Visual Perception Survey Project
-- Perceiving Motion and Events

Chienchih Chen

Yutian Chen

1. Image Motion

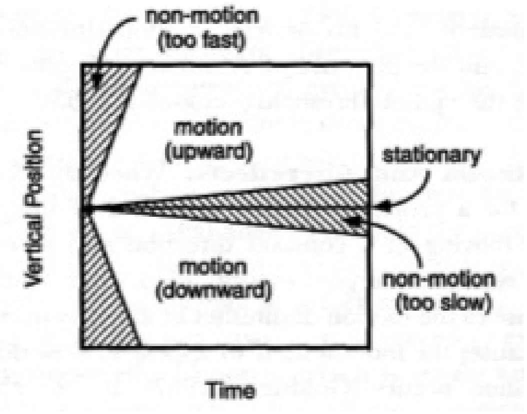
When a moving object is viewed by an observer with eyes, head, and body fixed, its projected image moves across the retina. In this instance it is natural to equate object motion with image motion. But much of the time, our eyes, head, and body are in motion, and this fact invalidates any attempt to equate object motion directly with image motion.

Image motion depends on both eye movements and object motion. As the eye moves in one direction, the image sweeps across the visual field in the opposite direction. But, there is image motion when there is no object motion at all. The opposite can also occur, as when a moving object is tracked by the eye: It's image on the retina will actually be stationary even though the object itself is in motion. For this reason, it is important to distinguish clearly between image motion and object always depends on eye, head, and body movements as well as image motion.

Continuous Motion

An environmental object is considered to be moving if its position changes over time. Because real objects do not disappear from one place and reappear in another, these changes in object position are continuous events.

Not all such events produce visual experiences of motion, however; some occur too rapidly, and others too slowly. In terms of space-time diagrams, the boundary conditions for continuous motion perception are defined by the slope or orientation of space-time contours with respect to the temporal axis. A moving object is perceived as stationary if the slope of its space-time contour is too shallow. If the slope of its space-time contour is too steep, it is perceived not as moving, but as spatially smeared across its trajectory.



Adaptation and Aftereffects

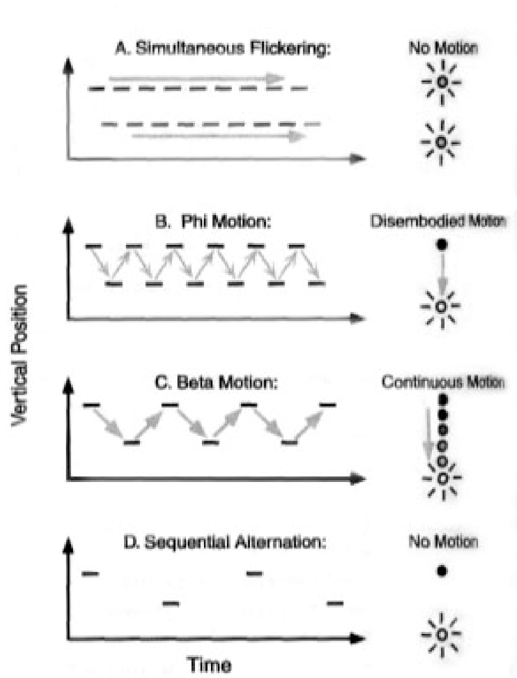
When an observer stares for a prolonged period at a field of image elements moving at a constant direction and speed, the visual system undergoes motion adaptation; that is, its response to the motion diminishes in intensity over time. It is related to motion aftereffects: by viewing of constant motion for a prolonged period, after the motion has stopped, subsequent motion perception is significantly altered.

Apparent Motion

Modern technology provides a host of situations in which realistic motion perception arises from the rapid presentation of completely static images. This technique rely on the fact that the human visual system can be fooled into perceiving continuous motion from a sequence of "snapshots" or "frames" presented at the proper rate.

As the rate decreases from very fast to very slow, the perception of a display alternately presenting two lights across a fixed distance changes through the following sequence of different experiences:

1. simultaneous flickering
2. Phi motion
3. Beta motion
4. Sequential alternation



Motion Picture Technology

Apparent motion lies at the heart of all motion picture technology. Movies, for example, are projected from a strip of film that contains a sequence of many static photographic images, each of which is slightly different because of the time at which it was exposed. The frames on the film are flashed by a projector at a rate of 24 frames per second, just lies within the range that produces beta motion, the most convincing perception of smooth continuous movement.

The Correspondence Problem of Apparent Motion

There is no problem in simple, one-point apparent motion, because each frame contains only a single object, making the correspondence unambiguous. But the correspondence problem arises whenever two or more objects are present in apparent motion display. The most potent factor in solving the correspondence problem is proximity: the distance between potentially corresponding elements, the closest elements will be perceived as corresponding.

