Limitations of Non Model-Based Recognition Schemes

Yael Moses and Shimon Ullman

Dept. of Applied Mathematics and Computer Science,
The Weizmann Institute of Science, Rehovot 76100,
Israel

Abstract. Approaches to visual object recognition can be divided into *model-based* and *non model-based* schemes. In this paper we establish some limitations on non model-based recognition schemes. We show that a consistent non model-based recognition scheme for general objects cannot discriminate between objects. The same result holds even if the recognition function is imperfect, and is allowed to mis-identify each object from a substantial fraction of the viewing directions. We then consider recognition schemes restricted to classes of objects. We define the notion of the *discrimination power* of a consistent recognition function for a class of objects. The function's discrimination power determines the set of objects that can be discriminated by the recognition function. We show how the properties of a class of objects determine an upper bound on the discrimination power of any consistent recognition function for that class.

1 Introduction

An object recognition system must recognize an object despite dissimilarities of images of the same object due to viewing position, illumination conditions, other objects in the scene, and noise. Several approaches have been proposed to deal with this problem. In general, it is possible to classify these approaches into *model-based* vs. *non model-based* schemes. In this paper we examine the limitations of non model-based recognition schemes.

A number of definitions are necessary for the following discussion. A *recognition function* is a function from 2-D images to a space with an equivalence relation. Without loss of generality we can assume that the range of the function is the real numbers, R. We define a *consistent recognition function* for a set of objects to be a recognition function that has identical value on all images of the same object from the set. That is, let s be the set of objects that f has to recognize. If v_1 and v_2 are two images of the same object from the set s then $f(v_1) = f(v_2)$.

A *recognition scheme* is a general scheme for constructing recognition functions for particular sets of objects. It can be regarded as a function from sets of 3-D objects, to the space of recognition functions. That is, given a set of objects, s, the recognition scheme, g, produces a recognition function, $g(s) = f$. The *scope* of the recognition scheme is the set of all the objects that the scheme may be required to recognize. In general, it may be the set of all possible 3-D objects. In other cases, the scope may be limited, e.g., to 2-D objects, or to faces, or to the set of symmetric objects. A set s of objects is then selected from the scope and presented to the recognition scheme. The scheme g then returns a recognition function f for the set s. A recognition scheme is considered consistent if $g(s) = f$ is consistent on s as defined above, for every set s from the scheme's scope.

A model-based scheme produces a recognition function $g(s) = f$ that depends on the set of models. That is, there exist two sets s_1 and s_2 such that $g(s_1) \neq g(s_2)$ where the inequality is a function inequality. Note that the definition of model-based scheme in our discussion is quite broad, it does not specify the type of models or how they are used. The schemes developed by Brooks (1981), Bolles & Cain (1982), Grimson & Lozano-Pérez (1984,1987), Lowe (1985), Huttenlocher & Ullman (1987), Ullman (1989) and Poggio & Edelman (1990) are examples of model-based recognition schemes.

A non model-based recognition scheme produces a recognition function $g(s) = f$ that does not depend on the set of models. That is, if g is a non model-based recognition scheme, then for every two sets s_1 and s_2, $g(s_1) = g(s_2)$, where the equality is a function equality.

Non model-based approaches have been used, for example, for face recognition. In this case the scope of the recognition scheme is limited to faces. These schemes use certain relations between facial features to uniquely determine the identity of a face (Kanade 1977, Cannon *et al.* 1986, Wong *et al.* 1989). In these schemes, the relations between the facial features used for the recognition do not change when a new face is learned by the system. Other examples are schemes for recognizing planar curves (see review Forsyth *et al.* 1991).

In this paper we consider the limitations of non model-based recognition schemes. A consistent non model-based recognition scheme produces the same function for every set of models. Therefore, the recognition function must be consistent on every possible set of objects within the scheme's scope. Such a function is *universally consistent*, that is, consistent for objects in its scope.

A consistent recognition function of the set s should be invariant to at least two types of manipulations: changes in viewing position, and changes in the illumination conditions. We first examine the limitation of non model-based schemes with respect to viewing position, and then to illumination conditions.

