ARTIFICIAL GENERAL VISION

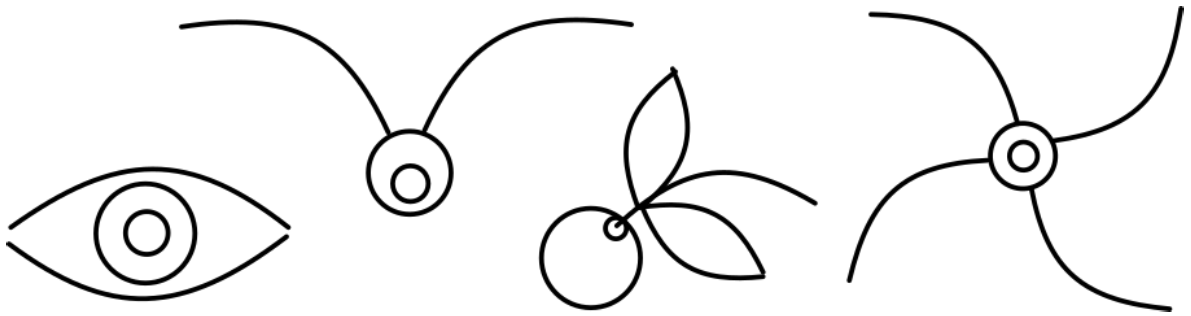
HUMAN-LIKE ACTIVE VISION

Mykola Rabchevskiy
May 16 1

In the natural environment, the main channel for obtaining information by humans and animals is vision. An AGI system that claims to have a level of ability comparable to that of a human must have a vision system close to the human one. To determine what it means "***comparable to human vision***", we will analyze the specifics of human vision.

The most essential feature is the ***dominant role of contours*** over color and brightness. A human can easily identify an object regardless of color and brightness; a pink elephant remains an elephant, a light elephant against a darker background is perceived as identical to a dark elephant against a light background, and an elephant in an outline drawing is identified as simply as in a halftone. Any drawing tutorial teaches a sequence of actions that mimics the analysis of images that our brains perform on a subconscious level - from the main outlines to the outlines of the details and only then - adding color and brightness to the areas bounded by the contours:

orientation of the contour fragments, building a ***structured description*** (model) of the objects and the scene as a whole. The same pieces of the contours are "***assembled***" by the brain into different objects depending on the relative position, orientation, and relative size of the parts:



If a combination of contours as a whole is not recognized as a known object, it can nevertheless be described as a new combination of known components; for example, the image of Pegasus is interpreted as a "horse with wings". The fact that the ***object is unknown does not prevent the construction of its structured description***; if the same thing comes back into the field of vision, it is identified by comparing the memorized description as an ***object that has already been seen before***. The process is similar to a situation with an unknown word in the text: we may not know the meaning of a word, but this prevents us from remembering and recognizing it when we meet it again.

Finally, the third aspect is that the process of analyzing a visible scene — building a structured description and updating it — is performed by the brain ***permanently and simultaneously with a change in the environment***; the visual model includes ***not only what falls into the field of view at the moment, but also what is seen before***, but is now out of sight. The brain builds a visual picture of the entire environment, updating it with information extracted from the current field of view; the gaze (and, accordingly, the field of view) moves according to what our attention is directed to. Closing our eyes, we can mentally "see" the entire environment and take objects without opening our eyes (not always successful due to the lack of visual control over the movement of the hand).

The process as a whole is entirely consistent with the conceptual scheme of building a model of the current situation described earlier in NUTS AND BOLTS OF THE DECISION MAKING.

Comparing the above with the currently popular image recognition systems based on neural networks, it is easy to conclude that image analysis by neural networks does not correspond to human vision. Some of the missing capabilities can be compensated for by combining the capabilities of neural networks with other algorithms. Still, the

fundamental aspect - the use of contours as the primary source of information - in this case, remains "overboard."

Detection of contours by algorithms for ***filtering*** pixel images has long been known, but the result is again formed as a pixel array, not as a set of contour fragments suitable for manipulating with them.

Obviously, the effectiveness of using a structured model of a visual scene depends radically on the complexity of comparing elementary fragments of contours for identity, even though they can differ in size and orientation. This, in turn, depends on the way the information about the elementary contour is presented. This, in turn, depends on the form of presenting information about an elemental fragment of the contour. Ideally, the shape, size, position, and orientation are specified separately so that the description is a set of "orthogonal" characteristics that can be defined and compared independently of each other.

Taking a smooth curve for an atomic fragment of a contour, it is easy to find that the ***minimum circumcircle*** naturally gives us two of the required characteristics: ***size*** (radius) and ***location*** (coordinates of the center). This is ensured by the uniqueness of the minimum circumcircle for any fragment of the contour as well as any combination of such fragments, including hierarchical models of objects in the visual scene.

