EasyChair Preprint

№ 10

Representation, Reconstruction and Recognition of 3D Objects

Mark Burgin and Allen Klinger

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

March 8, 2018

Representation, Reconstruction and Recognition of 3D Objects

Mark Burgin and Allen Klinger

University of California, Los Angeles
Los Angeles, CA 90095

Abstract. In this paper, we further develop the approach to image recognition based on 3D structural information representation and utilization of super-recursive algorithms. The approach suggested in this paper synthesizes feature recognition techniques with structural shape representations. In addition, problems of object identification, weak object identification and viable object identification are considered.

Keywords: perception, recognition, algorithm, computer vision, image

1. Introduction

Visual perception depends on two things: object recognition and image identification. The first is a general process while the latter involves applying multiple aspects of knowledge. Banerjee, et al, (1990) gives an explanation of the basics of computer vision. Here we isolate several aspects of that overall problem, leading to a new framework. That framework is presented here; it clarifies different steps toward the goal of recognizing objects and shows the importance of super-recursive algorithms in certain practical imaging situations. One particular conclusion then is that computations from visual input have thus far met with limited success in part because of a failure to recognize the continuous (rather than halting) nature of image processes.

This paper considers image recognition under the following conditions. Initial information consists of: first, a set R_A of two-dimensional (2D) images of a three-dimensional (3D) object (shape) A; and second, a collection $\boldsymbol{C}$ of sets R_i where each consists of 2D images of a 3D object (shape) A_i ($i = 1, 2, 3, \dots, n$). The *recognition problem* is to find the set R_i that best fits information that we have about the object A, that is, the set R_A . When such a set R_i is found, we say that the object A is *recognized* as the object A_i .

A restricted form of the recognition problem is the *identification problem* where we seek a set R_i that coincides with the set R_A . When such a set R_i is found, we say that the object A is *identified* as the object A_i .

One more form of the recognition problem is the *weak identification problem* where we seek a set R_i that contains the set R_A as its subset, i.e., $R_A \subseteq R_i$. When such a set R_i is found, we say that the object A is *weakly identified* as the object A_i.

Weak identification differs from (and is weaker than) identification. If an object A is identified as the object A_i, then A is weakly identified as the object A_i as any set is a subset of itself. However, it is possible that an object A is weakly identified as the object A_i, but later some new information about A arrives. E.g., the system can acquire new 2D images of the object A. Any new images involve extending the set R_A to the set R_A' where that new set stops being a subset of the set R_i. The result causes A to no longer be weakly identified as the object A_i.

One more form of the recognition problem is the *viable identification problem*. In that case, we seek a set R_i such that the set R_A contains the set R_i as its subset, i.e., $R_i \subseteq R_A$. When such a set R_i is found, we say that the object A is *viably identified* as the object A_i. Viable identification is weaker than identification: if an object A is identified as the object A_i, then A is viably identified as the object A_i because any set is a subset of itself. However, not every object A is viably identified as the object A_i can be identified as the object A_i because the set R_A can be strictly larger than the set R_i.

We first consider techniques and categories that apply to image recognition.

2. Models and Structural Representation

To understand an object composed of the concatenation of 3D shapes, we usually consider segmentation or region-growing methods that terminate. Frequently that is only possible when one adds restrictive assumptions. Indeed a common approach to object recognition has been to assume that objects have certain invariant properties present in all views (Banerjee, et al, 1990). Then to create *object recognition* one forms a function from the set of object views to the set of real numbers. This is the base of one main approach to image recognition. In it, the whole recognition process is broken down into extraction of a number of different properties followed by a final decision based on these properties. Similarities between properties of the given object and properties of the previously viewed objects are determined and used in the recognition process. It is possible to extract a huge quantity of properties and different conditions are used to stop the process and to go to the next stage. That is why, for its efficient realization, this stage requires *super-recursive algorithms* (Burgin, 2005) with *any-time algorithms* (Zilberstein, 1996) as their delimiters. Another way to describe this is *building discriminating feature spaces*.

However, in the majority of cases, chosen properties are not expected to be entirely invariant but only to lie within a certain range. This brings us to feature spaces and the concept of approximate or fuzzy invariance, which is considered in neoclassical analysis (Burgin, 1995). Properties of different objects may have partially overlapping ranges. However, using a sufficient number of properties allows one to identify each object from a given sample. In this approach, an object is treated as a point in the feature space. Invariant properties and features are often hierarchically organized to make recognition algorithms simpler.