In examining the effects of viewing position, we will consider objects consisting of a discrete set of 3-D points. The domain of the recognition function consists of all binary images resulting from scaling of orthographic projection of such discrete objects on the plane. We show (Section 2) that every consistent universal recognition function with respect to viewing position must be trivial, i.e. a constant function[1]. Such a function does not make any distinctions between objects, and therefore cannot be used for object recognition. On the other hand it can be shown that in a model-based scheme it is possible to define a nontrivial consistent recognition function that is as discriminating as possible for every given set of objects (see Moses & Ullman 1991)

The human visual system, in some cases, misidentifies an object from certain viewing positions. We therefore consider recognition functions that are not perfectly consistent. Such a recognition function can be inconsistent for some images of objects taken from specific viewing positions. In Section 3.1 we show that such a function must still be constant, even if it is inconsistent for a large number of images (we define later what we consider "large"). We also consider (Section 3.2) imperfect recognition functions where the values of the function on images of a given object may vary, but must lie within a certain interval.

Many recognition schemes deal with a limited scope of objects such as cars, faces or industrial parts. In this case, the scheme must recognize only objects from a specific class (possibly infinite) of objects. For such schemes, the question arises of whether there

[1] A similar result has been independently proved by Burns *et al.* 1990 and Clemens & Jacobs 1990.

exists a non-trivial consistent function for objects from the scheme's scope. The function can have in this case arbitrary values for images of objects that do not belong to the class. The existence of a nontrivial consistent function for a specific class of objects depends on the particular class in question. In Section (4) we discuss the existence of consistent recognition function with respect to viewing position for specific classes of objects. In Section (4.1) we give an example of a class of objects for which every consistent function is still a constant function. In Section (4.2) we define the notion of the function *discrimination power*. The function discrimination power determines the set of objects that can be discriminated by a recognition scheme. We show that, given a class of objects, it is possible to determine an upper bound for the discrimination power of any consistent function for that class. We use as an example the class of symmetric objects (Section 4.3).

Finally, we consider grey level images of objects that consist of n small surface patches in space (this can be thought of as sampling an object at n different points). We show that every consistent function with respect to illumination conditions and viewing position defined on points of the grey level image is also a constant function.

We conclude that every consistent recognition scheme for 3-D objects must depend strongly on the set of objects learned by the system. That is, a general consistent recognition scheme (a scheme that is not limited to a specific class of objects) must be model-based. In particular, the invariant approach cannot be applied to arbitrary 3-D objects viewed from arbitrary viewing positions. However, a consistent recognition function can be defined for non model-based schemes restricted to specific class of objects. An upper bound for the discrimination power of any consistent recognition function can be determined for every class of objects.

It is worth noting here that the existence of invariant features to viewing position (such as parallel lines) and invariant recognition function for 2-D objects (see review Forsyth *et al.* 1991) is not at odds with our results. Since, the invariant features can be regarded as model-based recognition function and the recognition of 2-D objects is a recognition scheme for a specific class of objects (see section 4).

2 Consistent function with respect to viewing position

We begin with the general case of a universally consistent recognition function with respect to viewing position, i.e. a function invariant to viewing position of all possible objects. The function is assumed to be defined on the orthographic projection of objects that consist of points in space.

Claim 1: Every function that is invariant to viewing position of all possible objects is a constant function.

Proof. A function that is invariant to viewing position by definition yields the same value for all images of a given object. Clearly, if two objects have a common orthographic projection, then the function must have the same value for all images of these two objects.

We define a *reachable sequence* to be a sequence of objects such that each two successive objects in the sequence have a common orthographic projection. The function must have the same value for all images of objects in a reachable sequence. A *reachable object* from a given object is defined to be an object such that there exists a reachable sequence starting at the given object and ending at the reachable object. Clearly, the value of the function is identical for all images of objects that are reachable from a single object.

Every image is an orthographic projection of some 3-D object. In order to prove that the function is constant on all possible images, all that is left to show is that every two objects are reachable from one another. This is shown in Appendix 1. □

We have shown that any universal and consistent recognition function is a constant function. Any non model-based recognition scheme with a universal scope is subject to the same limitation, since such a scheme is required to be consistent on all the objects in its scope. Hence, any non model-based recognition scheme with a universal scope cannot discriminate between any two objects.

