Towards a System for the Interpretation of Moving Light Displays

Richard F. Rashid
Computer Science Department
University of Rochester
Rochester, NY 14627

TR53

May 1979

Abstract:

"Lights" is a system under development for the interpretation of simple moving light displays of jointed objects against a stationary background. The displays being studied differ from those examined by previous researchers in that (1) objects are represented by a relatively small number of points, (2) objects are not rigid, and (3) the viewing geometry is such that highly varying degrees of perspective distortion occur. An algorithm is presented which segments the points of an MLD of a wire-frame man into body parts. The relationship of this algorithm to previous theories of MLD perception and actual human performance is discussed.

The preparation of this paper was supported in part by the Defense Advanced Research Projects Agency, monitored by the ONR, under Contract No. N00014-78-C-0164.

I. Motivation

While the study of human perception of motion by psychologists has deep historical roots [16], only recently have researchers in computer science begun to look at machine perception of multiple frame images. One reason has been the large amount of computer space and time required to store and process sequences of movie frames. Another is the belief that single frame vision must precede motion vision; this has had the effect of reducing motion understanding to the previously unsolved problems of picture understanding.

There are, however, strong reasons to believe that the perception of motion is a fundamental component of the human visual system and not an attribute derived from the integration of successive perceptual events. Research into the human visual system has demonstrated that the retina and lateral geniculate nuclei do not encode static information but transmit changes in state to the cortex. Without microscopic eye movements called micro-saccads, static images projected on the retina quickly fade and disappear [19].

It is even possible to demonstrate 'pure' motion understanding. Veridical three dimensional perception of complex objects in motion can be achieved in the absence of all visual information save the position and velocity of certain highlighted points. Over the last twenty years psychologists have taken advantage of this fact by using displays of moving lights to study human motion perception. Such displays isolate and present pure geometric evidence of motion divorced from other factors such as texture, color and lighting.

This paper describes a system under development for the interpretation of simple moving light displays of jointed objects against a stationary background. A 'moving light display' (MLD) is defined to be a sequence of binary images representing points of one or more moving objects in an actual or synthesized scene. Object joints and corners are typically the points chosen for display.

MLDs possess a number of advantages over greyscale movies as a starting point for research in computer motion perception. Their reduced data rate (usually less than 30 points per frame) makes it possible to investigate their properties with relatively modest hardware. Moreover, the problem of MLD understanding does not easily reduce to static frame analysis. Individual frames cannot normally be recognized by a human subject in a psychological experiment [13] (see figure 1).

Another advantage of MLDs is that they can be easily synthesized, thus allowing a great variation of movement patterns and parameters for investigation. Data for the research described here currently comes from a program (based loosely on [8]) which simulates a range of human walking motion in 3-D.

II. Human performance in interpreting MLDs

Presented with a movie screen on which a small number of moving points are projected, a human observer will invariably try to place a three dimensional interpretation on their movements. This is true even when abundant evidence of two dimensionality is present, such as the edge of the screen and the sound of the projector. Psychologists researching this phenomenon have remarked on the strength of this illusion:

> It is a common characteristic of all the experiments I have described that the observer is evidently not free to choose between a Euclidean interpretation of the changing geometry of the figure and a projective interpretation. [15]

Two points opposite an imaginary center and tracing an ellipse evoke the perception of a rigid rotating rod. Four points describing the motion of the corners of a square create the perception of a moving plane. Frequently subjects describe 'seeing' an imaginary set of lines (called "subjective contours") connecting the moving points.

For simple MLDs consisting of a small number of points, it could be argued that such three dimensional interpretations result from a tendency to consider all points in an MLD to be related by rigid connections. The appearance of depth would then derive from an attempt to 'explain' the two dimensional display. Experiments have shown, however, that the effect also holds for more complicated MLDs.