Another approach to object recognition is based on invariant properties. Specifically this applies to the theory of high-order invariants (Gibson, 1979). In it, invariant object properties are reflected in high-order invariants, which are then used in object recognition.

One more approach to object recognition is based on decomposition of objects into their constituent parts and reconstruction of object shapes using some geometrical primitives. We assume that any object can be decomposed into (and reconstructed from) a small set of generic components. The components are obtained by mapping properties and features onto a structural description of an object. The mapping has to be stable, i.e., preserved across different views.

As Biederman (2007) writes, shape is the major route by which we gain knowledge about our visual world. Hence we now turn to the shape recognition problem. One of the most popular approaches uses structural encoding of geometrical shapes. To be able to apply mathematical tools, we now formalize this problem.

The goal is to recognize a given object. We assume that an object is (or at least, has) a physical body and this body has some shape.

As in (Banerjee, et al, 1990), a geometrical 3D *shape position* is a subset of the 3D space $\boldsymbol{R}^3$ that satisfies the following conditions:

(1) A is compact and does not have isolated points.

(2) The interior of A is not empty.

These conditions are similar to but not the same as conditions from (Banerjee, et al, 1990). Their third condition is implied by our first condition because $\boldsymbol{R}^3$ is a metric space (Kuratowski, 1966).

Taking an equivalence relation E between geometrical 3D shape positions, we define the *pure geometrical* 3D *shape* of an object as an equivalence class of the relation E. Usually, this equivalence relation E is defined by admissible shape transformations. Often such transformations as translations (shifts), rotations, and dilation are considered. For instance, according to (Banerjee, et al, 1990), two objects have the same shape if and only if one is a translation, dilation and rotation of the other. Unfortunately these conditions are insufficient in some cases. For instance, face recognition demands taking into account local transformations to represent different expressions of the same face. A face image conveys much more than just shape: e.g., viewing position, illumination conditions, and facial expression. Thus, any face recognition system must take into account the changes in face appearance induced by these factors, representing these changes by local transformations.

However, in many cases the object's shape does not give a complete characterization. Some object A may have many other additional properties and characteristics besides shape, e.g., color or weight. Attaching these characteristics to the shape of A, we obtain a *labeled geometrical* 3D *shape* of the object A. To do this, we take a *geometrical* 3D *shape state*, which is a geometrical 3D *shape position*, plus the values of the label, i.e., the values of those properties and characteristics that are attached to shapes. Thus, an object recognition program can use both shape information and additional parameters.

As we know shapes of many objects are extremely complex when they are examined with unlimited precision. That is why a natural approach to shape representation, analysis, recognition, and understanding deals with finite ensembles of data. There are three main types of such representations: parametric, imaging, and structural.

In the *parametric approach*, a shape R is described by a system of – mostly numerical – parameters that characterize object features, such as eyes, mouth, and nose.

In the *imaging approach*, a shape R is portrayed by a system of generally 2D images (its views).

In contrast, a shape R can be represented via a *structural approach*, through simpler items from a set of primitive elements. The following representation procedure is organized in these terms.

A base B of primitive elements is chosen (constructed). An example of such a base is the system of geons suggested by Biederman (1987).

Geons, or *geometric icons*, are simple 3D forms, such as cubes, spheres, cylinders or cones that satisfy the following conditions:

1. They are view-invariant, i.e., they can be identified from different angles.

2. They are stable and resistant to visual noise, making recognition of geons robust to occlusion and degradation.

3. They are discriminable, i.e., each geon can be potentially distinguished from others almost from all viewpoints.

Note that from some viewpoints different geons can look the same. For instance, from an end-on view, a cylinder and cone look like a sphere.

A shape that is built of geons is called a *geon complex*.

Geon complexes only approximate the majority of real-life shapes. Thus, the process of finding the best approximating geon complex may require a super-recursive algorithm. This is true because on each finite step we cannot be sure in all cases that we already have the best approximation for all cases.

Hoffman and Richards (1986) suggested another object decomposition/reconstruction scheme based on description and recognition of contours. Here the shape primitives are called codons.

