992. Space efficient 3D model indexing. In *IEEE Conference on Computer Vision and Pattern Recognition,* pp. 439-444.

Kanade, T. 1977. *Computer recognition of human faces*. Birkhauser Verlag.

Kaya, Y. and Kobayashi, K. 1972. A basic study of human face recognition. In S. Watanabe, Ed., *Frontiers of Pattern Recognition,* pp. 265-289.

Koenderink, J. J., and Van Doorn, A. J. 1991. Affine structure from motion. *Journal of the Optical Society of America,* 8(2):377–385.

Lamdan, Y., Schwartz, J.T. and Wolfson, H.J. 1987. Affine invariant model-based object recognition. *IEEE Transaction on Robotics and Automation*, 6:578-589.

Lamdan, Y. and Wolfson, H. 1988. Geometric hashing: a general and efficient recognition scheme. In *Proceedings of the 2nd International Conference on Computer Vision*, pp. 238-251.

Longuet-Higgins, H. C. 1981. A computer algorithm for reconstructing a scene from two projections. *Nature,* 293:133-135.

Lowe, D. G. 1987. Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, 31:355-395.

Marcelja, S. 1980. Mathematical description of the responses of simple cortical cells. *J. Optical Soc.,* 70:1297-1300.

Marr, D. and Hildreth, E. 1980. Theory of edge detection. *Proc. R. Soc. Lond. B,* 207:187-217.

Moses, Y. 1993. Face recognition: generalization to novel images. Ph.D Thesis, Weizmann Institute of Science.

Moses, Y., Edelman, S. and Ullman, S. 1996. Generalization to novel images in upright and inverted faces. *Perception,* 25:443-461.

Moses, Y., and Ullman, S. 1992. Limitation of Non-model-based recognition schemes. In *Proc. European Conference on Computer Vision,* pp. 820-828.

Nixon, M. 1985. Eye spacing measurements for facial recognition. *SPIE Application of Digital Image Processing VIII,* 575:279-285.

Pollen, D., and Ronner, S. 1983. Visual cortical neurons as localized spatial frequency filters. *IEEE Transactions on System, Man and Cybernetics,* SMC-13: 907-916.,

Rothwell, C. A., Forsyth, D. A., Zisserman, A. and Mundy, J. L. 1993. Extracting projective structure from single perspective views of 3D point sets. In *Proceeding of International Conference on Computer Vision,* pp. 573-582.

Rothwell, C.A., Zisserman, A., Forsyth, D.A. and Mundy, J.L. 1992. Canonical frames for planar object recognition. In *European Conference on Computer Vision,* pp. 757-772.

Shashua, A. 1992. Illumination and view position in 3D visual recognition. In J.E. Moody, J. E. Hanson, and R.P. Lippman, Eds., *Advances in Neural Information Processing Systems 4*, Morgan Kaufman, pp. 68-74.

Torre, V., and Poggio, T. 1986. On edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 8:147-163.

Tsai, R.Y. and Huang, T.S. 1984. Uniqueness and estimation of three dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 6:13–27.

Ullman, S. 1979. *The interpretation of visual motion*. MIT Press.

Ullman, S. 1989. Aligning pictorial descriptions: an approach to object recognition. *Cognition*, 32:93-254.

Ullman, S. and Basri, R. 1991. Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 13:992-1005.

Viola, P., and Wells III, W. M. 1995. Alignment by maximization of mutual information. In *Fifth International Conference on Computer Vision*, pp.16-23.

Warrington, E.K, and Taylor, A.M. 1978. Two categorical stages of object recognition. *Perception,* 7:152-164.

Weinshall, D. 1993. Model-based invariants for 3D vision. *International Journal on Computer Vision,* 10(1):27-42.

Wong, K.H., Law, H.M. and Tsang, P.W.M. 1989. A system for recognising human face. In *Proc. ICASSP*, pp. 1638-1642.

Yuille, A. L., Cohen, D.C. and Hallinan, P.W. 1992. Feature extraction from faces using deformable templates. *International Journal of Computer Vision,* 8(2):99-111.

Zisserman, A., Forsyth, D., Mundy, J., Rothwell, C., Liu, J. and Pillow, N. 1995. 3D Object Recognition Using Invariance. *Artificial Intelligent*, 78(1-2):239-288.