In a series of experiments, Johansson has demonstrated the power of twelve moving lights to evoke the illusion of a walking man. His MLDs were created on video tape through the use of high intensity lights and adjustments of video contrast. Subjects performed a variety of tasks wearing glass bead reflectors on their major joints (shoulders, elbows, wrists, hips, knees and ankles), and the resulting MLDs of human body motion display considerable complexity. Nevertheless, less than .2 seconds are required for perfect recognition of an MLD as a moving man [14]. Only .4 of a second is necessary for discrimination of different human movements, e.g. walking left, walking right, walking backward, etc.

In an MLD of a moving man it is not the case that all points are seen as connected. Rather, the pattern of motion is correctly analyzed as a combination of motions of rigid parts connected by joints. A considerable amount of information can be recovered from such MLDs. Cutting has recently demonstrated the ability of subjects to recognize the sex of a walker [7]. It is also possible to recognize the gait of a friend [6].

III. Theories of MLD interpretation

Early workers on the kinetic depth effect tended to frame its explanation in terms of traditional 'cue theory'. Wallach and O'Connell [21], for example, speak of simultaneous changes in two visual dimensions (size and orientation) as a cue for rotation in depth.

Working from the Gibsonian viewpoint [10] that the eye is sensitive to continuous three dimensional perspective transformations, Johansson and his colleagues Borjesson and von Hofsten have attempted to explain the interpretation of MLDs in terms of a low level "spatio-temporal differentiation and integration" [14]. The outer layers of the visual system, according to this theory, extract a hierarchy of coordinate systems that permit the interpretation of motion patterns according to a simple vector analysis.

In his 1976 paper Johansson describes the theory as it applies to the interpretation of the hip-knee-ankle system of an MLD of a man walking parallel to the viewing plane. The hip is identified as moving in the coordinate system of the stationary background. The knee moves in the coordinate system of the hip and the ankle in the coordinate system of the knee. Each point's total motion is seen as the composition of a movement relative to its particular coordinate system with the motion of that coordinate system relative to the next in the hierarchy.

Johansson suggests that the selection of a coordinate system for a point depends upon its two dimensional velocity. The lowest velocity point is interpreted relative to the stationary background and so on down the hierarchy. Unfortunately, this criterion does not always work even in his simple example. At certain points of the walker's step, e.g. when his foot is in contact with the floor, the movement of the ankle is actually less than the movement of either the hip or knee.

Despite their difficulty in defining rules for the determination of a coordinate hierarchy, Johansson et al. have presented a large body of data to corroborate their claim that the human visual system is performing a kind of vector decomposition in the analysis of MLDs. Their theory has led to the correct prediction of several MLD effects [2,3,4].

A radically different approach has recently been suggested by Ullman [20]. He has demonstrated that three distinct orthogonal projections of four non-coplanar points provide sufficient information to reconstruct mathematically the three dimensional structure of the object defined by the points (subject to a possible reflection). Using this "structure from motion" theorem, Ullman has written a computer program capable of deriving the structure of multiple rigid objects in motion. He has also suggested an algorithm for the interpretation of MLDs of certain objects viewed by perspective transformation. He divides an object into rigid groups of four non-coplanar points, iteratively classifying overlapping groups of points in order to extract their relative three dimensional location. The accuracy of this algorithm depends on the distances between the points selected in each step of the analysis. They must be close enough to each other (relative to the viewing distance) so that to a first approximation they are viewed by orthogonal projection.

A mathematical approach to the interpretation of stick figures in Badler's 1975 thesis [1] can also be applied to MLDs. Badler proved that two points traversing parallel three dimensional paths and viewed by polar projection can be localized in three space. This is a much stronger statement than that made by Borjesson and von Hofsten's two-point vector analysis which would simply classify two points moving in convergent paths on the image plane as "moving in depth".

IV. Lights: A system for the interpretation of MLDs

None of these theories of MLD perception provide an algorithmic basis for the interpretation of complex images. Ullman's approach cannot cope with the low degree of connectivity, the perspective distortion, or the non-rigidity of MLDs such as those of human motion created by Johansson. Johansson, on the other hand, presents only a partial solution, leaving out important details such as the determination of connectivity and coordinate bases. Lights is an ongoing project aimed at producing a computer program capable of MLD recognition and interpretation. In its present form it is able to track and cluster points belonging to independently moving objects. Within a cluster, Lights analyzes the relative motions of object points. It then performs an initial segmentation into groups representing independently moving subparts.