3 Imperfect recognition functions

Up to now, we have assumed that the recognition function must be entirely consistent. That is, it must have exactly the same value for all possible images of the same objects. However, a recognition scheme may be allowed to make errors. We turn next to examine recognition functions that are less than perfect. In Section 3.1 we consider consistent functions with respect to viewing position that can have errors on a significant subset of images. In Section 3.2 we discuss functions that are almost consistent with respect to viewing position, in the sense that the function values for images of the same object are not necessarily identical, but only lie within a certain range of values.

3.1 Errors of the recognition function

The human visual system may fail in some cases to identify correctly a given object when viewed from certain viewing positions. For example, it might identify a cube from a certain viewing angle as a 2-D hexagon. The recognition function used by the human visual system is inconsistent for some images of the cube. The question is whether there exists a nontrivial universally consistent function, when the requirements are relaxed: for each object the recognition function is allowed to make errors (some arbitrary values that are different from the unique value common to all the other views) on a subset of views. The set should not be large, otherwise the recognition process will fail too often.

Given a function f, for every object x let $E_f(x)$ denote the set of viewing directions for which f is incorrect ($E_f(x)$ is defined on the unit sphere). The object x is taken to be a point in R^n. We also assume that objects that are very similar to each other have similar sets of "bad" viewing directions. More specifically, let us define for each object x, the value $\Phi(x, \epsilon)$ to be the measure (on the unit sphere) of all the viewing directions for which f is incorrect on at least one object in the neighborhood of radius ϵ around x. That is, $\Phi(x_0, \epsilon)$ is the measure of the set $\bigcup_{x \in B(x_0,\epsilon)} E_f(x)$. We can now show that even if $\Phi(x, \epsilon)$ is rather substantial (i.e. f makes errors on a significant number of views), f is still the trivial (constant) function. Specifically, assuming that for every x there exist an ϵ such that $\Phi(x, \epsilon) < D$ (where D is about 14% of the possible viewing directions), then f is a constant function. The proof of this claim can be found in Moses & Ullman (1991).

3.2 "Almost consistent" recognition functions

In practice, a recognition function may also not be entirely consistent in the sense that the function values for different images of the same object may not be identical, but only close to one another in some metric space (e.g., within an interval in R). In this case,

a threshold function is usually used to determine whether the value indicates a given object.

Let an *object neighborhood* be the range to which a given object is mapped by such an "almost consistent" function. Clearly, if the neighborhood of an object does not intersect the neighborhoods of other objects, then the function can be extended to be a consistent function by a simple composition of the threshold function with the almost consistent function. In this case, the result of the general case (Claim 1) still holds, and the function must be the trivial function.

If the neighborhoods of two objects, a and b, intersect, then the scheme cannot discriminate between these two objects on the basis of images that are mapped to the intersection. In this case the images mapped to the intersection constitute a set of images for which f is inconsistent. If the assumption from the previous section holds, then f must be again the trivial function.

We have shown that an imperfect universal recognition function is still a constant function. It follows that any non model-based recognition scheme with a universal scope cannot discriminate between objects, even if it is allowed to make errors on a significant number of images.

4 Consistent recognition functions for a class of objects

So far we have assumed that the scope of the recognition scheme was universal. That is, the recognition scheme could get as its input any set of (pointwise) 3-D objects. The recognition functions under consideration were therefore universally consistent with respect to viewing position. Clearly, this is a strong requirement. In the following sections we consider recognition schemes that are specific to classes of objects. The recognition function, in this case must still be consistent with respect to viewing position, but only for objects that belong to the class in question. That is, the function must be invariant to viewing position for images of objects that belong to a given class of objects, but can have arbitrary values for images of objects that do not belong to this class.

The possible existence of a nontrivial consistent recognition function for an object class depends on the particular class in question. In Section (4.1) we consider a simple class for which a nontrivial consistent function (with respect to viewing position) still does not exist. In Section (4.2) we discuss the existence of consistent functions for certain infinite classes of objects. We show that when a nontrivial consistent function exist, the upper bound of any function discrimination power can be determined. Finally, we use the class of symmetric objects (Section 4.3) in order to demonstrate the existence of consistent function for an infinite class of objects and its discrimination power.

