an **Article** from **SCIENTIFIC AMERICAN**

AUGUST, 1984 VOL. 251, NO. 2

The Mechanical Manipulation of Randomly Oriented Parts

It is one of the main obstacles to the broader application of robots in industry. A computer system can now "see" an object at the top of a bin of mixed parts and direct a mechanical arm to pick it up

by Berthold K. P. Horn and Katsushi Ikeuchi

Consider the fine coordination between the eye and the hand of a young child who picks a cookie out of a jar. Although the cookies are roughly uniform in size and shape, the pile of cookies at the top of the jar is a jumble of visual cues, a rugged topography from which the child must extract enough information to determine what part of the visual or tactile field can be ascribed to the single, target cookie. As the child learns to take a cookie without crushing or breaking the ones around it, the child comes to realize that not every orientation of the hand can be successful. For example, seizing the edge of the cookie between thumb and forefinger works only if the center of the cookie is on or near the line connecting the opposing points of pressure. A much more reliable strategy is to determine the attitude, or orientation, of the cookie visually and then turn the hand to one of the positions best suited for picking it up. Finally, having grasped the cookie in one attitude or another, the child must transform the spatial coordinates of the cookie that pertain to the hand into the coordinates that pertain to the mouth.

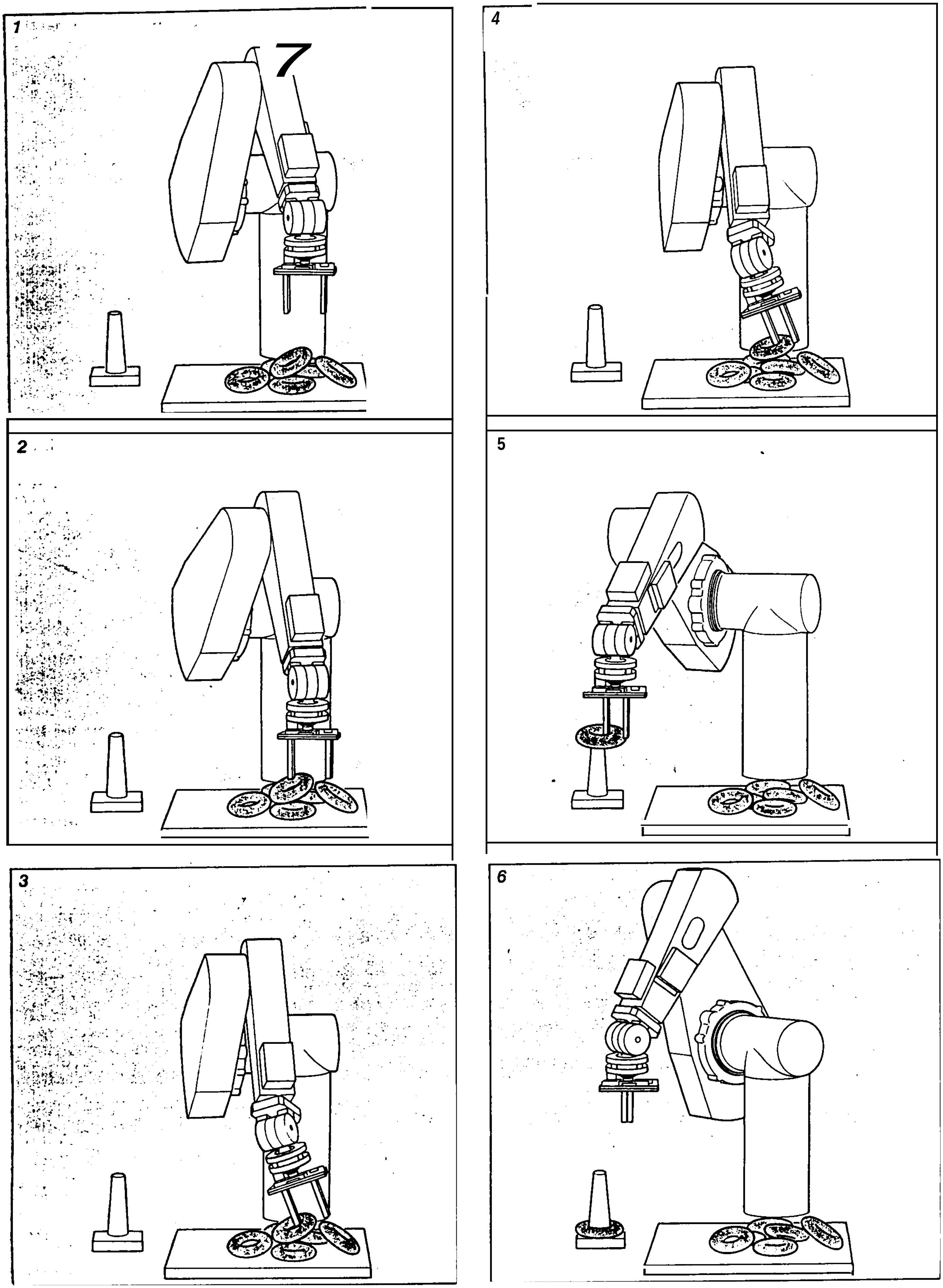
Until recently such a complex set of coordinated actions was beyond the capability of mechanization that seeks to replicate some of the functions of factory workers. The robot now working in the factory is fundamentally a playback machine for motions in space. To carry out a task the robot must first be "trained" by a person already skilled in the task. The "arm" of the robot is guided through a series of motions, and the sequence of robot configurations needed to follow the trainer is recorded on a tape or other memory device. When the tape is played back, it directs the robot to execute the same sequence of motions. The ability of the robot to record spatial motion has been exploited by choreographers to make a permanent record of dance movements, but without notable success. Nevertheless, the playback robot has found a niche in the factory because many industrial tasks are so highly repetitive that they can be done as a sequence of fixed motions. Mechanical manipulators have therefore been applied to spot welding, machine loading, painting, deburring, seam welding, sealing and other tasks that are boring or hazardous.

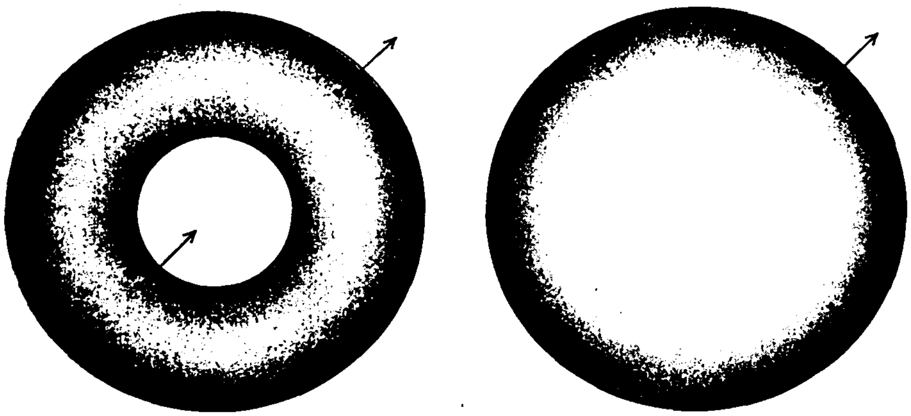
There is much factory work that cannot readily be adapted to a fixed routine of movement. In manual assembly, for example, it is common to have parts stored in trays or bins surrounding the work station. There the blind playback robot is virtually useless because it can tolerate very little uncertainty in the position of a part it must handle. An obvious solution to the problem is to avoid jumbling the parts together in the first place, or in other words to maintain a controlled orientation from the time they are made. There is a trend among manufacturers in favor of this solution: parts can be organized on carrier, or attached to pallets on which they can be mechanically manipulated without the need for sensing. Nevertheless, the solution has its costs. The carriers or pallets must be designed and manufactured, often to close tolerances. Moreover, the pallets are usually heavy, they take up a large amount of space and they often have to be redesigned when the part they carry is modified. Indeed, the design of the part itself may have to be altered for the sake of automatic feeding. Suffice it to say there are many circumstances in which the volume of production has not presented enough economic incentive for the manufacturer to depart from more traditional, manual methods.