Input

The primary stimuli for Lights are synthesized MLDs of two 'men' walking along different paths on a plane. The distance of each man from the hypothetical viewer varies from about two to four times the man's height, creating an overall change in perspective distortion of 2:1. Both men are seen to take about five steps in five seconds. Each frame is displayed for about 25 milliseconds. The point of visual fixation remains constant throughout the MLD (see figure 1).

These MLDs were created by a program (written in SAIL) based on a model of human walking movement developed by Cutting [8]. Taken alone, the motions of the shoulders and hips define two ellipses having different major and minor axes. The arms and legs swing as double pendulums from the shoulders and hips and the entire body moves forward with each step. Speed of stride may be varied. As the speed is increased, a forward lean and accentuated arm and leg swinging are added. Other stimulus parameters include hip and shoulder excursion, speed, size, and three dimensional path and orientation. The path of movement is defined either by a SAIL procedure which takes the current distance 'walked' and returns a three dimensional coordinate or by a chain-coded path on a plane interactively specified on a screen (CRT) with a computer 'mouse'. The direction faced by the man is tangent to the path at all times.

Although referred to as a 'walking man', the underlying model is actually that of a wire-frame figure, since no attempt is made to occlude points on the basis of body part widths. Nevertheless, the net effect is a stimulus universally identified by human observers as a walking (albeit transparent) man. A full description of the program can be found in [18].

Tracking

For each frame, the input to the interpretation program is an unlabeled set of coordinate pairs corresponding to the points of the MLD. The first problem, therefore, is that of tracking points from one frame to the next.

In the general case, the problem of tracking is the correspondence problem which has been looked at extensively by researchers in stereo vision. Recently Ullman has suggested a network relation algorithm appropriate for moving rigid objects [20].

The MLDs under study here contain a small number of points and depict objects with parts in relative motion. Our solution to the tracking problem is to select a correspondence which minimizes the difference between the expected position of a point (based on its velocity averaged over the preceding two frames) and the actual position of a point in the next frame. A conflict occurs if a point in the next frame satisfies the minimum distance criterion for two or more points in the current frame. This conflict is resolved by selecting the correspondence which minimizes the sum of the distances for all pairings. For the first frame no velocity is known and the algorithm reduces to finding the closest neighbor.

In practice this approach has worked extremely well. For MLDs derived from analytic functions (e.g. a man walking in a circle or straight line) perfect tracking is the rule. When the stimulus is generated by a chain-coded path, discontinuities of motion can cause tracking errors detectable only by later program stages. In all cases, tracking errors occur during frames which also cause difficulty for the human tracking system.

In one example, a roughly triangular path was drawn. Unfortunately the best hand rounding of the corners still left them too sharp for smooth motion at the turns. Nevertheless, the display was considered acceptable. It was shown to a number of graduate students over the span of a few weeks, and all reported seeing a 'normal' man walking along a triangular path with sharp turns. When the tracking program was run using this MLD as input, it mismatched the right knee with the left ankle after the first turn in the triangle. These two points and the left knee displayed in isolation provided the author with the same impression of a switchover noted by the tracking routine. Alerted to this illusion and re-shown the original MLD, all students saw the "ankle turn into the knee" even though that was inconsistent with their interpretations of the rest of the display. This human tendency to ignore tracking errors unless they are explicitly pointed out suggests a strategy for handling such difficulties. When confronted with a possible conflict, the interpretation program can simply suspend judgment on the identity of questionably matched points, waiting for a clear interpretation to present itself in later frames.

Object Separation

Separation of MLD points into groups belonging to different objects is the next stage of Lights' interpretation process. The underlying assumption is that independently moving objects can be differentiated on the basis of their projected movement and position. When this assumption is violated, as in the case of two dancers arm in arm or soldiers marching side by side, the claim is that an MLD provides insufficient data to separate the objects. Higher level knowledge must be employed to distinguish them.