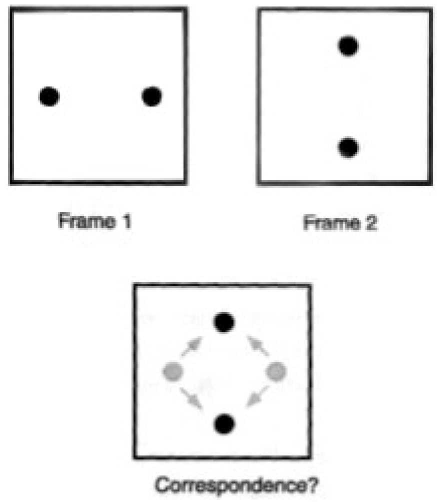
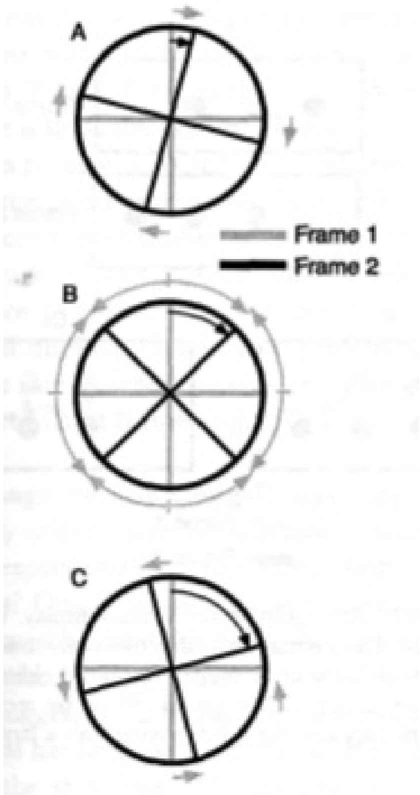
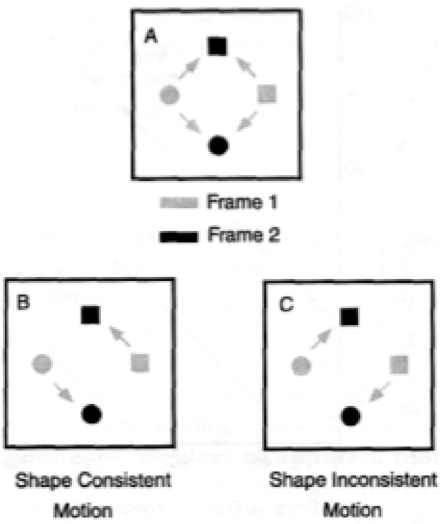
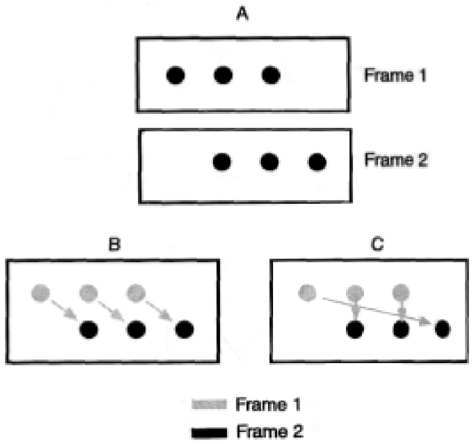


Figure A shows a wheel turns slowly. In the second frame(black lines), the spokes have turned only a short distance clockwise, as indicated by the black arrow. If the visual system solves the correspondence problem across frames in terms of pairing the closest spokes, it should produce clockwise apparent motion. But there comes a rotation rate at which a dramatic change takes place. In Figure B, the distance between the spokes in the first frame and those in the second is the same when measured in the clockwise and counterclockwise directions. The correspondence based on distance is therefore maximally ambiguous, and the motion can be seen in either direction. As the wagon picks up still more speed, the spokes have traveled even farther from one frame to the next, so the closest spokes are clearly the ones in the counterclockwise direction. Solving the correspondence problem based on distance therefore predicts that the wheel will appear to turn backward.



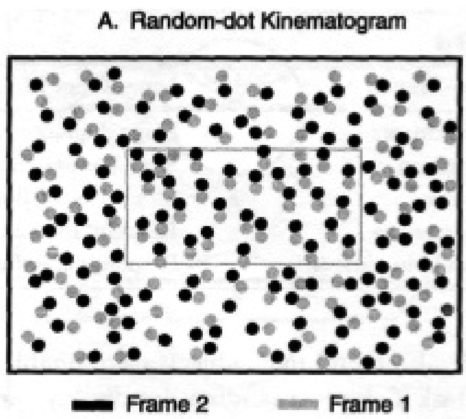
Logically, one would expect that factors like sizes, shapes, orientations, or colors might affect the solution to the correspondence problem. Surprisingly, several experiments indicated that these factors did not matter. This means the visual system solved the correspondence problem almost exclusively according to distance criteria.

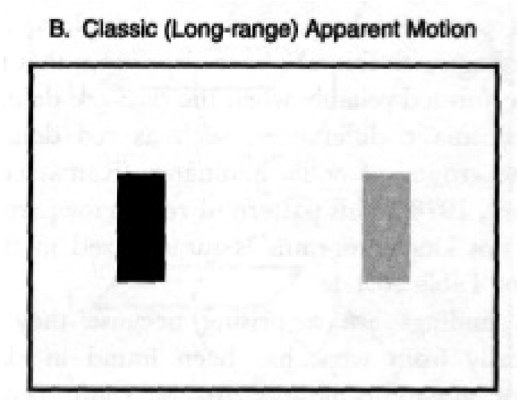




Short-Range versus Long-Range Apparent Motion

In 1974 English psychophysicist Oliver Braddick suggested a dual process theory of motion perception that has been very influential. He called the two processes the short-range and long-range motion systems. Braddick generated a random dot kinematograms, which are motion stimuli composed of many randomly positioned elements. The first frame consisted of a completely random array of thousands of tiny black dots on a white background as shown in the gray dots of Figure A. Generate the second display from the first by (1) displacing all of the dots within a central rectangular area by some particular distance and direction and (2) replacing the "background" dots by a completely new set of random dots, uncorrelated with the first frame.



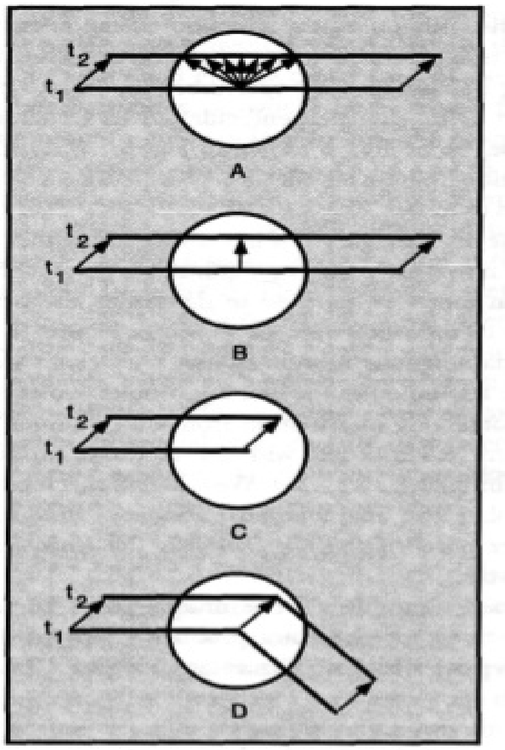


If the visual system is sensitive to the systematic displacement of the dots within the rectangular area, the rectangle should be perceived to move coherently as a unit against a background of randomly moving dots, and the orientation of the rectangle should be easy to discriminate. If the correspondence cannot be detected, however, subjects should be unable to perceive the rectangle and therefore fail to determine its orientation. Braddick found that for subjects to perform the orientation discrimination task accurately, the displacements had to be small and the alternation rate had to be rapid. He also found that subjects were unable to perform the task if the two random dot display were presented to different eyes or if a bright uniform masking field was presented between the two displays.