Curvature versus distance along the curve can define the shape of the smooth curve regardless of position and orientation. Using the distance ratio to the full length of the curve instead of the actual distance eliminates the dependence of the description on the size. In this case, the comparison of the shapes of two atomic contour fragments is reduced to compare two ***curvature functions specified on the interval [0,1]***. Comparison of functions is a very resource-intensive operation, so it makes sense to use its discrete analog instead of such a representation. For this, a ***smooth curve is represented as a sequence of circular arcs*** (with the possibility of degeneration into a straight line segment); each of the sequence segments is characterized by the angle between the tangents (or normals) at the beginning and end of the segment. Since the angle is a dimensionless quantity, such a sequence determines the shape of a smooth curve and does not depend on size, location, and orientation. The mirrored version of the curve has the opposite signs of the values of the segment angles.

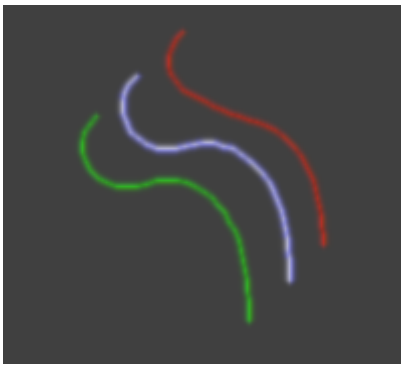
Suppose the number of segments forming an atomic fragment of the contour is set as a power of two. In that case, we get a helpful possibility of a simple transition from a more accurate to a coarser model, reducing the number of segments by half and adding the

value of the angle parameters of the combined pairs of segments. If two fragments represented by exact models are not identical, we can compare their coarser models; repeating the process, we can find the level of accuracy at which they turn out to be similar if some similarity takes place. The comparison process can be carried out in reverse order, starting with the coarsest model from one segment.

Finally, by discretizing the values of the angles, we can reduce the description to a sequence of abstract symbols. The symbols encoding the angles are analogous to letters, and the definition of a fragment of a contour is equivalent to a word.

The last of the orthogonal set of parameters characterizing the atomic fragment of the contour should determine the ***orientation of the fragment***. As such, you can use the ***direction of the start segment***, that is, the angle between the tangent to the contour fragment at its origin and the coordinate axis.

Reconstruction of the visual representation of a fragment of the contour is reduced to the sequential rendering of segments with the corresponding standard transformations of the coordinates of the points (scale, move, rotate). The figure below illustrates how the appearance of the contour fragment changes as the model gets rougher; the original contour of 16 segments is shown in green, simplified versions with 8 and 4 segments are shown in white and red:



When building a structural visual model of an object, a set of things (each can be an atomic outline or a compound object) are combined into one more complex entity. As a result, a hierarchical structure (tree) is formed. The combination of elements as a whole is characterized by the minimum circumcircle over all members` minimum circumcircles and the orientation of the "main" component (the first element in the case of an ordered list).

For each pair of objects of the same level (set of siblings), we can define three dimensionless quantities characterizing the relative position and orientation. The relative position of the second object of the pair relative to the first is specified in a polar

coordinate system with the origin in the center of the minimum circumcircle of the first object and the axis along the initial tangent. The dimensionless distance between objects is equal to the ratio of the actual distance to the sum of their minimum circumcircle radii, providing the requirement for the metric:

distance(A, B) == distance (B, A).

Mutual orientation is defined by the angle between the start tangents. So the structure of a set of objects of one level of the hierarchy can be described by a matrix, each of the cells containing the three dimensionless parameters listed above.

The comparison of such structures, which is the essential operation of visual identification of objects, is reduced to comparing trees and comparing composite objects to compare the mentioned matrices.

The algorithm for detecting and constructing atomic fragments of the contour will be described in the next vision-related chapter.

An important aspect is the relationship of the described processes of visual scene analysis with the analysis of the dynamics of objects in the natural environment. The trajectories of moving objects (see NUTS AND BOLTS OF THE DECISION MAKING), like the fragments of contours, are smooth functions; information update in both cases is permanent and controlled by the attention module. The difference is that the dynamic model and the curvature of the trajectory require information about the time intervals corresponding to the trajectory segments. In a dynamic model, extrapolation is used for forecasting; in visual analysis, extrapolation is used in contouring. The commonality allows software components to be used in both cases - as is probably the case in our brains.

SUMMATION

- *Human vision uses contours as the most informative element of the visual scene.*
- *The construction of the contour model plays the role of compression of the information of the visual scene with a high compression rate.*
- *An elementary fragment of a contour can be represented by data that determine the shape, position, and orientation independently of each other.*
- *Analysis of the visual scene based on the analysis of contours allows you to build a compact symbolic model convenient for logical analysis.*

- *The structural model can be built for unidentified objects and situations and used in the future to identify them.*
- *Analysis of visual information by neural networks implements a radically different scheme that significantly narrows the possibilities of intelligent analysis of the situation.*

Subscribe

← Previous Next →

Write a comment…

Олег May 16

General NONSENCE!

Reply

Ready for more?

Subscribe