We have now developed a computer system that can determine the position of a part with an arbitrary shape in a randomly arranged pile. The system requires only a few electronic images of the pile of parts. The images are mathematically transformed by the computer into a form that is readily compared with a mathematical model of the part stored in the computer memory. The mathematical model is rotated by the computer until it closely matches the attitude of the object to be grasped. The results are applied to direct a mechanical arm to pick up the part. Such a flexible sensing system may be able to substantially extend the range of applications of industrial robots.

We are, of course, not the first to develop a sensing system that can be employed to guide the motions of a machine. Indeed, the first stage in our procedure is common to many other kinds of machine vision: we record a digitized image of the object on the image plane of an electronic camera. The image plane is made up of a large number of pixels, or picture elements, arranged in a regular pattern. The brightness of the object, which is called the gray level, is measured for each area that corresponds to a pixel in the image plane. The brightness values are quan-

MOTION OF A ROBOT that selects an object from a small pile of similar objects is depicted in a series of drawings based on photographs made by the authors. The object is a torus, or doughnut-shaped solid, which is difficult for most computer-controlled systems to recognize and pick up. The command that directs the arm of the robot along a ray in space is based on information provided by three images made by an electronic camera. A computer program determines the identity and orientation of an object and then finds the region that corresponds to the object in the image plane. The program also selects the points at which the robot is to grip the object. When an infrared beam passing from one side of the gripper to the other is interrupted, the motion of the arm along the ray is stopped. The gripper then maneuvers into position for the pickup, closes on the object and lifts it free. The object can be set down in any orientation.


1
7
2
3
4
5
6



ORIENTATION OF A SURFACE at a point is given by the direction perpendicular to a plane that is tangent to the surface at the point *(left)*. It can be represented by the coordinates of a point with the same orientation on the surface of a unit sphere called the Gaussian sphere, after Carl Friedrich Gauss *(right)*. Orientation can be defined for any point not on a crease or at a vertex of the surface. On the torus more than one point can have the same orientation.

tized, or rounded off, to one of as many as 256 gray levels.

In special cases it is sufficient to calculate certain properties of the object directly from the quantized image on the array. For example, in certain situations it is possible to distinguish points in the image that correspond to the object of interest from points that do not. Such a segmentation into object and background is usually based on differences in brightness. The resulting image is called a binary image because each pixel represents one of two states of the object: its presence or its absence. The binary image of an object is conceptually much like the pictures that are formed from an array of lights on a theatrical marquee or a stadium scoreboard.

Binary-image processing can be done with high-speed equipment of moderate cost. Unfortunately the binary image is often too crude a representation for it to serve as a guide to automatic manipulation. If the shape of the binary image is to conform even roughly with the silhouette of the real object, the contrast in brightness between object and background must be quite strong. If there is more than one object within the field of view. they must not overlap or touch; if the objects are not separate on the image plane of the camera, the silhouettes can change in unpredictable ways and the outline of the binary image may have little to do with the actual shape of a single object. Furthermore, unless the object has some rotational symmetry, the silhouette of the object can change in a complicated way when it is rotated in any plane except one that is parallel to the image plane. The information carried by the binary image of an object in an arbitrary configuration is in general too variable to be matched reliably with the representation of the object stored in the memory of a computer.

There has been substantial progress in machine vision since the first binary-image processors were demonstrated in the laboratory about 15 years ago. Nevertheless, the same strategic question about the design of such a system must still be faced: How can a symbolic description of the three-dimensional world be recovered from the quantized, gray-scale image recorded by an electronic camera? The form and detail required in such a description depend on its application. For picking randomly arranged objects out of a bin the description need give only the identity, position and attitude of the objects in space.

It is often thought that the identity, position and attitude of a part can readily be derived if the three-dimensional topography of the top of the bin of parts is known. It turns out the derivation is not straightforward, but the topography of the parts is still a first step in determining the description. The best-known cue for recovering three-dimensional topography from two-dimensional images is the depth perception afforded by stereoscopic vision. We can see in depth partly because we have two eyes that form images from slightly different viewpoints. A number of machine-vision systems attempt to exploit stereoscopic vision, but they are slow, complex and expensive, and they can deal only with certain kinds of images.

For practical applications machine vision does not have to emulate the admirable capabilities of biological vision. We have chosen instead to adopt a method invented at the Massachusetts Institute of Technology by Robert J. Woodham, which is called photometric stereo. The method determines the surface orientation of each small patch on the surface of an object but does not give the absolute distance to a point on the object. It turns out that for segmenting, identifying and finding the attitude of an object in space only local surface orientation is necessary.

The orientation of a surface at any point (except a point on a crease or at a corner) is defined by the direction of a line perpendicular to the plane tangent to the surface at that point. Every possible orientation of a surface corresponds to the orientation of some point on a sphere, and every separate point on the sphere has a different orientation. The line that gives the orientation of a point on an arbitrary surface is therefore parallel to the line that gives the orientation of some point on the sphere. It also follows that any spatial orientation can be specified by giving two coordinates, say the latitude and the longitude, of a point on a unit sphere called the Gaussian sphere, after the mathematician Carl Friedrich Gauss.