Ullman [20] has criticized the grouping of elements into bodies as a prelude to depth analysis. He makes the valid point that a gestaltist grouping of points by "common fate" is inadequate for the separation of complex MLDs. Potter's criterion [17], for example, groups two points whenever their velocity difference falls below a defined threshold. This will not work for the MLDs of two rotating cylinders used by Ullman. In such displays each cylinder contains points spanning a range of velocities and both may contain points moving at exactly the same speed.

The fact that simple rules for grouping points do not work should not be taken as sufficient grounds for abandoning the idea of low level object grouping. Ullman was quick to give up object clustering because absolute structure determination was possible for his images. This solution is not available for the less restricted domain represented by MLDs of walking men.

A way around Ullman's objections can be found in the techniques of graph-theoretic cluster analysis. Single-linkage cluster analysis has been successfully used to handle a wide range of 'difficult' problems such as separating two touching Gaussian distributions of points and determining gradient clustering [22]. This technique, based on the computation of the minimal spanning tree (MST), is used by Lights to distinguish independently moving objects.

Let every point in an MLD frame be represented by the four-vector (X,Y,VX,VY), where X and Y are its projected position and VX and VY its projected velocity. A graph can be constructed which has each point as a node, with each node connected to all others by an edge of cost equal to their Euclidean distance. Information from previous frames is included by adding to this edge cost a function of the cost of the same edge in past frames. A minimal spanning tree can then be built [11] and the resulting graph can be segmented into clusters based on an appropriate cut function.

Figure 2 shows the result of this process on a frame of an MLD of two rotating cylinders viewed in orthogonal projection. Thirty points were placed on each cylinder in such a way that no boundary could be seen in a static view of the first frame. After seven frames the MST for these points was calculated based on a cumulative distance function. While the projected velocity of points moving nearly parallel to the viewing plane did differ greatly from that of points moving nearly perpendicular to it, no sharp divisions occured within a cylinder because the speed of a point was close to that of its neighbors. On the border between the two cylinders their different rotational velocities (four degrees per frame for the upper cylinder and two degrees per frame for the lower) resulted in a discontinuity which was found by the cluster analysis.

When a perspective rather than an orthogonal projection is used, changes in scale caused by varying degrees of distortion can detract from the usefulness of data collected in previous frames. To compensate for these changes and for the mismatch in the units measuring velocity and position, each dimension of the four dimensional feature space is scaled to unit variance and translated to zero mean. Single frame distances between features in this new space are combined with previous values to form a measure of the distance between points over a number of frames according to the function:

$$CD_n(i,j) = d(i,j) + CD_{n-1}(i,j)*.95,$$

where '$CD_n(i,j)$' is the combined distance between points i and j in frame n and d(i,j) is the Euclidean distance between points i and j in frame n.

Figure 3 shows the MST for two men, one walking in a circle, the other in a triangle, after 30 frames. Figure 4 below it shows another MST, this one calculated for two walking men traversing intersecting paths. In both cases a cut between the two groups of points could be made in 25 frames or less (about one-half step). Both were complicated by the fact that the projected positions of the two groups were initially close and by the fact that in both cases the men were made to walk 'in step' rather than show completely unrelated movement patterns. Greater independence of movement would hasten the clustering process.

The criterion for separating clusters was conservatively chosen. Two clusters were assumed to be unrelated when the cost of the MST edge separating them was over fifty percent larger than the average cost of the edges near its two endpoints. A cluster was required to have at least two points.

It should be noted that single-linkage clustering is but one of a group of clustering techniques including complete-linkage and average-linkage (King's method) clustering. Investigation is proceeding on the usefulness of these different clustering techniques and on the choice of cut criterion.

Intra-Object Relationships

An object in motion can be thought of as defining a moving coordinate system. Object parts move relative to that system and in turn define their own frames of reference. These two facts reflect not only the mechanics of motion but also its normal perception by a human observer.