4.1 The class of a prototypical object

In this section, we consider the class of objects that are defined by a generic object. The class is defined to consist of all the objects that are sufficiently close to a given prototypical object. For example, it is reasonable to assume that all faces are within a certain distance from some prototypical face. The class of prototypical objects composed of n points in space, can be thought of as a sphere in R^{3n} around the prototypical object.

The results established for the unrestricted case hold for such classes of objects. That is, every consistent recognition function with respect to viewing position of all the objects that belong to a class of a given prototypical object is a constant function. The proof for this case is similar to the proof of the general case in Claim 1.

4.2 Discrimination power

Clearly, some class invariants exist. A simple example is the class of eight-point objects with the points lying on the corners of some rectangular prism, together with the class of all three-point objects (since at least 4 points will always be visible of the eight-point object). In this example the function is consistent for the class, all the views of a given object will be mapped to the same value. However, the function has a limited discrimination power, it can only distinguish between two subclasses of objects. In this section we examine further the discrimination power of a recognition function.

Given a class of objects, we first define a *reachability partition* of equivalence subclasses. Two objects are within the same equivalence subclass if and only if they are reachable from each other. Reachability is clearly an equivalence relation and therefore it divides the class into equivalence subclasses. Every function f induces a partition into equivalent subclasses of its domain. That is, two objects, a and b, belong to the same equivalent subclass if and only if $f(a) = f(b)$. Every consistent recognition function must have identical value for all objects in the same equivalence subclass defined by the reachability partition (the proof is the same as in Claim 1). However, the function can have different values for images of objects from different subclasses. Therefore, reachability partition is a refinement of any partition induced by a consistent recognition function. That is, every consistent recognition function cannot discriminate between objects within the same reachability partition subclass.

The reachability subclasses in a given class of objects determines the upper bound on the discrimination power of any consistent recognition function for that class. If the number of reachability subclasses in a given class is finite, then it is the upper bound for the number of values in the range of any consistent recognition function for this class. In particular, it is the upper bound for the number of objects that can be discriminated by any consistent recognition function for this class. Note that the notion of reachability and, consequently, the number of equivalence classes, is independent of the particular recognition function. If the function discrimination power is low, the function is not very helpful for recognition but can be used for classification, the classification being into the equivalence subclasses.

In a non model-based recognition scheme, a consistent function must assign the same value to every two objects that are reachable within the scope of the scheme. In contrast, a recognition function in a model-based scheme is required to assign the same value to every two objects that are reachable within the set of objects that the function must in fact recognize. Two objects can be unreachable within a given set of objects but be reachable within the scope of objects. A recognition function can therefore discriminate between two such objects in a model-based scheme, but not in a non model-based scheme.

4.3 The class of symmetric objects

The class of symmetric objects is a natural class to examine. For example, schemes for identifying faces, cars, tables, etc, all deals with symmetric (or approximately symmetric) objects. Every recognition scheme for identifying objects belonging to one of these classes, should be consistent only for symmetric objects.

In the section below we examine the class of bilaterally symmetric objects. We will determine the reachability subclasses of this class, and derive explicitly a recognition function with the optimal discrimination power. We consider images such that for every point in the image, its symmetric point appears in the image.

Without loss of generality, let a symmetric object be $(0, p_1, p_2, ..., p_{2n})$, where $p_i = (x_i, y_i, z_i)$ and $p_{n+i} = (-x_i, y_i, z_i)$ for $1 \leq i \leq n$. That is, p_i and p_{n+i} are a pair of symmetric points about the $y \times z$ plane for $1 \leq i \leq n$. Let $p_i^r = (x_i^r, y_i^r, z_i^r)$ be the new coordinates of a point p_i following a rotation by a rotation matrix R and scaling by a scaling factor s. The new x-coordinates are: $x_i^r = s(x_i r_{11} + y_i r_{12} + z_i r_{13})$ and $x_{n+i}^r = s(-x_i r_{11} + y_i r_{12} + z_i r_{13})$. In particular, for every pair of symmetric points p_i and p_{n+i}, $(x_i^r - x_{n+i}^r)/(x_1^r - x_{n+1}^r) = x_i/x_1$ hold.