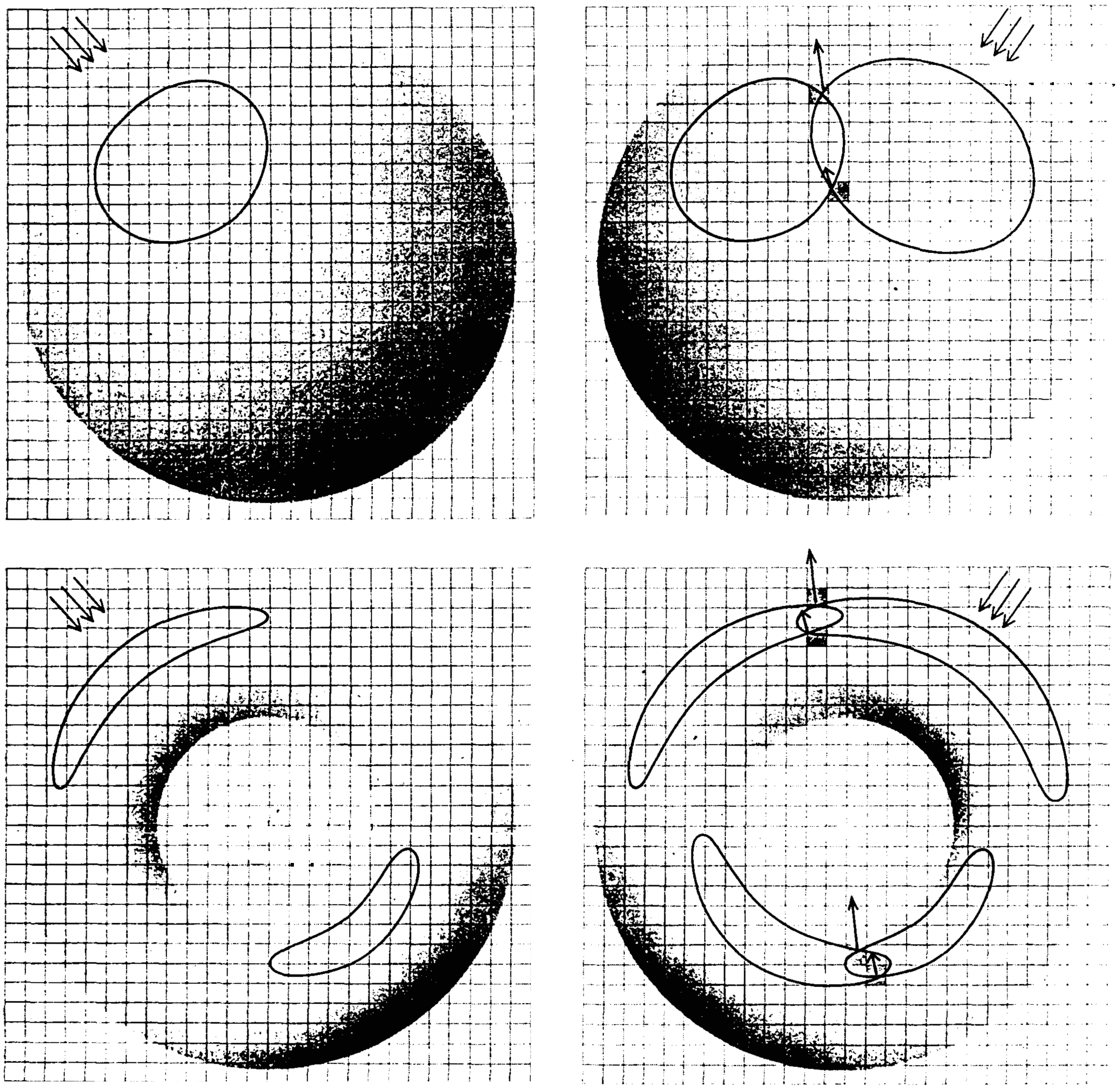
Suppose a Gaussian sphere is illuminated by a distant source of light. Furthermore, suppose the material on the surface of the sphere reflects all incident light and appears equally bright from all viewing directions. Since the light source is far away, the distance between the light and a point on the sphere does not vary significantly with the position of the point. The amount of light captured and reflected by a small patch on the surface of the sphere therefore depends only on the apparent ares of the patch as seen from the light source. The apparent area depends in turn on the inclination of the patch with respect to the light.

Since the brightness of the spherical surface is assumed not to change with viewing direction, the brightest part of the surface for any viewer is the small patch around the point where the surface orientation matches the direction of the incident light, or in other words the point for which the source of light is directly overhead. The brightness of the surface decreases with the distance, measured on the spherical surface, from the brightest point. Patches of equal brightness form concentric rings around the brightest point because they are all inclined at the same angle to the light.

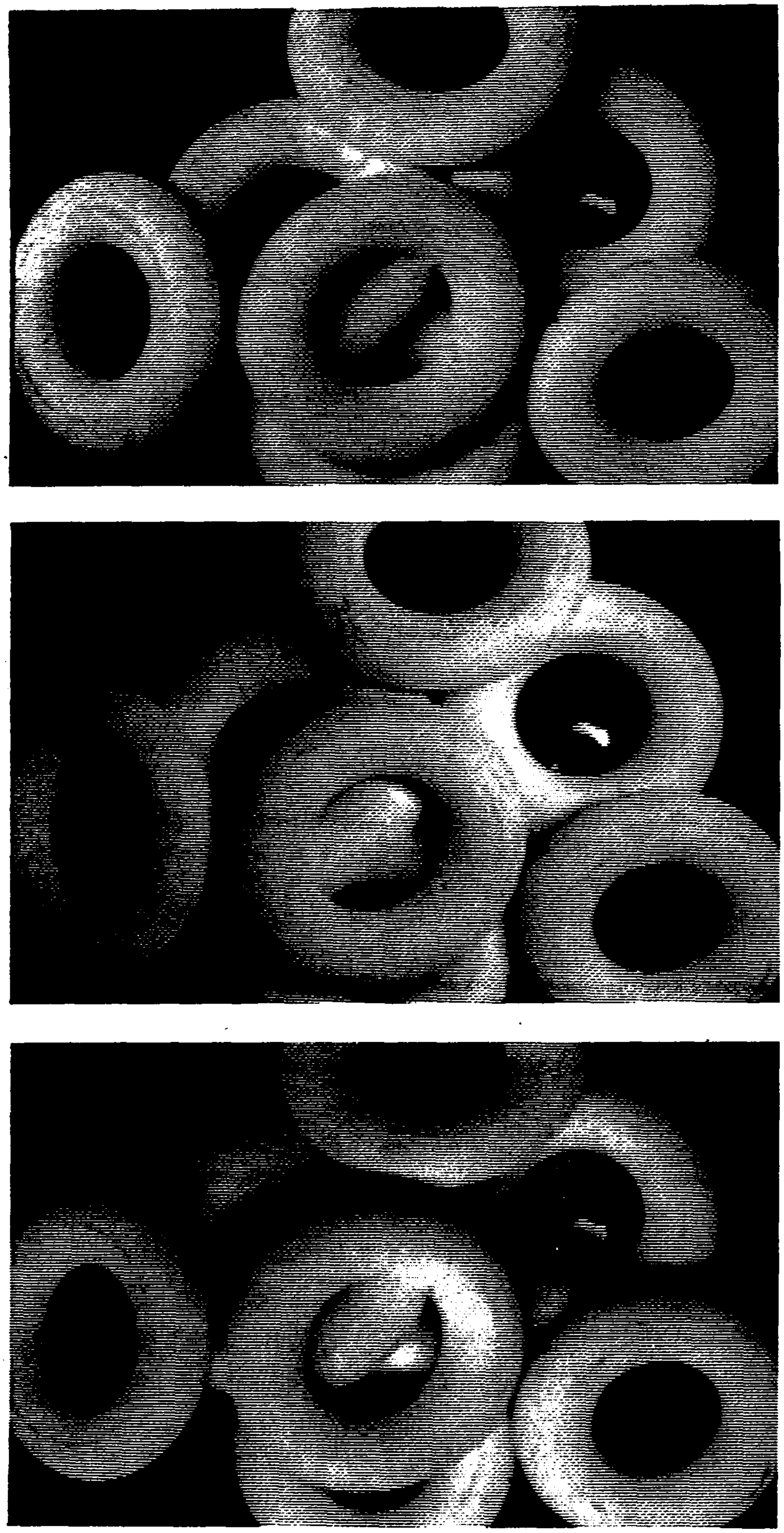
If the first light is turned off and a second distant light is turned on, the visible patches on the spherical surface are grouped into thin rings of equal brightness centered on a second point. Any small patch on the surface of the sphere that can be illuminated by both lights is thereby assigned two brightness values, one value for each light source. The first value limits the possible positions of the patch to a circle centered on the point directly under the first light; the second value assigns the patch to a second circle with a different center. The circles can intersect at no more than two points on the surface of the sphere. Hence for a given pair of gray levels there can be at most two corresponding points on the sphere, or in other words two orientations.