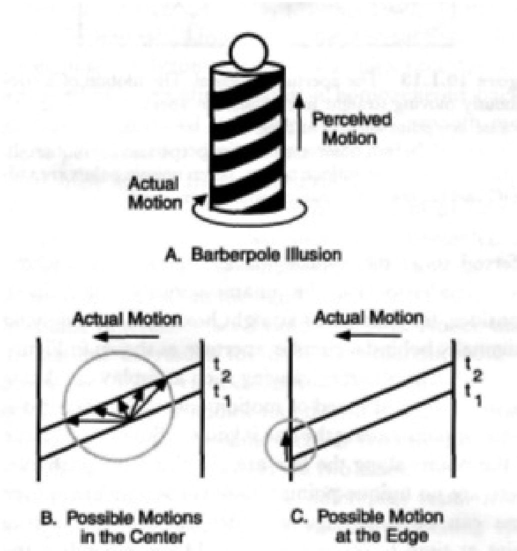
These findings are different from classical studies of apparent motion. When two large-scale objects are presented in alternation against a homogeneous background, as illustrated in Figure B, smooth motion is perceived for displacements of many degrees of visual angle and for frame durations up to 300 ms. This classical form of apparent motion is also perceived when the two displays are presented to different eyes, when a light masking field intervenes between them, and when the figures are defined by purely chromatic differences.

Braddick suggested that the differences result from the operation of two different motion processing systems. What he called the short-range motion system is responsible for performance on the random dot kinematograms, and is thought to occur fairly early in visual processing, before information from the two eyes has been integrated and before shape and color have been extensively analyzed. In contrast, the long-range motion system is thought to be responsible for classical phenomena of apparent motion with large-scale individual figures, and is thought to occur much later in processing than the short-range system.

The Aperture Problem



The aperture problem refers to local ambiguity in the direction and speed of motion whenever the portion of the image that is visible within a restricted region lacks unique points whose correspondences can be unambiguously determined. A observer viewing such a display can know the direction and speed of motion only if the correspondence of points along the line is known. But because none of the points along the line are visually distinguishable, there are no unique points whose correspondences over time can be unambiguously determined. If one or both ends of a line are visible within the apperture, their correspondence is unambiguous and can therefore be used to determine the perceived motion of the whole line. This kind of unambiguous motion of unique points to other parts of the same objects the unique-point heuristic.



A good example of the unique-point heuristic is barberpole illusion. A barberpole consists of a cylinder painted with a helix of red stripes on a white background. It rotate continuously around its central vertical axis, so all points on its surface actually move laterally at all times. Yet a rotating barberpole produces the illusion that the stripes move vertically up the pole. This perception is illusory because the "endpoints" of the stripes are not actually the same from one time to another. The correspondence arrived at by the unique-point heuristic is erroneous, the motion that is perceived will be correspondingly erroneous. If the stripes had a clearly visible texture -- for example, if they were made up of many red dot on a white background so that correspondences over time were unambiguous -- the barberpole illusion should disappear.

Physiological Mechanisms

The Mango and Parvo Systems. M(mango) cells respond more rapidly to changes in stimulation than P(Parvo) cells. They are highly sensitive to luminance contrast, have large receptive fields, and are low in spatial resolution. These general properties of the M cells gives us hint are the first step in the visual system's analysis of image motion.

Cortical Analysis of Motion. The processing of image motion is greatly expanded in area V1, where the first cells that are specifically sensitive to directed motion are found. There is a selective response to different directions in V1 sells.

Neuropsychology of Motion Perception. An intriguing neurological case has been reported of a highly specific deficit in human visual motion perception. The patient was admitted to the hospital complaining of the inability to perceive motion. She reported that the world appeared to her as a series of frozen snapshots. Brain imaging tests indicated that the lesion in her brain was located in the border region between occipital and temporal cortex, clearly outside of primary visual cortex. By some research, motion in depth appears to be processed at a later stage where her cortex was damaged, probably in MT and MST.

Computational Theories

Directionally selective cells are clearly present in visual cortex, and they obviously play a crucial role in our ability to perceive motion. But why do they respond as they do is not yet known with certainty. Computational theories mentioned below of how motion sensitive elements might be constructed have provided important additional insights.

Delay-and-Compare Networks. They compare what happens in one region of the retina with what happened shortly before in a nearby area, using some form of time delay mechanism.

Edge-Based Models. By analyzing the change in illumination over time in conjunction with an edge detector, a motion detector could be constructed.

Integrating Local Motion. IF a straight edge moves across an aperturelike local receptive field, its precise direction and speed are ambiguous. If two edges are part of the same rigidly translating object, however, they must both be moving in the same direction. Locally analyzed motions from a single, rigid object can be integrated by finding the vector that is common to their vectors.

2. Object Motion

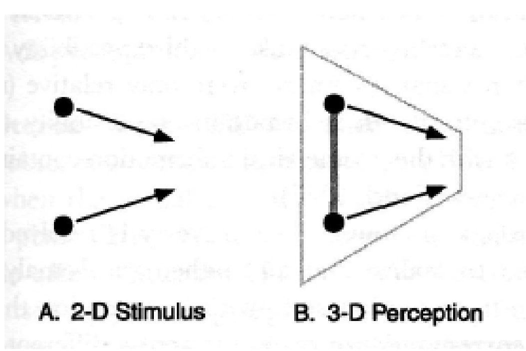
The information of image motion is not sufficient for us to handle events in our life, because what is ultimately needed is the knowledge about motion of objects in 3D world – **object motion**. No matter when we fix our eyes onto a flying bird or take a glance at it, the changes of retina distribution of luminance, i.e. image motion, are different, but we still perceive the same motion of the bird, which is called **motion constancy**. By combining the image motion, eye movement and other information such as depth of objects, our visual system can achieve at least approximate motion constancy.