And yet, particularly in the case of MLDs, this correspondence between object and percept seems singularly fortuitous. An infinite number of motions of points in space can produce a single MLD, and once a three dimensional interpretation of structure is arrived at, it does not necessarily resolve such questions as "what parts of an object are connected?" and "how are unconnected parts related?".

An informal experiment was devised to see how a group of graduate students and faculty members interpreted ambiguous connectivity information in MLDs. A display was constructed similar to the 'walking man' displays discussed earlier but with the difference that the man remained rigid throughout his motion about a circular path (see figure 5). The result corresponded roughly to a scene in which a mannequin is wheeled around in a circle or rotated on a lazy susan. Not only was the display understood as a rigid group of points moving through space, it was recognized immediately as a man in a fixed position. Other displays of rigid objects showed this same tendency to evoke a single perception of connectivity, despite the fact that all their points were equally 'connected' in the sense that an imaginary rod could be extended between them.

Certainly in the case of the rigid man moving in a circle, part of the explanation must lie in the sophisticated pattern matching abilities of the human mind. This may not, however, be the only reason. It may also be the case that the mechanisms used to interpret the structure of an object seen in an MLD are sensitive to certain relationships in the stimulus pattern, resulting in a tendency toward certain interpretations.

Whether or not this represents a credible theory of human vision, it is the case that the relationship calculation done by Lights on the points of an MLD (see previous section) can suggest connectivity in the underlying objects. Figure 6 shows the MST for three rigid objects: a man, a cube and a jack. There is a high degree of similarity between the connectivity preferred by most observers and the connections favored by the relationship function on which the MST was based.

Initially it was hoped that this kind of clustering alone would lead to a natural breakdown of the object into subparts according to the following algorithm:

(1) Separate individually moving objects using MST clustering.
(2) Recalculate the MST for each object so

defined.

(3) Use this graph to define subparts.

Figures 7, 8, 9, and 10 show the results of this algorithm for the component objects of the MLD of two men walking in along triangular and circular paths.

Unfortunately, the groupings obtained from this algorithm did not always correspond with the correct division of the object. The reason is that the clustering algorithm is meant to identify closely related points from their two dimensional projection of position and velocity. For it to work properly the relationships between the three dimensional motions of connected points must be preserved. This will often not happen if the object as a whole is spinning or twisting in space.

The calculation of similarity should most properly be done relative to the coordinate system defined by the moving object. Two facts define that system: (1) the movement of its origin, and (2) its changing orientation relative to the stationary background (the orientation itself is not important because there is no one 'correct' orientation for the object's coordinate system).

An attempt is made by Lights to compensate for these factors. The centroid of the points defining an object is used as an approximation for the origin of the object's frame of reference (psychologists have also used the centroid, see [4]). Some compensation for the rotation of an object is achieved through a modification to the Euclidean distance function to allow points with equal but opposite velocity to be considered 'close' together. The resulting clusters more accurately reflect object composition.

The division of an object into its related parts is still subject to uncertainty. In the case of the walking man in particular, pseudo-relationships are often suggested based on similarity of motion of the arms and legs on opposite sides of the body. These graphs are useful nonetheless as a starting point for the next stages of the interpretation process - the recovery of three dimensional relationships and the matching of the stimulus to a known model.

3-D and Model Matching

This last part of the Lights system is speculative as it has not yet been implemented. The basic idea is to take the suggested relationship graph provided by the previous stages and elaborate a network of possible three dimensional interpretations based on [4]. Illegal or inconsistent relationships can feed back to previous levels for special consideration. Ultimately a consistent set of relationships can be used to try and match the stimulus with a known model.

V. Current State of Lights and Future Work

Although much has been done, Lights has not yet achieved its goal of MLD recognition and interpretation. The various components described here (tracking, object separation, and subpart separation) run in stages with no feedback. Eventually they will be integrated with the remaining system components into a multiprocess, message passing system conforming to the PLITS paradigm (see [9]).

The major areas in which research is not yet complete are: (1) determining the relationships between point motions that lead to specific three dimensional interpretation, and (2) model matching. In addition the problems of occlusion found in MLDs of actual human motion have not yet been resolved. Occlusion primarily affects the tracking and model matching phases of the interpretation process and should not present problems to the object separation and subpart determination parts of the system.