Suppose a table of values is constructed in which the brightness measurements made on the sphere are matched with the orientations to which they correspond. If a new object of arbitrary shape is put in place of the sphere, its surface orientations can be determined directly from the table. For each small patch on the new object a pau of brightness measurements are made, one measurement for each light source that was previously turned on to calibrate the surface orientations of the sphere. The orientations that correspond io each brightness measurement are then simply read from the table of values. The procedure is fast because the brightness measurements for all the surface patches of the new object can be obtained simultaneously from two images and because the data manipulation needed to determine the orientation of each surface patch from its brightness values is trivial. Moreover, the method can work for almost any object, no matter how complicated its surface or how strange the arrangement of the lights.

One obvious problem with the procedure is that the surface orientation of a patch is not uniquely determined. A



BRIGHTNESS MEASUREMENTS of the light reflected from any **small** patch of a surface can specify the orientation **of** the patch. A sphere is placed in the field of view of an electronic camera in order to calibrate the computer system. The orientations of the surface points of the sphere are known, and **so** each brightness measurement, **or gray** level, recorded by a picture element in the image plane **of** the **camera** can be associated with a known orientation. When the sphere is illuminated **by** one source of light, the contours of constant brightness **on** the sphere are concentric circles, one **of** which **is** shown *in color* at the upper left. A second brightness measurement with a light source from **a** different direction gives a second set **of** concentric circles (*upper right*). The two measurements specify a pair **of** gray levels for each picture element. Only two patches on the sphere, which correspond to the two picture elements where two circles intersect, can have a given pair **of** gray levels. **A** third measurement gives a unique orientation of a patch on the sphere. When the same measurements are made for an unknown object, the orientation **of** a surface patch **is** determined **from** the calibrated gray levels ***(lower left, lower right).***



REFLECTED LIGHT from a random pile of objects is shown for sources of light from three directions. The differences in shading for a given region of the surface are subtle to the eye, but they can readily be detected by electronic sensors. The photographs were made by the authors.

third light source can remove the remaining ambiguity, but the information it provides is far greater than that needed to distinguish two orientations. Instead of being content with overkill, one can exploit the three sources of light to derive additional information about surface properties. For example, if a surface reflects only a fraction of the incident light, and if that fraction, which is called the albedo of the surface, varies from point to point, each of the three brightness measurements gives rise to an equation with three variables. The variables are the two coordinates and the albedo of each point on the surface of a sphere. The system of three equations can be solved for the variables, provided the three lights and the illuminated object do not all lie in a plane.

If the brightness measured by each pixel of the camera is rounded off to one of, say, 16 values, there are 16^3, or 4,096, possible combinations of brightness values for each pixel when the brightness is measured for three sources of light. Most of the combinations, however, are not to be found in the lookup table. For example, no surface orientation of the sphere would correspond to the combination in which all three brightness values are maximum, unless all three sources of light were to impinge on the surface from the same direction. In that case, however, surface orientations could not be uniquely defined by the variations in lighting. Brightness combinations absent from the lookup table are nonetheless detected by the camera at some pixels, and such "impossible" combinations can be quite valuable in segmenting the image, or dividing it into regions that correspond to different objects.

One cause of anomalous brightness combinations is the shadowing of one object by other objects in the pile. A crude way to detect shadows is to assume that gray levels darker than a certain threshold in at least one image indicate a shadow. A second cause of anomalous brightness combinations is mutual illumination, the reflection of light from one object onto another; it is particularly common when objects of high albedo face one another. We assume that if the gray levels are brighter than the shadow thresholds, most observed combinations that are not found in the lookup table are caused by mutual illumination. The effect is generally seen near the edges of objects and along boundaries where objects tend to obscure one another; it can therefore be exploited for image segmentation. We also look for discontinuities in surface orientation and for high surface inclination, both of which tend to mark regions where one object obscures another.

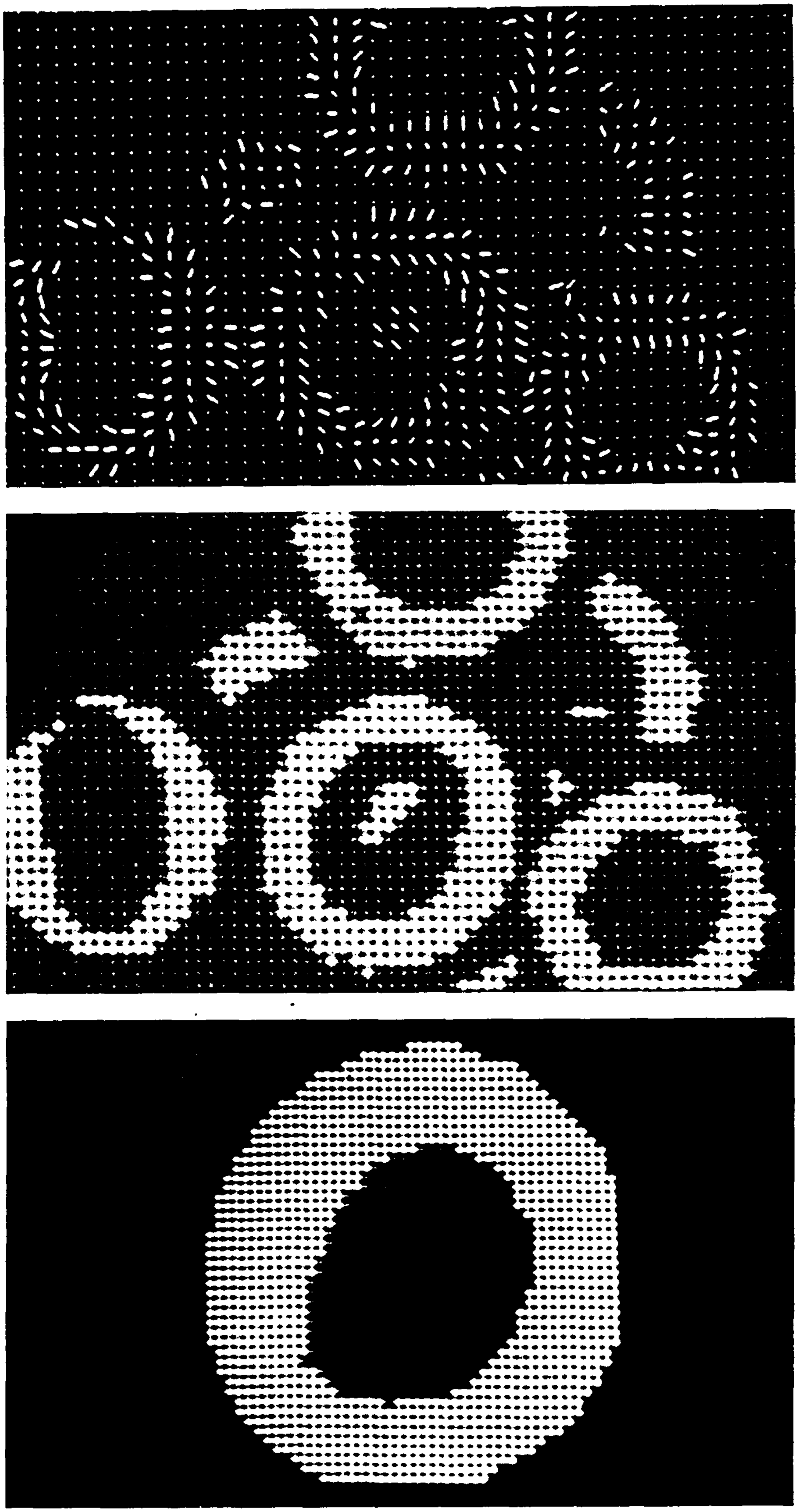
Once a connected object of interest has been tentatively identified in the field of view, our goal is to match the observed object with one of the prototype objects that is abstractly represented in the memory of the computer. When the match is made, the observed object is identified. The only data available for making the match, however, are the position and orientation of surface patches on the object in view. That information can be represented by constructing a line perpendicular to the surface of the object at each point that corresponds to the center of one of the pixels into which the image is divided.