In the same manner it can be shown that the ratios between the distances of two pairs of symmetric points do not change when the object is rotated in space and scaled. We claim that these ratios define a nontrivial partition of the class of symmetric objects to equivalence subclasses of unreachable objects. Let d_i be the distance between a pair of symmetric points p_i and p_{n+i}. Define the function h by
$h(0, p_1, ..., p_{2n}) = (\frac{d_2}{d_1}, \frac{d_3}{d_1}, ... \frac{d_n}{d_1})$.

Claim 2: Every two symmetric objects a and b are reachable if and only if $h(a) = h(b)$. (The proof of this claim can be found in Moses & Ullman (1991).)

It follows from this Claim that a consistent recognition function with respect to viewing position defined for all symmetric objects, can only discriminate between objects that differ in the relative distance of symmetric points.

5 Consistent recognition function for grey level images

So far, we have considered only binary images. In this section we consider grey level images of Lambertian objects that consist of n small surface patches in space (this can be thought of as sampling an object at n different points). Each point p has a surface normal N_p and a reflectance value ρ_p associate with it. The image of a given object now depends on the points' location, the points' normals and reflectance, and also on the illumination condition, that is, the level of illumination, and the position and distribution of the light sources.

An image now contains more information than before: in addition to the location of the n points, we now have the grey level of the points. The question we consider is whether under these conditions objects may become more discriminable then before by a consistent recognition function. We now have to consider consistent recognition functions with respect to both illumination condition and viewing position. We show that a nontrivial universally consistent recognition function with respect to illumination condition and viewing position still does not exists.

Claim 3: Any universally consistent function with respect to illumination condition and viewing position, that is defined on grey level images of objects consisting of n surface patches, is the trivial function.

In order to prove this claim, we will show that every two objects are reachable. That is, there exists a sequence of objects starting with the first and ending with the second object, and every successive pair in the sequence has a common image. A pair of objects has a common image if there is an illumination condition and viewing position such that the two images (the points' location as well as their grey level) are identical. The proof of this claim can be found in Moses & Ullman (1991).

We conclude that the limitation on consistent recognition functions with respect to viewing position do not change when the grey level values are also given at the image points. In particular, it follows that a consistent recognition scheme that must recognize

objects regardless of the illumination condition and viewing position must be model-based.

6 Conclusion

In this paper we have established some limitations on non model-based recognition schemes. In particular, we have established the following claims:
(a) Every function that is invariant to viewing position of all possible point objects is a constant function. It follows that every consistent recognition scheme must be model-based.
(b) If the recognition function is allowed to make mistakes and mis-identify each object from a substantial fraction of viewing directions (about 14%) it is still a constant function.

We have considered recognition schemes restricted to classes of objects and showed the following: For some classes (such as classes defined by prototypical object) the only consistent recognition function is the trivial function. For other classes (such as the class of symmetric objects), a nontrivial recognition scheme exists. We have defined the notion of the discrimination power of a consistent recognition function for a class of objects. We have shown that it is possible to determine the upper bound of the function discrimination power for every consistent recognition function for a given class of object. The bound is determined by the number of equivalence subclasses (determined by the reachability relation). For the class of symmetric objects, these subclasses were derived explicitly.

For grey level images, we have established that the only consistent recognition function with respect to viewing position and illumination conditions is the trivial function.

In this study we considered only objects that consist of points on surface patches in space. Real objects are more complex. However, many recognition schemes proceed by first finding special contours or points in the image, and then applying the recognition process to them. The points found by the first stage are usually projections of stable object points. When this is the case, our results apply to these schemes directly. For consistent recognition functions that are defined on contours or surfaces, our result do not apply directly, unless the function is applied to contours or surfaces as sets of points. In the future we plan to extend the result to contours and surfaces.

Appendix 1

In this Appendix we prove that in the general case every two objects are reachable from one another.