Naïve experience supports the existence of approximate velocity constancy. If we watch the traffic on a freeway, the closer vehicles don't seem to move faster than farther ones, although this is what's happening on the retina. A more controlled experiment (Rock, Hill, & Fineman, 1968) shows the perception of speed does yield approximate constancy. Subjects were required to adjust the speed of two objects at different distances so that they had the same velocity. The distant one was four times farther the close one. When they were allowed to use only one eye, they roughly set the speed of the farther object four times faster than the closer one. This can be explained by their setting the speed of image motion to be the same. However, when subjects viewed the objects binocularly, with good depth information, they fulfilled this task easily. This experiment shows that as long as good depth information is available, observers can perceive objects' real-world velocity.

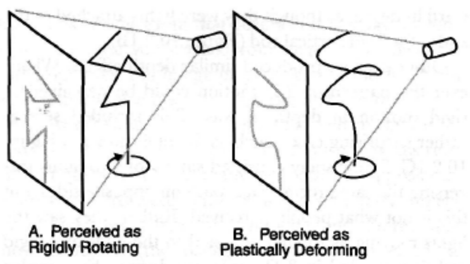
Depth and Motion

From the experiment above, we know that information about depth is very important for object motion perception. What if the depth is unknown? How would our visual system interpret image motion in terms of objects in 3-D world under such circumstance?

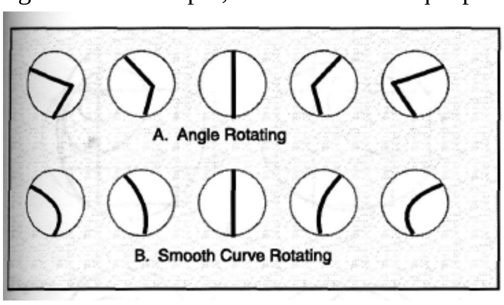
A study conduced by Gunnar Johansson (1950) discovered some powerful depth effects in motion perception. In one experiment, the subjects were shown two dots moving back and forth in synchrony in a plane as shown in the following figure. What most subjects perceived was not this simple retinal motion (A), but two dots moving forward and backward in depth, as though they were attached to a rigid stick (B). This experiment suggests an interesting notion called **rigidity heuristic** which states as: All else being equal, if there is an interpretation in which rigid motion can be perceived, it will be.



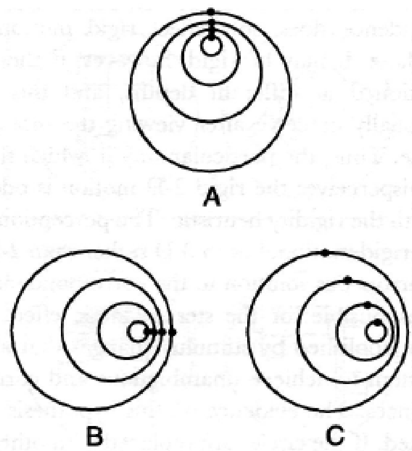
This heuristic can be further proved by another phenomenon known as the **kinetic depth effect (KDE)**. A shadow-casting technique was used by Wallach and O'Connell to project an angular 3-D wire object to a 2-D image. Subjects could only see the images on the plane. When the wire was rotating around its axis, they perceived a rigid 3-D wire rotating in depth which was thought as merely a 2-D figure when stationary. Apparently, observers could combine images at different time to form a perception of a rigid object with the rigid heuristic. However, if the wire was smoothly curved without angles, observers only reported plastically deforming curves in the projected plane.



This experiment leads to two important conclusions. One is the previous mentioned rigidity heuristic. People tend to interpret 2-D changing images as rigid moving objects in 3-D space. Another one is the significance of correspondence problem to the object motion perception. The **structure-from-motion theorem** shows that mathematically if one knows the correspondence of points from each view to the next and if one assumes that the object is rigid, then it is possible to recover both the 3-D location and the motion of the object from four non-coplanar points in three distinct orthographic projections. But if the correspondence problem is not solved correctly, things will become quite interesting. The figure below shows the different perception due to correspondence problem. A rotating wire is observed through an aperture. When a sharp angle rotates in depth, its vertex is a unique point

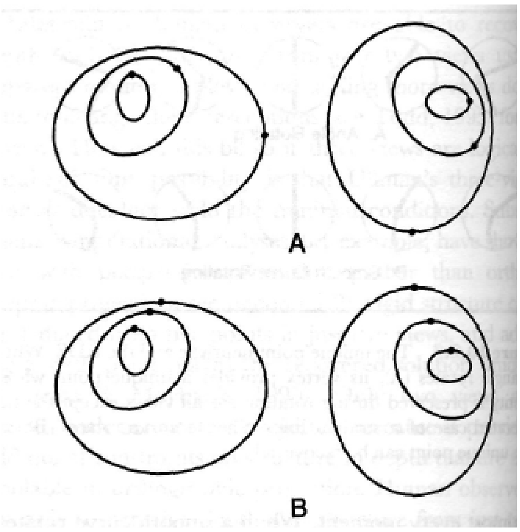


which can be tracked through different views. Then, the 3-D rotation is perceived. When a smooth curve rotates in depth, since there is no unambiguously unique pint, only 2-D deforming could be perceived.