Suppose all the perpendicular lines have the same length. The shape of the observed object can be represented by the length and direction of the perpendicular lines as seen in perspective: the lines on the surface patches that face the viewer are represented as points, and the lines on the surface patches that slope away from the viewer vary in length with the sine of the inclination of the patch. The resulting figure resembles a surface covered with the quills of a porcupine; it is called a needle diagram.

It is costly and computationally inefficient to compare the needle diagram of the observed object directly with a needle diagram of a prototype object. Oddly enough, it is much more efficient to temporarily disregard the information that gives the relative position of various surface patches and focus instead on the surface orientations alone. A mathematical representation of the surface orientations called the extended Gaussian image, or EGI, is constructed from the needle diagram. The prototype objects are stored in computer memory in a similar mathematical form.

The EGI of any object is a sphere on which are plotted the relative contributions of each orientation of the surface of the object to the area of the surface as a whole. In order to identify the object selected in the field of view, the EGI of a prototype is abstractly rotated within the computer until it matches the EGI of the observed object as closely as possible. The same procedure is repeated for each prototype stored in memory. The observed object is assumed to be the prototype that gives the best overall match; the match simultaneously gives the attitude of the EGI for the object.

To understand how the EGI of an object is constructed, remember that any point on the surface of the object can be associated with a point having the same orientation on the Gaussian sphere. Similarly, a patch on the surface of the object can be associated with a patch on the surface of the Gaussian sphere by matching each point on the object with its corresponding point on the sphere. For example, wherever the surface of the object is relatively level like the flat side of an egg, the corresponding patch on the Gaussian sphere



NEEDLE DIAGRAM (*top*) represents the orientation of surface patches on the random pile of objects shown in the photographs on the opposite page. The orientation corresponding to each picture element in the camera is given by the direction of a needle, or line segment of constant length. The needles are shown as if they were attached to the surface at right angles like the quills of a porcupine and viewed from the camera. The computer divides the image into connected segments (*middle*), and one of the segments is isolated for further processing (*bottom*).

encompasses a relatively small fraction of the surface of the sphere. On the other hand, wherever the surface of the object curves relatively sharply like the end of an egg, the corresponding patch on the Gaussian sphere is relatively large.

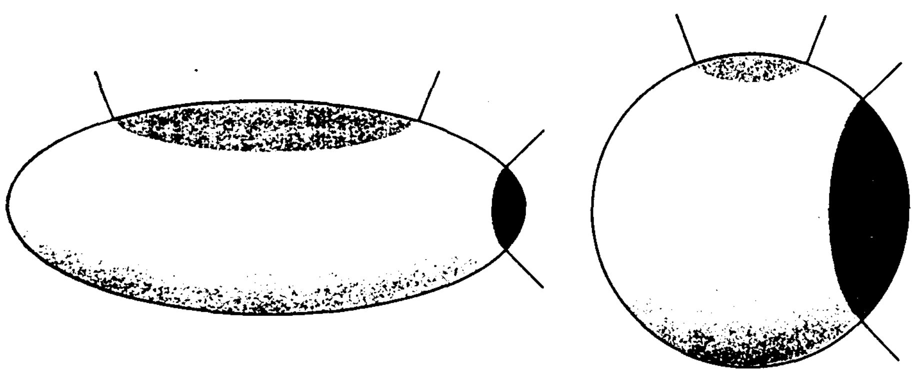
Imagine now that the egg is covered by a material of uniform density. To construct the EGI of the egg, the material from every patch on the surface of the egg is compressed or spread out in such a way that it fits exactly into the corresponding patch on the Gaussian sphere. The material on the flat region of the egg is compressed like a lump of clay in order to fit into a relatively small region on the Gaussian sphere. The material on the end of the egg must be spread out so that it fills a relatively large region on the Gaussian sphere. As the patches on the object of interest become progressively smaller the density of the material on the Gaussian sphere can vary continuously over its surface.

The visible hemisphere of the EGI, which corresponds to the visible surface of the observed object, can be numerically approximated from the needle diagram. The surface of the Gaussian sphere is tessellated, or divided into cells, and each cell corresponds to some small range of possible orientations. Every pixel of the needle diagram whose orientation falls within the range of orientations corresponding to one of the cells is assigned to that cell.

In determining the mass of material that is to be assigned to the cell, one must remember that the surface area of the observed object projected onto a pixel depends on the inclination of the surface with respect to the viewer. A surface that is steeply inclined away from the viewer is foreshortened and appears smaller than it would if it were viewed head on; one can correct for the effect because the angle of inclination is known from the needle diagram. The mass on each cell on the Gaussian sphere is then equal to the total mass of the parts of the observed surface that are visible in the pixels assigned to the cell. Since the material that covers the observed surface has a uniform density, its mass over any patch of the surface is directly proportional to the area of the patch. Hence the mass on each cell is also equal to the area of the parts of the observed surface to which the cell corresponds. The mass distributed over all the cells is equal to the total area of the observed surface. The tessellated Gaussian sphere is a quantized, or discrete, approximation of the EGI. It is called the orientation histogram.

The distribution of mass on the orientation histogram and, for that matter, the distribution on the EGI lead to a number of mathematical results that are useful in matching the observed object with a prototype. It is straightforward to calculate the center of mass for any visible hemisphere of the orientation histogram. (Note that this quantity has nothing to do with the center of mass of the visible part of the real object.) Since the orientation histogram of each prototype object in the memory of the computer is known over the entire Gaussian sphere, the center of mass can be calculated for any visible hemisphere and stored in memory. We generally do the calculation for each hemisphere that is visible when one of the cells in the tessellation is viewed head on.