First note that the projection of two points, when viewed from the direction of the vector that connects the two points, is a single point. It follows that for every object with $n-1$ points there is an object with n points such that the two objects have a common orthographic projection. Hence, it is sufficient to prove the following claim:

Claim 4: Any two objects that consists of the same number of points in space are reachable from one another.

Proof. Consider two arbitrary rigid objects, a and b, with n points. We have to show that b is reachable from a. That is, there exists a sequence of objects such that every two successive objects have a common orthographic projection.

Let the first object in the sequence be $a_1 = a = (p_1^a, p_2^a, ..., p_n^a)$ and the last object be $b_1 = b = (p_1^b, p_2^b, ..., p_n^b)$. We take the rest of the sequence, $a_2, ..., a_n$ to be the objects: $a_i = (p_1^b, p_2^b, ..., p_{i-1}^b, p_i^a, ..., p_n^a)$. All that is left to show is that for every two successive objects in the sequence there exists a direction such that the two objects project to the same image. By the sequence construction, every two successive objects differ by only one point. The two non-identical points project to the same image point on the plane perpendicular to the vector that connects them. Clearly, all the identical points project to the same image independent of the projection direction. Therefore, 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. □

References

1. Bolles, R.C. and Cain, R.A. 1982. Recognizing and locating partially visible objects: The local-features-focus method. *Int. J. Robotics Research*, **1**(3), 57-82 .
2. Brooks, R.A. 1981. Symbolic reasoning around 3-D models and 2-D images, *Artificial Intelligence J.*, **17**, 285-348.
3. Burns, J. B., Weiss, R. and Riseman, E.M. 1990. View variation of point set and line segment features. *Proc. Image Understanding Workshop, Sep.*, 650-659.
4. Cannon, S.R., Jones, G.W., Campbell, R. and Morgan, N.W. 1986. A computer vision system for identification of individuals. *Proc. IECON 86 0, WI.*, **1**, 347-351.
5. Clemens, D.J. and Jacobs, D.W. 1990. Model-group indexing for recognition. *Proc. Image Understanding Workshop, Sep.*, 604-613.
6. Forsyth, D., Mundy, L., Zisserman, A., Coelho, C., Heller A. and Rothwell, C. 1991. Invariant Descriptors for 3-D object Recognition and pose. *IEEE Trans. on PAMI.* **13**(10), 971-991.
7. Grimson, W.E.L. and Lozano-Pérez, T. 1984. Model-based recognition and localization from sparse data. *Int. J. Robotics Research*, **3**(3), 3-35.
8. Grimson, W.E.L. and Lozano-Pérez, T. 1987. Localizing overlapping parts by searching the interpretation tree. *IEEE Trans. on PAMI.* **9**(4), 469-482.
9. Horn B. K.P. 1977. Understanding image intensities, *Artificial Intelligence J.*. **8**(2), 201-231
10. Huttenlocher, D.P. and Ullman, S. 1987. Object recognition using alignment. *Proceeding of ICCV Conf., London*, 102-111.
11. Kanade, T. 1977. Computer recognition of human faces. *Birkhauser Verlag. Basel and Stuttgart.*
12. Lowe, D.G. 1985. Three dimensional object recognition from single two-dimensional images. *Robotics research Technical Report 202, Courant Inst. of Math. Sciences, N.Y. University.*
13. Moses, Y. and Ullman S. 1991. Limitations of non model-based recognition schemes. *AI MEMO No 1301, The Artificial Intelligence Lab., M.I.T.*
14. Phong, B.T. 1975. Illumination for computer generated pictures. *Communication of the ACM* , **18**(6), 311-317.
15. Poggio T., and Edelman S. 1990. A network that learns to recognize three dimensional objects. *Nature*, **343**, 263-266.
16. Ullman S. 1977. Transformability and object identity. *Perception and Psychophysics*, **22**(4), 414-415.
17. Ullman S. 1989. Alignment pictorial description: an approach to object recognition. *Cognition*, **32**(3), 193-254.
18. Wong, K.H., Law, H.H.M. and Tsang P.W.M, 1989. A system for recognizing human faces, *Proc. ICASSP*, 1638-1642.

This article was processed using the LaTeX macro package with ECCV92 style