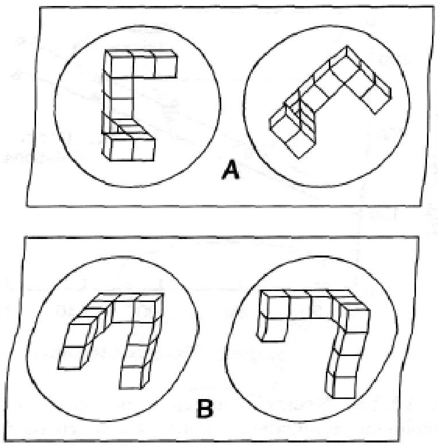
Another well known phenomenon related to different perceptions of motion because of different solutions to the correspondence problem is called **stereo-kinetic effect**. Put a picture like part A in the figure above without solid dots on a turntable. When rotating the turntable, we will experience motion of a 3-D object, a rotating cone protruding out of the turntable plane or a rotating conical tune receding into the plane. The reason why we perceive 3-D motion rather than veridical 2-D image rotation lies in the incorrect interpretation of the correspondence problem. Since each point is equivalent on a circle, we have no strong cue to find corresponded points. Part B and C show two possible solutions. The actually corresponding points are shown in part B and we should perceive veridical rotation in the turntable plane, but we don't have stimulus evidence supporting this answer over others. Another possibility is shown in part C where the uppermost points are corresponded to each other. In this case, 3-D motion is perceived. In the experiment, most observers report the second case and it would be expected if the visual system prefers translations over rotations in the absence of obvious evidence for rotation. If we paint dots on the rotating circles like part B, then correspondence problem will be solved in a unique way and a lot more report of 2-D rotation perception would arise.

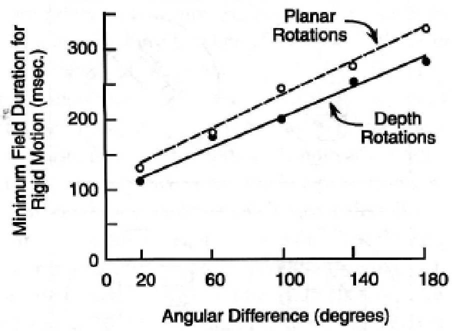
Because of the ambiguity in the solution of correspondence problem, even non-rigid 3-D motion can be perceived from 2-D rigid motion. The next figure shows a experiment similar to the preview stereo-kinetic effect example. Now, ellipses are placed on the turntable. There are also more than one possible solutions for corresponding points, but none of them will give lead to a rigid motion in depth. Instead, a deforming rotating cone or tune is perceived by some subjects with the solution of part B.



Long-Range Apparent Motion

In the preview discussion about short-range and long-range apparent motion, we have seen that the long-range apparent motion is associated with object-level process. In this section we can further find that it's fundamentally a 3-D, object-based perception. Shepard and Judd (1976) studied the phenomenon of **apparent rotation**. As illustrated below, two sets of displays were shown to induce the perception of picture-plane rotation (A) and 3-D rotation in depth (B) respectively. The apparent motion is vivid only within a bounded range of alternation





rate. Shepard and Judd studied the fastest alternation rate at which observers reported smooth rigid motion of a single object. From the result shown above, the minimum field duration for rigid motion is linear to the angular difference. With larger angular difference, more time between two frames is required for observers to perceive apparent rotation. This suggests that the process underlying apparent rotation is an analog process, that is, a continuous process that requires time to go through intermediate orientations, just as real objects do when they rotate. Also, this experiment shows that the linear increase in picture-plane rotation and 3-D rotation is virtually identical, which suggests that the apparent motion takes place in an internal representation of 3-D space.

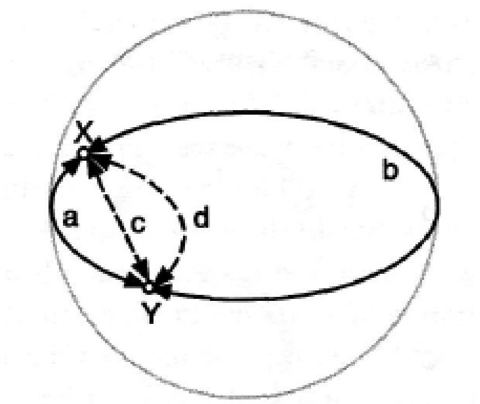
To explain these discoveries in the long-range apparent motion, Carlton and Shepard proposed a geometrical model based on the following three ideas.

1. *Motion hyperspace*. Any continuous motion can be presented by a continuous path in a

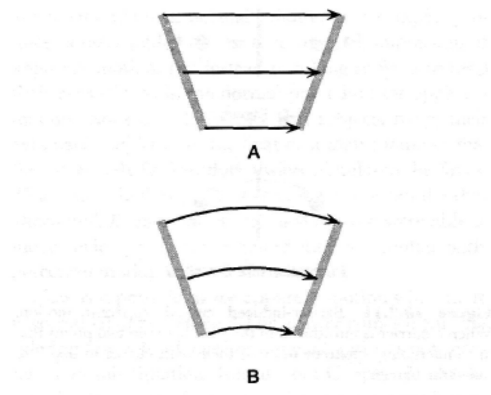
high-dimensional space (or **hyperspace**). Rigid rotations (e.g. path a, b in the following figure) lie on a 3-D spherical surface (or **hypersphere**). Any other path (e.g. path c, d) lying off the surface corresponds to non-rigid motion.

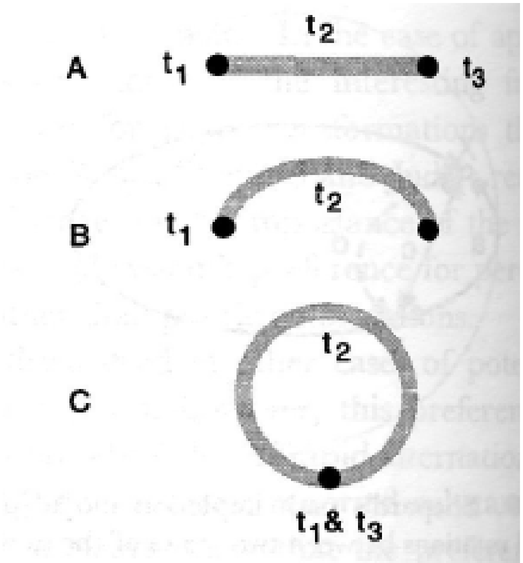
2. *Path impletion*. The two displays of the same object correspond to 2 points (X, Y in the illustration) on the hypersphere. The visual system picks one of many possible paths between these two points (impletion) when interpreting the apparent motion, corresponding to one of many possible motions.