Consider the plane that divides the Gaussian sphere into a visible hemisphere and an invisible one. The center of mass of the visible hemisphere lies at some distance D above the plane in the direction of the viewer. The product of the mass of the visible hemisphere and D is called the first moment of the mass about the dividing plane. Since the mass of the visible hemisphere is equal to the surface area of the object to which the hemisphere corresponds, the first moment is equal to the area of the visible surface of the object times D.



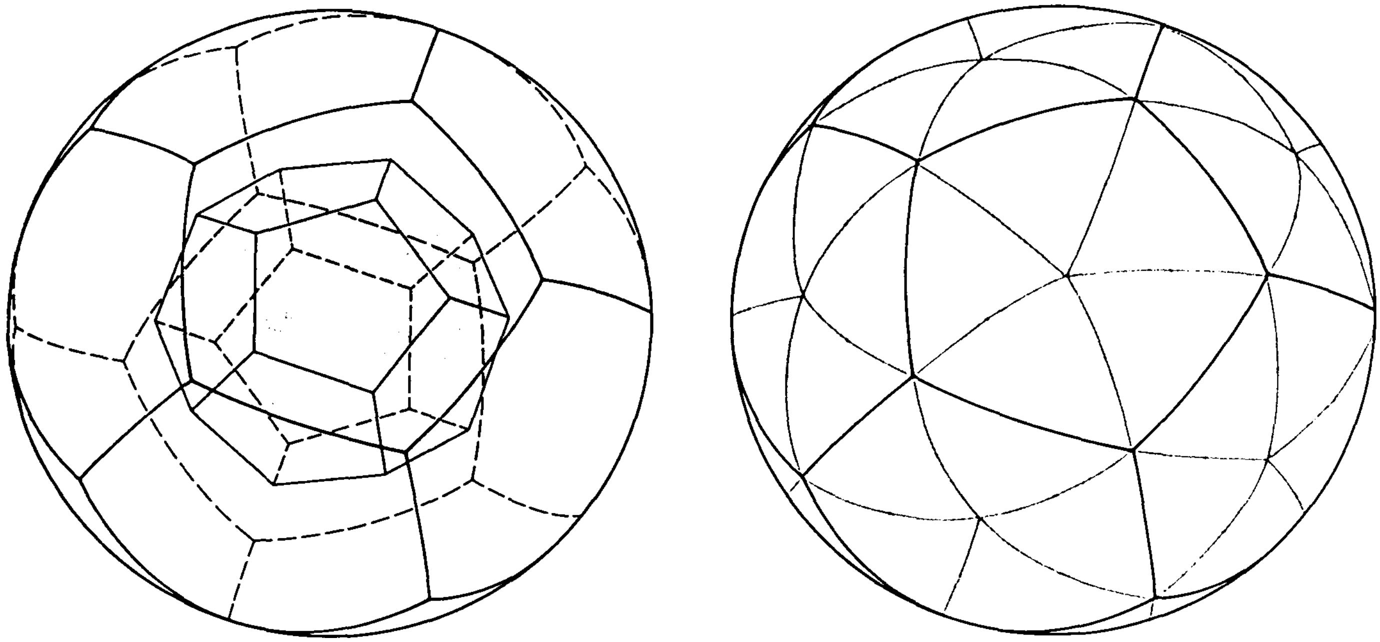
PATCH ON THE SURFACE of an object can be associated with a patch on the surface of the Gaussian sphere Every point in the patch on the object is matched with the point on thesphere that has the same orientation. The patch on the sphere is a large proportion of the total surface area when the corresponding patch on the object is strongly curved *(gray);* the patch on the sphere is small when the corresponding patch on the object is relatively flat *(color).*

There is another way to represent the first moment of the mass of the visible hemisphere. Consider the mass assigned to any cell in the tessellation of the Gaussian sphere. The individual cell's first moment about the dividing plane is the product of its mass and its distance from the plane. Since the cell lies on the surface of the unit sphere, its distance from the plane is readily calculated. If the cell directly faces the viewer, its first moment is equal to its mass. If the cell is inclined from the viewer, its first moment is reduced by a factor that depends on its inclination; the factor is equal to the cosine of the angle between the orientation of the cell and the viewing direction. Remember that the mass of the cell is equal to the area of the parts of the surface to which it corresponds. When those parts of the surface are viewed, their actual area is also reduced by the cosine of the angle between their orientation and the viewing direction. It follows that the first moment of the cell about the dividing plane is equal to the cross-sectional, or apparent, area of the surface to which it corresponds.

The first moment of the mass of the entire hemisphere is equal to the sum of the first moments of all the visible cells; in other words, the first moment is equal to the cross-sectional area of the visible surface of the object. As we have shown, however, the first moment is also equal to the product of the actual area of the surface and D. The result is that D, which is a number defined for the orientation histogram, is equal to the ratio of the cross-sectional area of the observed object to its actual surface area. The ratio can be calculated directly from the needle diagram of the observed object.

The position of the center of mass for any hemisphere of the orientation histogram determines the value of D. Hence the observed ratio of the cross-sectional area of an object to its actual surface area can be compared with the values of D associated with various attitudes of the prototype objects. Although the value of D does not unambiguously give the attitude of the prototype matching that of the observed object, it does save computation. Any hemisphere of the orientation histogram for which the center of mass is not at least approximately in the right position need not be scrutinized further.

In general the attitude of an object can be specified by giving the direction of some axis that passes through it and the amount of rotation of the object about that axis. Since the number of directions for the axis and the number of rotations for the object are both infinite, one cannot compare the EGI of the observed object with all possible attitudes of the EGI of a prototype. Our matching procedure depends on sampling a finite number of the attitudes of the EGI that



TESSELLATION OF THE GAUSSIAN SPHERE can be done by projecting a regular dodecahedron onto the surface of the sphere (*left*). The dodecahedral tessellation is ideal for matching a representation of the unknown object with a prototype stored in the memory of the computer, except that the 12 pentagonal cells in the tessellation are too large. Smaller cells can then be generated *(right)* by dividing each pentagon into five triangles *(red);* the subdivision could be continued indefinitely by dividing each large triangle into smaller ones *(blue)*.

can be derived from the tessellation of the Gaussian sphere.

The tessellation is constructed to meet several independent criteria. In order to distribute the surface orientations represented by the needle diagram in an unbiased way, the cells of the tessellation should ideally have the same area and the same shape. If the cells are relatively "rounded" polygons, such as pentagons or hexagons, instead of sharply angled ones, such as triangles, the range of orientations assigned to each cell can be minimized. Moreover, it is desirable that when one cell is rotated into the former position of another, the rest of the cells on the Gaussian sphere are permuted from their initial positions. In this way the rotation of the sphere can be represented in the computer simply by permuting the masses associated with the cells. All these criteria can be met by projecting the regular dodecahedron, which is a polyhedron with 12 pentagonal faces, onto the sphere.

Unfortunately, with only 12 pentagonal cells on the Gaussian sphere the tessellation is too coarse for the comparison with an unknown object. A finer tessellation can be constructed by dividing each of the 12 cells into five triangles; each triangle can in turn be subdivided into four smaller triangles. The resulting tessellation has 240 cells, and so about 120 cells cover any hemisphere of the sphere that is to be matched with the orientation histogram of the object.