In a general case, when we have a base *B* of primitive elements, a construction (shape) built of elements from *B* is called a *B-complex*. For instance, in algebraic topology, simplicial complexes are very useful and thus, popular (Spanier, 1966). They are built of simplexes.

B-complexes can be represented by various structures such as 3D images, systems of 2D images, and diagrams (cf., for example, (Irani and Ware, 2003)). For instance, node-link diagrams are used extensively for many applications, including planning, communications networks, and computer software. The defining features of these diagrams are mostly circular or rectangular nodes connected by linear or arrow-headed lines, the links. A set of guidelines for such diagrams is derived from perception theory and these collectively define the concept of the geon diagram suggested by Irani and Ware (2003).

B-complexes are constructed as approximations for real-life shapes, providing structural representation of these shapes. However, even images are not solely shapes; they have many other properties and characteristics, e.g., color and brightness. For instance, an image of a face depends not only on its shape, but also on the viewing position, illumination conditions, and facial expression (Moses, et al, 1993; Rotshtein, et al, 2007). Thus, any face recognition system must overcome face-appearance changes introduced by such factors. Attaching these characteristics to a *B*-complex, we obtain a *labeled B-complex*, an approximate version of a labeled shape.

3. Cognitive Identification and Recognition Operations

To begin we seek a formulation that considers object recognition as a cognitive operation**.** Here an object may be a person, building, vehicle, geometrical shape, text in some natural language, image on the screen, etc. The following general schema provides a mathematical description of such a process.

Consider two collections. The first, **CO**, is a *discrimination collection* of the problem where **CO** = $\{ A_i ; i = 1, 2, 3, \dots , n\}$. The second is the *model collection* **CM** = $\{ M_i ; i = 1, 2, 3, \dots , n\}$ of the problem. Besides, an object A and its model M_A are also given. We assume that all models have the same type, i.e., they belong to the set **M** of all models of a given type, and that there is a *correlation* or *similarity measure* m on the set **M** with values in a partially ordered set L. In other words, $m: \mathbf{M} \times \mathbf{M} \to L$ is a mapping, which reflects how close two models are to one another. Usually, L is either the interval [0, 1] or the interval [-1, 1] or the whole real line $\boldsymbol{R}$.

Consider the example of a set R_A of two-dimensional (2D) images of a 3D object (shape) A. Then R_A is a model of A. This type of models, i.e., systems of 2D images of a 3D object, corresponds to the classical problem of 3D image recognition given related 2D images. Another type of models for a 3D object consists of its structural representations considered in Section 2. For instance, *B*-complexes give structural object models (descriptions). Labeled *B*-complexes give a parastructural object models (descriptions).

Note that the set **M** of models in general and even a model M_A of one object A can be potentially infinite. This is true because as a rule there are infinitely many admissible transformations of a geometrical shape.

The *recognition problem* is to find an object A_i such that the correlation measure m (M_i , M_A) is the highest for all M_i from the collection **CM**. When such a model M_i is found, we say that the object A is *recognized* as the object A_i .

The *highest-correlation measure* operation appears in the subsequent diagram. We see there four central states and four transition operations, two of which depend on correlation or similarity measures:

$$\begin{array}{ccc} M_{Sp} & \xrightarrow{\;c\;} & M_p \\ m_{Sp}\big\uparrow & & \big\uparrow m_p \\ Sp & \xrightarrow[\;s\;]{} & p \end{array} \qquad (1)$$

This recognition-problem model encompasses other situations and models. For instance, Ulman (1984) suggests that the process of object recognition is the inversion of the following correspondence. Given an object A, a large set of possible views of A is assigned to A. Object recognition starts with one or several, but not too many, views of an object A and goes to finding the original object A or, at least, its name, from these views. Assuming that views of an object constitute its model, we come to the already introduced model for the recognition problem.

Recognition problem difficulties arise because of two possibilities. Sometimes the object model is too simple. This leads to weak-relevance. Alternatively, if it is possible to achieve high relevance it may be at computational costs. In many cases this makes the model very complex and almost intractable. For instance, the set of possible views of a given object is large. However different views of an object can be widely dissimilar.

A restricted form of the recognition problem is the *identification problem* where we seek a model M_i that coincides with the model M_A . When such a model M_i is found, we say that the object A is *identified* as the object A_i . In the categorical setting, this means that the correspondence c in the diagram (1) is the identity mapping.