3. *Maximum speed*. Rigid apparent motion will be perceived if and only if the alternation rate is sufficiently slow that the shortest path along the rigid-rotation surface can be traversed at or below a maximum speed. Otherwise, a shorter path off the hypersphere will be followed, leading to non-rigid motion perception.



Another phenomenon than can be explained by Carlton and Shepard's theory is **curved apparent motion**. In an apparent motion display, two lines at different orientation are alternated. One possible perception (A) would be a translation along a straight line path accompanied by a rotation around the center point. This trajectory would minimize the work required to move the object. Another possibility (B) is a global rotation around a point at the crossing of these two lines when extended. This trajectory minimizes motion in Carlton and Shepard's kinematic geometry and would result in a curved path apparent motion. The experiment showed that people perceived the second interpretation, supporting the geometrical model.





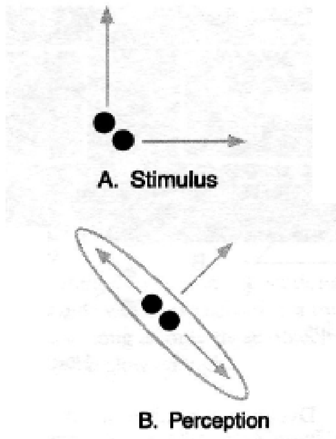
Another kind of curved path apparent motion is known as **path-guided apparent motion**. In this example, the left dot, middle grey path and right dot are displayed repeatedly in a sequence. Observers perceive a dot moving between these two dots along the grey path whenever it's straight, curved or even a circle. Shepard interpreted this result as evidence that the visual system has internalized the structure of motion that objects undergo in the real world to a remarkable degree allowing it to fill in the most likely motion in apparent motion displays.

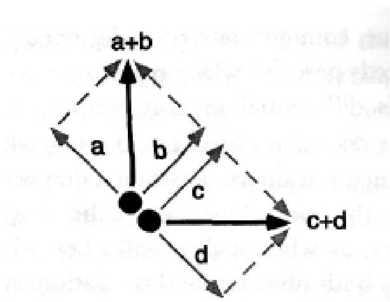
Dynamic Perceptual Organization

So far we have been talking about the motion perceptual of one object. How do you determine the organization we perceive in a scene with multiple moving objects? What regions are perceived to be part of the same object or group of objects and how they move with respect to each other?

We have a *common fate*: the tendency to group together units that move with the same velocity. Flocks of birds and schools of fish are naturally grouped together not only by similar shape, but also by the common velocity. Much more common are cases in which parts of a single moving object are separated in the retinal image by occluding objects so that the visual system much “put them back together”. The images of rigid connected objects tend to move in the same direction and at the same speed. When a train is running through a tunnel, even if only the head and tail are viewed we can easily consider them to belong to the same object by the same speed. The common fat can even destroy the best natural camouflage when an animal moves relative to its well color- or texture-matched environment.

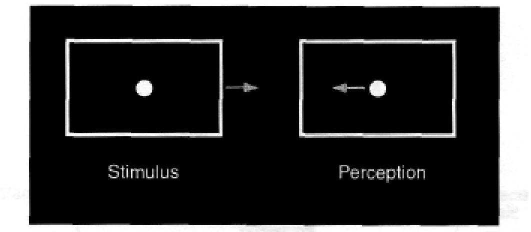
Another common fate is known as an organizational principle in motion perception. Gunnar Johansson studied the experiment of **configural motion** in 1950. A classic example of L configuration is shown in part A of the following figure. Two dots are moving harmonically in vertical and horizontal direction respectively. When each dot is viewed in isolation, this is exactly how they are perceived. However, when viewed together, these two dots are perceived to move in a group along a diagonal direction and meanwhile they are moving toward and away from each other along an opposite diagonal.





It appears that the visual system performs a vector analysis of the motion of objects into common motion and relative motion. Many ways of decomposition are possible and it’s not clear yet why only one is perceived. We don’t know if the common motion is extracted first and the relative motion is the residual or relative motion is analyzed first and the common component is residual.

Another phenomenon we often experience related to the dynamic perceptual organization is **induced motion**. When a cloud moves past the moon at night, it looks like the moon is moving through the cloud rather than the cloud is



moving. A controlled experiment illustrates this phenomenon more clearly. In a dark room, only a luminous dot and a large luminous rectangle are shown to the subjects. The dot is stationary and the rectangle moves slowly left and right. Subjects perceive the dot as moving and the rectangle as stationary. The majority of facts fit an account based on the following two assumptions:

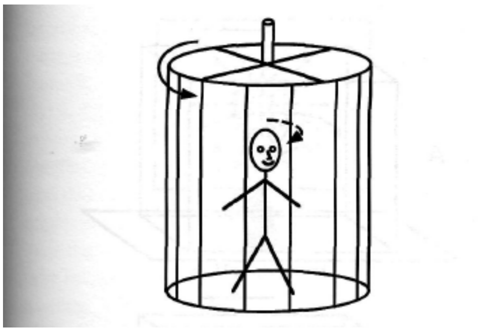
1. *Sensitivity to relative motion*. The visual system is more sensitive to relative motion than absolute motion. In the experiment, the rectangle moves fast enough for subjects to observe relative motion against the dot but still below the threshold of its absolute motion perception.
2. *Stationarity of the surrounding object*. With the perception of only relative motion, the visual system uses the heuristic of assuming that the larger or surrounding object is stationary to come to the conclusion that the dot rather than the rectangle is moving.

If the rectangle moves faster above the threshold of absolute motion perception or the dot is comparable in size to the rectangle, the induce motion would not take place.

3. Self-Motion and Optic flow

Induced Self-Motion

In discussing the induced motion, we have noticed one of the most important accounts is that the visual system tends to perceive the larger, surrounding object as stationary, given there is no evidence to the contrary. This heuristic can lead to a kind of induced motion that plays a special role in specifying our relation to the environment, called **induced self-motion**. The following figure is an example of induced self-motion. A stationary person seated inside a drum that is rotating counterclockwise soon experiences himself or herself as rotating clockwise and the drum as stationary. This experience is so compelling that many people become dizzy and nauseous, very much as they would if they were actually rotating.