There are many technical refinements to the matching process that we shall not discuss here. In general we align the orientation histogram of the observed object with the orientation histograms of the various prototypes. One way to measure the success of the match is to find the square of the difference in mass for each corresponding cell. The best match is the one for which the sum of the squares for all pairs of corresponding cells is a minimum. In practice we find that about 720 trials are needed to find the attitude of the EGI of an unknown object. The accuracy is between five and 10 degrees of arc.

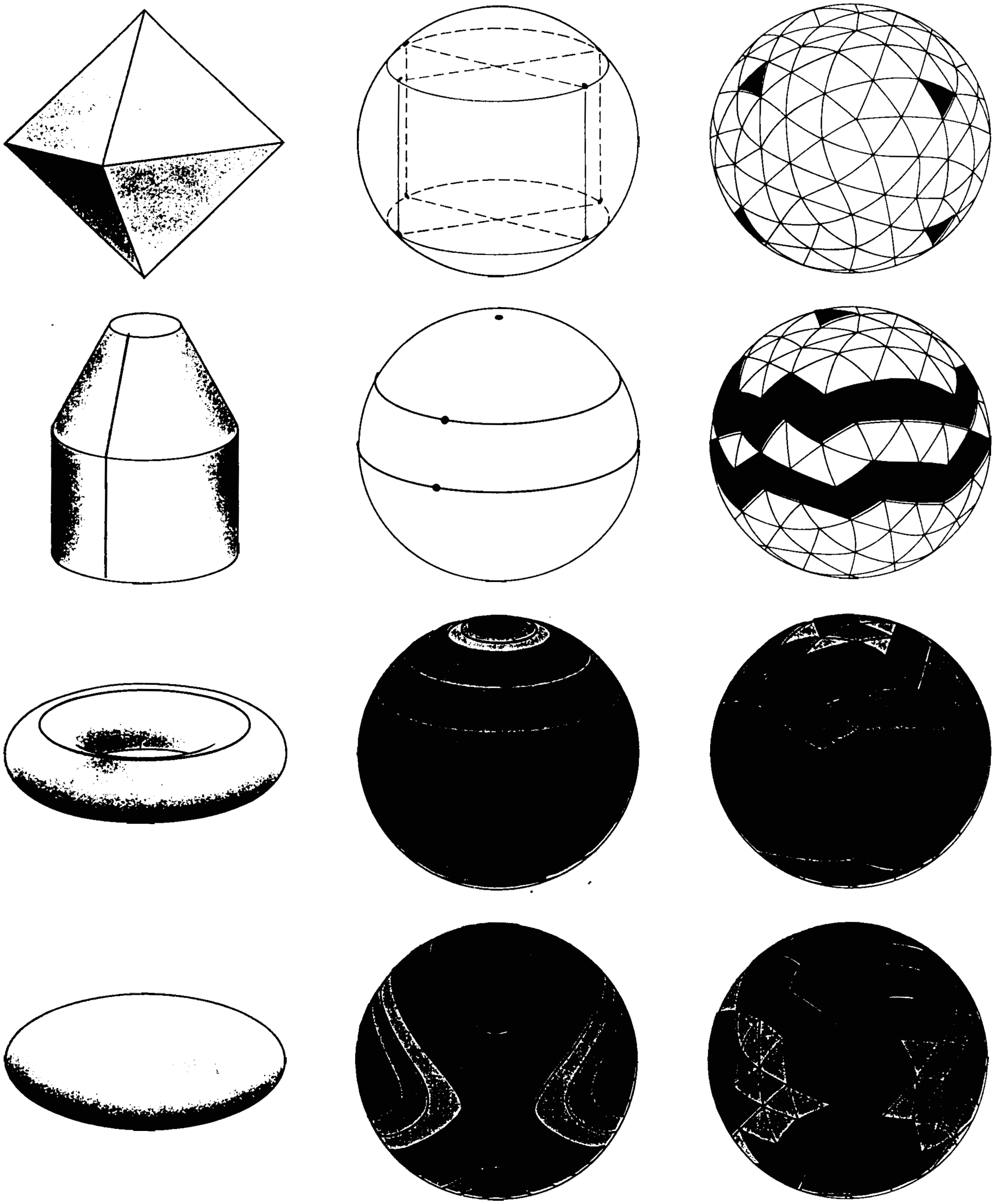
Once the identity of the-observed object and the attitude of its EGI are known, the attitude of the object in space is also determined; the control of the robot arm is then relatively straightforward. The computer must determine which points on the surface of the object are most suitable for grasping. In part the decision is dictated by the shape of the object, but it is also desirable to choose points for grasping that are high on the object in order not to interfere with neighboring objects in the bin.

It is worth noting that the calculations we have described do not give the position of the object. Position can be roughly defined by the point in the center of the object region in the image plane of the camera. We find the position more accurately by calculating a needle diagram from the known orientation of the prototype. The calculated needle diagram can then be matched with the observed needle diagram.

The position of the object in the camera image defines a ray, or direction, from the camera. In order to command the arm of the robot to move along the ray, it is necessary to transform the spatial coordinates measured with respect to the camera into spatial coordinates measured with respect to the arm. We establish a general rule for the transformation by calibrating a few fixed points. The gripper of the robot moves a surveyor's mark, or two-by-two checkerboard, to several fixed points in two planes parallel to the image plane of the camera. For each point the spatial coordinates of the arm, which are determined by the sensors of the robot, are matched with the spatial coordinates measured by the camera. After the calibration is made each point in the image can be associated with a point in each of the two planes. The two points define a ray in the spatial coordinates of the arm.

The arm of the robot begins moving along the ray from some convenient height above the objects in the bin. Since photometric stereo does not give information about the absolute distance to the object along the ray, we installed a sensor on the gripper of the robot. The sensor is actuated by a modulated infrared beam of light that propagates from one side of the gripper to the other. When the beam is interrupted, the arm is stopped. The hand is then reoriented if necessary to match the attitude of the object, the gripper is closed and the object is lifted free.

Our system takes about a minute to switch the lights on and off, record the images, match the observed data with the prototypes and send the proper commands to the manipulator. There is no inherent reason the cycle time could not be much shorter. The calculations are