Definitions above imply the following result.

Proposition 1. If an object A is identified as A_i , then A is recognized as A_i .

A model M_A of an object A is called *attributive* if it is a set R_A of entities that characterize A. These entities are attribute values of A. For instance, in relational databases, object characteristics are used as its attributes (Elmasri and Navathe, 2000). In the case of 3D object recognition, two-dimensional (2D) images are attributes of a three-dimensional object (shape) A. All models M_i are also sets R_i that consist of attribute values of objects A_i . Attributive models allow us to consider two additional identification problems.

In the *weak identification problem*, we seek a set R_i that contains the set R_A as its subset, i.e., $R_A \subseteq R_i$. When such a set R_i is found, we say that the object A is *weakly identified* as the object A_i . Weak identification is weaker than identification because if an object A is identified as the object A_i , then A is weakly identified as the object A_i as any set is a subset of itself. However, it is possible that an object A is weakly identified as the object A_i , but later some new information (in our case, some new 2D images of the object A) about A comes, extending the set R_A to the set R_A' and the new set the set R_A' stops being a subset of the set R_i. As a result, now A cannot be weakly identified as the object A_i .

In the *viable identification problem*, we seek a set R_i such that the set R_A contains the set R_i as its subset, i.e., $R_i \subseteq R_A$. When such a set R_i is found, we say that the object A is *strongly identified* as the object A_i . The viable identification is weaker than identification because if an object A is identified as the object A_i , then A is viably identified as the object A_i as any set is a subset of itself. However, not every object A is viably identified as the object A_i can be identified as the object A_i because the set R_A can be strictly larger than the set R_i .

4. Recognizing 3D Shape From Structure

The first stage of the structural recognition of 3D shapes is to build an adequate structural representation (model) M_A of a given 3D shape (object) A. Here we consider B-complexes as models, e.g., geon complexes.

To build a representation (model) M_A , the recognition program RP needs to know what primitives to use and how to combine them. To find construction primitives, the program RP employs a simplified version PRP of an image recognition program. This program PRP identifies shape primitives and their connections in the given object A, using input information, e.g., the set of given views (2D images) of A. Finding the necessary shape primitives and their connections, the recognition program RP utilizes a construction program to build a B-complex D_1 as the first approximation to the model M_A .

Then shapes A and D_1 are compared, using a distance measure d defined for all plausible 3D shapes. To compare shapes of two 3D objects, it is possible to use measures suggested by Banerjee, et al (1990). The measurement procedure at first normalizes positions and volumes of the compared shapes. Then after normalizing their 3D orientation using the characteristic planes, they are superimposed on each other. The resulting volume of mismatch is taken to be a shape distance between the two 3D objects. It is formalized in the following formula

$$d(A, B) = m((A \setminus B) \cup (B \setminus A))$$

Here A and B are two 3D shapes, \ means the set-theoretical difference of two sets, and m is some measure in $\boldsymbol{R}^3$.

In the analog domain, this shape distance d satisfies metric properties.

Note that in a general case, the function d is non-computable. Consequently, it is more efficient to use super-recursive algorithms because they are more powerful than recursive algorithms (Burgin and Klinger, 2005). As result they allow one to compute value of the distance d for more shapes than recursive algorithms can do.

The chosen measure d allows the recognition program to estimate the distance between shapes A and D_1 . If the distance $d(A, D_1)$ is sufficiently small, then the B-complex D_1 is taken as the geometrical model of the object A. Otherwise, the recognition program RP builds a new B-complex D_2 and compares it to A. This process continues until a sufficiently close to A B-complex D_n is constructed. Then the B-complex D_n is taken as the geometrical model of the object A. After obtaining the model M_A , this model is compared to models from the set $\mathbf{M}$ and the correlations between each pair of models is measured.

To build an adequate correlation or similarity measure $m: \mathbf{M} \times \mathbf{M} \rightarrow L$ in the set $\mathbf{M}$, we can use different approaches depending on the type of models in $\mathbf{M}$.