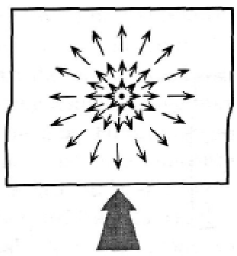


There are several differences between classical induced motion and induced self-motion. One is that classical induced motion requires the larger, surrounding object to move below the threshold of absolute motion, or observers will perceive veridical motion of the moving object. However, in the experiment of rotating drum, the drum rotates quite fast, well above the threshold and observers still have the perception of self-motion. Second, there is an initial period of veridical perception: You first perceive the drum as rotating and yourself as stationary. After a few seconds, the experience changes to one of self-rotation inside a stationary drum. A few explanations are proposed for these phenomena including motion adaptation and contradiction with information about self-motion from other sources such as vestibular input. But none of them can explain the induced self-motion perfectly yet.

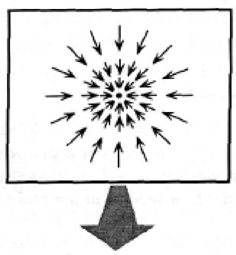
Perceiving Self-Motion

Although the experiment of induced self-motion looks like a trick, the perception of self-motion is indeed very essential for our daily life. When you are sitting in a cruising car at a steady speed, you only perceive the motion via visual information. In this case, the heuristic to take the environment as stationary gives you veridical perception of the motion. People tell the direction and speed of their motion mostly by the changing visual environment.

Optic flow is the global patterns of retinal motion when one moves relative to the environment. For example, the following figure (A) and (B) illustrate the optic flow in two simple cases. The retina image expands or contracts around one point as one walks towards or away from that point while fixing his or her eyes on it. A more complex case is shown in the third figure. That stands for the optic flow when one moves toward "x" while tracking "o" in an environment containing horizontal ground and two vertical walls. There is no true focus of expansion.

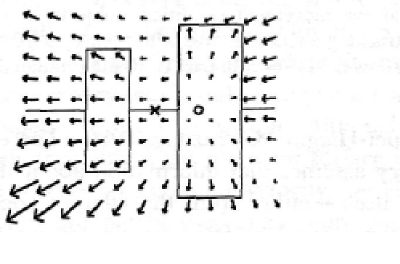


A. Motion Toward



B: Motion Away

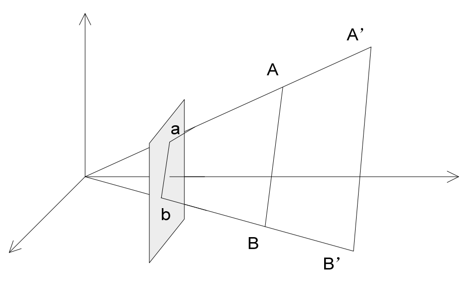
Flow field from moving toward X while tracking O.



It's very easy for us to tell the direction of movement from optic flow in part A and B, but is it possible to give the answer in the third case? There are two different explanations about how people perceive their heading accurately. The **retinal image theory** believes that the optic flow alone is sufficient for the visual system to extract information of both moving direction and eye movement. In contrast, the **extraretinal theory** assumes that information about the eye movement is taken into account by effectively subtracting its contribution to the flow filed. Several experiments were carried out to test which explanation is correct. Some experiments supported retinal image theory with slow eye movement, while some supported extraretinal theory with fast eye movement. Further experiment showed in certain environment, the optic flow alone is sufficient even under fast eye movement. A commonly accepted answer is not obtained yet.

In contrast to human's ability to tell the accurate direction, the speed of self-motion is not available solely from optic flow because speed is defined as the ratio of change in location over time, and the absolute distance cannot be computed from pure optic information. From the figure below, assuming the projection of a point moves from a to b on the retinal, its corresponded actual path can be AB or A'B'. As the distance of object increases by k times, the speed at which the object moves also increases by k times. In absence of the information about absolute distance, we're unable to tell the absolute speed.

However, we can specify another quantity called as **time to contact** which is the length of time it will take the observer to reach the surface



toward which he or she is heading under present motion conditions. It's invariant over scale transformations because it affects distances and speeds in the same way. Perceptually, this can be understood by the formula, time=distance/speed.

Virtual Reality and Ecological Perception

One problem about optic flow is Gibson's claim of *direct perception*: The observer's visual system does not need to add any extra information to arrive at an accurate perception of the world. He believes that optic flow provides so much information that it's sufficient for one to explore the layout of surfaces in the 3-D environment. Whether this theory is true or false depends on one considers from a logical or ecological point of view.

Logically, this claim is false because the 2-D image is not sufficient to reconstruct any 3-D environment. The technology of **virtual reality** demonstrates the inaccuracy of Gibson's claim. The virtual reality displays could create the same scene as perceived in a real world from two screens. So, it's not possible to tell whether we should interpret what we see as a 3-D world or just 2-D image screens solely based on the visual information.

Ecologically, the direct perception is correct because under natural viewing conditions in natural surroundings, people perceive their path through an optically rich 3-D environment with truly impressive accuracy. Natural environment provides us the heuristic to solve the ambiguity and to interpret the optic flow in a veridical unique way. Thus, the visual system usually gives us the true perception of our world under ordinary conditions.

4. Understanding Events

Even after moving images have been organized into coherent objects and their motions have been interpreted within appropriate reference frames including movements of the observer through the environment, further processing is required to understand the event that gave rise to the pattern of retinal motion.