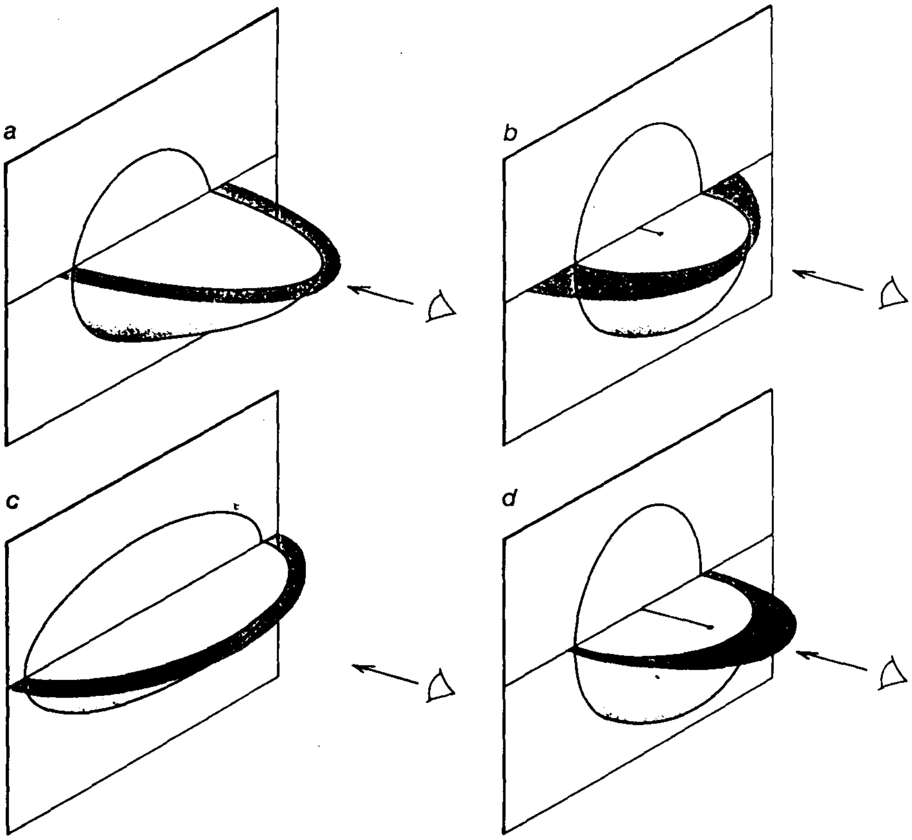


EXTENDED GAUSSIAN IMAGE (EGI) of an object can be pictured as a distribution of material over the surface of the Gaussian sphere The material is initially spread evenly over the surface of the object, Each patch of material on the surface is then moved onto the sphere and compressed or spread out like clay to fit into the corresponding patch on the sphere. The EGI is shown in the middle column of the illustration for various objects. The regions of highest density are shown in red, and regions of lower density are shown in orange, yellow, green, blue and purple. For example, all the points on a face of a polyhedron have the same orientation, and so all the material from that face is concentrated at one point on the Gaussian sphere. The surfaces of a cone and a cylinder are each mapped into a circle on the Gaussian sphere; a line on the cone and a line on the cylinder parallel to the axis of rotation are each mapped into a point. The computer "perceives" the objects as they are shown in the column at the right; there the EGI is quantized on a tessellated Gaussian sphere simple, and special-purpose hardware could be built to speed the matching. Our aim was solely to demonstrate the feasibility of our approach to the problem, not to show how fast it can work. Most of the time the robot picks up one of the objects from the pile on its first approach. Occasionally the fingers of the gripper bump into another object before they reach the target; the arm then backs out of the field of view and the process is started again from scratch.

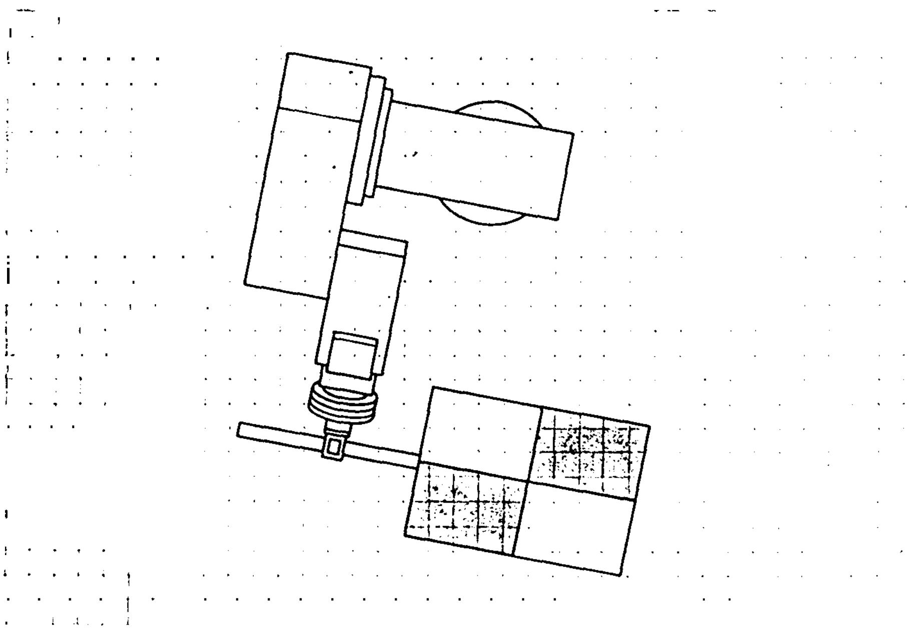
There are numerous ways our system could be modified, and many improvements will undoubtedly be made before it is adopted by industry. We have recently added a system designed by H. Keith Nishihara of M.I.T. that simulates stereo vision for determining the topography of a surface. The advantage of the additional system over photometric stereo used alone is that it gives some information about absolute depth. In another experiment we substituted laser range sensors for the lights and camera employed in photometric stereo. Both methods enable the robot to avoid moving along rays that intersect objects in the foreground of the target.

For many industrial applications, of course, the robot is too slow and its versatility is not needed. In such cases "hard" automation is the rule: special-purpose machinery is designed to orient a part. For example, small parts such as screws and other objects with a cylindrical geometry can be dumped into a vibratory bowl that can reject all configurations of the objects except the configuration needed. Large or heavy parts, however, as well as parts with complex shapes are not well suited to vibrational sorting. Moreover, a huge production volume may be necessary to justify the cost of such machinery..

We believe the system we have described is flexible and robust enough to be adapted to industrial tasks. It can reliably recognize objects and determine their attitude in space. The cameras and other necessary hardware are relatively inexpensive because only a few thousand pixels are scanned for each field of view. The computer program is largely devoted to pattern matching, and the patterns for the prototypes can be derived directly from data already present in programs for computer-aided design. Photometric stereo can readily be applied in the factory because it requires no special lighting conditions; extended sources of light can be placed in almost arbitrary positions. provided the gray-level calibrations are made after the lights are fixed in place. Moreover, the method is not limited to materials with particular light-reflecting properties. It remains to be seen what additional improvements will be made before machine vision is extensively applied to the mechanization of work.



CENTER OF MASS of the material *(color)* that covers a hemisphere of the extended Gaussian image depends on the area of the cross section of the object to which the EGI corresponds. For example, when an egg is viewed end on (*a*), the center of mass of the corresponding hemisphere of its EGI is relatively close to the plane that divides the Gaussian sphere in half *(6)*. When the relatively flat surface of the egg is viewed (*c*), the center of mass of the corresponding hemisphere of its EGI is farther from the dividing plane (d). It can be proved that the distance from the center of mass to the dividing plane is equal to the ratio of the area of the cross section to the area of the visible surface of the object. Because that ratio is known, many orientations of the EGI of the prototype can be eliminated from further comparison with the unknown object.



COORDINATION BETWEEN EYE AND HAND of the robot is arranged by calibrating the spatial coordinate system of the robot arm with the spatial coordinate system of the electronic camera. A surveyor's mark is moved to a series of fixed points io the coordinate system of the arm, and the images of the points are given a second set of coordinates measured with respect to the camera. The robot and the surveyor's mark are shown as they appear to the camera through a coordinate grid. The computer calculates a transformation whereby the coordinates of a point in one system can be determined from the coordinates of the point in the other.