When $\mathbf{M}$ consists of structural object representations (descriptions) in terms of the shape primitives, e.g., geons, it is possible to take the measure suggested by Biederman (1987) as the correlation measure. Namely, if M_A and M_B are two shapes constructed from shape primitives and treated as the union of these primitives, then

$$m_1(M_A , M_B) = | M_A \cap M_B | - | M_A \setminus M_B | - | M_B \setminus M_A | \qquad (2)$$

or in a more general form

$$m_2(M_A , M_B) = f(| M_A \cap M_B |) - g(| M_A \setminus M_B | + | M_B \setminus M_A |) \qquad (3)$$

Here f and g are positive functions and $|C|$ denotes the number of primitives in the B-complex, e.g., geon complex, C.

However, when two shapes (e.g., geon complexes) have the same primitives, they may be essentially different as shapes having different connections. Thus, it is necessary to develop measures $m_1$and m_2 taking into account connections. Namely, we assume that sets M_A and M_B in formulas (2) and (3) include not only shape primitives but also their connections.

After an adequate correlation measure m is constructed, the last step of the recognition program PR is comparison of the model M_A to all models M_{Ai} . When the model M_i with the best correlation with the model M_A is found, the given object A is recognized as the object A_i . However, new data can cause it to become necessary to repeat the recognition process.

5. Conclusion

We have distinguished several aspects of the computer vision process, comparing them with each other as well as super-recursive processes. Constructing measures that evaluate system value of partial computations is a needed step toward practical utilization of the categories and principles described here.

References

1. Banerjee, D.K ., Parui, S.K., and Majumder, D.D. (1990) Shape based object recognition, *Lecture Notes in Computer Science,* v. 444, pp. 363-374
2. Biederman**,** I. (1987) Recognition by components: A theory of human image understanding, *Psychological Review*, v. 94, No. 2, pp. 115-147
3. Biederman I. (1985) *Human image understanding: Recent research and a theory*. Comp. Vision, Graphics and Image Processing, 32, pp.29–73
4. Biederman, I. (2007) Recent psychophysical and neural research in shape recognition, in *Object Recognition, Attention, and Action*, Ch. 5, pp. 71-88
5. Brooks R. (1981) Symbolic reasoning among 3-dimensional models and 2-dimensional images, Art. Int., v. 17, pp. 285–349
6. Burgin, M. (1995) Neoclassical Analysis: Fuzzy Continuity and Convergence, *Fuzzy Sets and Systems*, v. 75, pp. 291-299
7. M. Burgin, *Super-recursive Algorithms,* Springer, New York/Berlin/Heidelberg, 2005
8. Burgin, M. and Klinger, A. Three Aspects of Super-recursive Algorithms and Hypercomputation or Finding Black Swans, *Theoretical Computer Science*, v. 317, No. 1/3, 2004, pp. 1-11
9. Burgin, M. and Klinger, A. Experience, Generations, and Limits in Machine Learning, *Theoretical Computer Science*, v. 317, No. 1/3, 2004, pp. 71-91
10. Elmasri, R. and Navathe, S.B. (2000) *Fundamentals of Database Systems*, Addison-Wesley Publishing Company, Reading, Massachusetts
11. Gibson, J.J. (1979) *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston
12. Hoffman, D. and Richards, W. (1986) Parts of recognition, in *From Pixels to Predicates*, Ablex Publishing Corp., Norwood, N.J.
13. Irani, P. and Ware, C. (2003) Diagramming information structures using 3D perceptual primitives, *ACM Transactions on Computer-Human Interaction* (TOCHI), v. 10, No. 1, pp. 1 - 19

14. Kuratowski, K. *Topology*, v. 1, Academic Press, Waszawa, 1966

15. Moses, Y., Ullman, S. and Edelman, S. *Generalization to Novel Images in Upright and Inverted Faces*, Weizmann Science Press of Israel, Jerusalem, Israel, 1993

16. Rotshtein, P., Geng, J.J., Driver, J. and Dolan, R.J. (2007) Role of Features and Second-order Spatial Relations in Face Discrimination, Face Recognition, and Individual Face Skills: Behavioral and Functional Magnetic Resonance Imaging Data, *Journal of Cognitive Neuroscience*, v.19, No. 9, pp. 1435-1452

17. Spanier, E.B. Algebraic Topology, McGraw-Hill Book Company, New York, 1966

18. Ulman, S. (1984) Visual routines, Cognition, v. 18, pp. 97-159

19. Zilberstein, S. (1996) Using Anytime Algorithms in Intelligent Systems, *AI Magazine*, v. 17(3), pp.73-83