Biological Motion

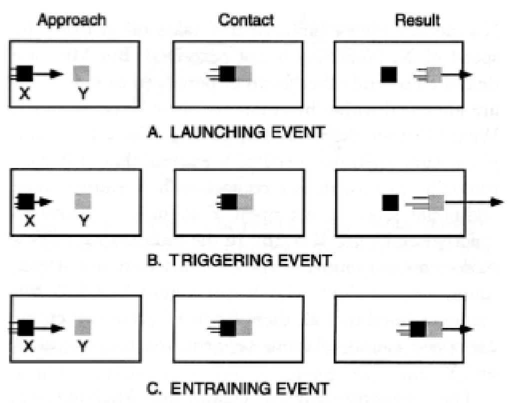
When viewing simple two-point configurations people have a strong propensity to interpret the moving points as though they were attached to a rigid rod moving in depth. This phenomenon led to predict that people would be able to perceive the movement of a human body from just the motions of the joints. To test this hypothesis, an actor filmed in the dark with small lights attached to his joints so that nothing was visible except the lights.

When the actor was seated motionless in a chair, observers perceived a meaningless configuration of points, rather like a constellation of stars. But when the actor start to walk, he was immediately and unmistakably perceived as a person in motion.

Perceiving Causation

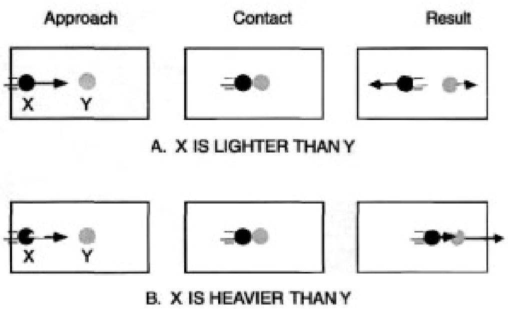
We see perceptually organized structures with important interrelations. One particularly important component of motion events is their causal structure.

Launching effect is basically when one object hit another, the object been hit will start to move in a slower than original speed. The **triggering effect** is the object been hit move in a faster than original speed. The **entraining effect** is the object doesn't stop after hitting another object.



Perceiving Mass Relations

We cannot literally see whether one object is heavier than another when they are stationary. But, when they collide, more information becomes available through the dynamics of their interaction. If one ball rolls into another and ricochets backward, it appears to be substantially lighter than the ball that it hits. If it hits the second ball and keeps rolling forward, it appears to be substantially heavier.

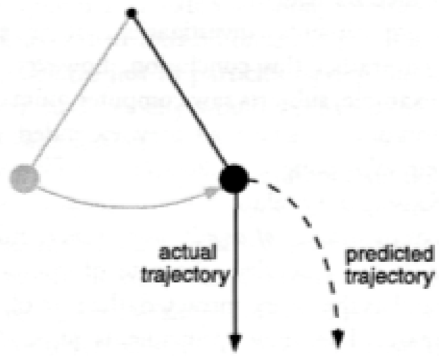


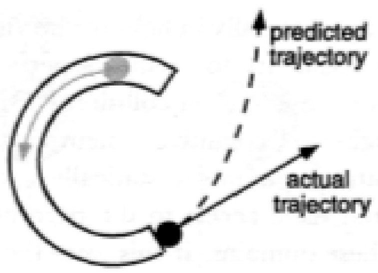
From above, we appear to make our judgments on the basis of two simpler heuristics:

1. **Ricochet heuristic**: When the incoming ball ricochets at a higher velocity than the forward motion of the initially stationary ball, the stationary ball is heavier than the incoming one.
2. **Clobbering heuristic**: When the stationary ball moves off with high velocity, the incoming ball is heavier than the stationary one.

The two heuristic conflict when the ricocheting ball moved more slowly than the ball it struck. Therefore, people appear to be recovering information about mass from visual information, but not with the quantitative precision and sophistication.

Intuitive Physics





The high-level cognitive system that is responsible for generating solutions to statically stated problems simply does not have access to the perceptual system's expertise in these domains. Thus, dynamic animation displays produced substantially better performance than static displays. These support the view that the visual system is reasonably good at discriminating between physically natural and unnatural versions of dynamic events as they unfold over time but relatively poor at doing so when the problem is presented statically.

5. Conclusion

How do we perceive image motion and how to transfer image motion to real motion is the most fundamental problem in our visual systems. The continuous motion will affect the binocular nervous system, showing adaptation and aftereffects, simultaneous motion contrast, and autokinetic effect. Apparent motion is the most important phenomenon in our daily life. TV, movie, many kinds of signals are produced based on apparent motion. In apparent motion, there are some tricks like aperture problem will deceive our eyes. After receiving the stimulation from retina, the nervous system will process the signal and then generate motion information of objects. Finally, the brain will gather all information to determine if the object is moving, where it moves to, and how fast it moves. Thus generate image motion in our brain.

The object motion perception is a relative high level of visual process compared to image motion perception. The depth information is important for perceiving object motion in depth. Without independent knowledge about the depth, our visual system will try to interpret the retina image motion into 3-D object motion using rigidity heuristic. The absence of unique point may leads to ambiguous solutions to the correspondence problem, resulting in different motion perceptions. The method our visual system use to pick one out of various solutions is not completely clear yet. When more than one moving object is presented, we tend to group objects with similar movement together and compute common motion for the whole group. In the process of motion perception, hierarchical reference frames are extracted: the environment, the whole group and the objects. The visual system tends to assume the larger, surrounding frame to be stationary and the smaller, embedded one to be moving.

The perception of self-motion plays an important role in our daily life. We spend lots of time everyday processing the optic flow and learned how to tell the direction and speed of self-motion from visual information. The argument over whether other information such as eye movement and vestibule input is required in this process still continues.

After we perceive the motion of objects, we can acquire the relation between them. We can tell it’s a person who is dancing or walking but not some meaningless dots. And from the speed before and after two object collide we can infer the relative mass and causation. Also we have better physics knowledge from dynamic simulation than stationary snapshots.

Most studies on object motion perception were conducted in a psychophysical approach, and few physiological evidences have been obtained so far. Although many theories have been proposed to explain phenomena such as long-range apparent motion and induced self motion, most of them looks quite subjective and none could explain them generally, consistent with most cases. Also, many claims (e.g. theories in explaining curved path apparent motion) are difficult to be tested in a psychophysical way. In my opinion, physiological studies, though difficult in such a highly integrated level, are needed for comparing different theories and providing new clues for discovering these mysteries.