The major accomplishment of Lights in its present form is to show how an analysis of MLDs can proceed without making elaborate assumptions about object rigidity or type of projection. The techniques outlined here do not depend on exact mathematical criteria but will work for both perspective and orthogonal transformations and in the face of systematic distortion.

References

1. Badler, Norman, "Temporal scene analysis: conceptual descriptions of object movements", Univ. of Toronto Computer Science Technical Report No. 80, February, 1975.

2. Borjesson, Erik and von Hofsten, Claes, "Spatial determinants of depth perception in two-dot motion patterns", Perception and Psychophysics, vol. 11, no. 4, pp. 263-268, 1972.

3. Borjesson, Erik and von Hofsten, Claes, "Visual perception of motion in depth: application of a vector model to three-dot motion patterns", Perception and Psychophysics, vol. 13, no. 2, pp. 169-179, 1973.

4. Borjesson, Erik and von Hofsten, Claes, "A vector model for perceived object rotation and translation in space", Psychological Research, 38, pp. 209-230, 1975.

5. Braunstein, Myron L., "Depth perception in rotating dot patterns: effects of numerosity and perspective", Journal of Experimental Psychology, vol. 64, no. 4, pp. 415-420, 1962.

6. Cutting, James E. and Kozlowski, Lynn T., "Recognizing friends by their walk: gait perception without familiarity cues", Bulletin of the Psychonometric Society, vol. 9, no. 5, pp. 353-356, 1977.

7. Cutting, James E. and Kozlowski, Lynn T., "Recognizing the sex of a walker from dynamic point-light displays", Perception and Psychophysics, vol. 21, no. 6, pp. 575-580, 1977.

8. Cutting, James E., "A program to generate synthetic walkers as point-light displays", Behavior Research Methods and Instrumentation, vol. 10 (1), pp. 91-94, 1978.

9. Feldman, Jerome, "High level programming for distributed computing", to appear in CACM, 1979.

10. Gibson, J. J. and Gibson, E. J., "Continuous perspective transformations and the perception of rigid motion", Journal of Experimental Psychology, vol. 54, no. 2, pp. 129-138, 1957.

11. Gower, J. C. and Ross, G. J. S., "Minimum spanning trees and single linkage cluster analysis", Applied Statistics, vol. 18, no. 1, pp. 54-64, 1969.

12. Johansson, Gunnar, "Perception of motion and changing form", Scandinavian Journal of Psychology, vol. 5, pp. 180-208, 1964.

13. Johansson, Gunnar, "Visual perception of biological motion and a model for its analysis", Perception and Psychophysics, vol. 14, no. 2, pp. 201-211, 1973.

14. Johansson, Gunnar, "Spatio-temporal differentiation and integration in visual motion perception", Psychological Research, 38, pp. 379-393, 1976.

15. Johansson, Gunnar, "Visual motion perception", Scientific American, pp. 76-88, November, 1976.

16. Kolers, P. A., Aspects of Motion Perception, Oxford: Pergamon Press, 1972.

17. Potter, J. L., "Extraction and utilization of motion in scene description", Ph.D. thesis, University of Wisconsin, 1974.

18. Rashid, Richard, "Lights: a system for the interpretation of moving light displays", Ph.D. thesis, University of Rochester, in preparation.

19. Riggs, L.A., Ratliff, F., Cornsweet, J. C., and Cornsweet, T. N., "The disappearance of steadily fixated visual test objects", Journal of the Optical Society of America, 43, pp. 495-501, 1953.

20. Ullman, Shimon, "The interpretation of structure from motion", MIT, 1978.

21. Wallach, H. and O'Connell, D. N., "The kinetic depth effect", Journal of Experimental Psychology, 45, pp. 143-152, 1953.

22. Zahn, Charles, "Graph-theoretical methods for detecting and describing gestalt clusters", IEEE Transactions on Computers, vol. c-20, no. 1, January, 1971.

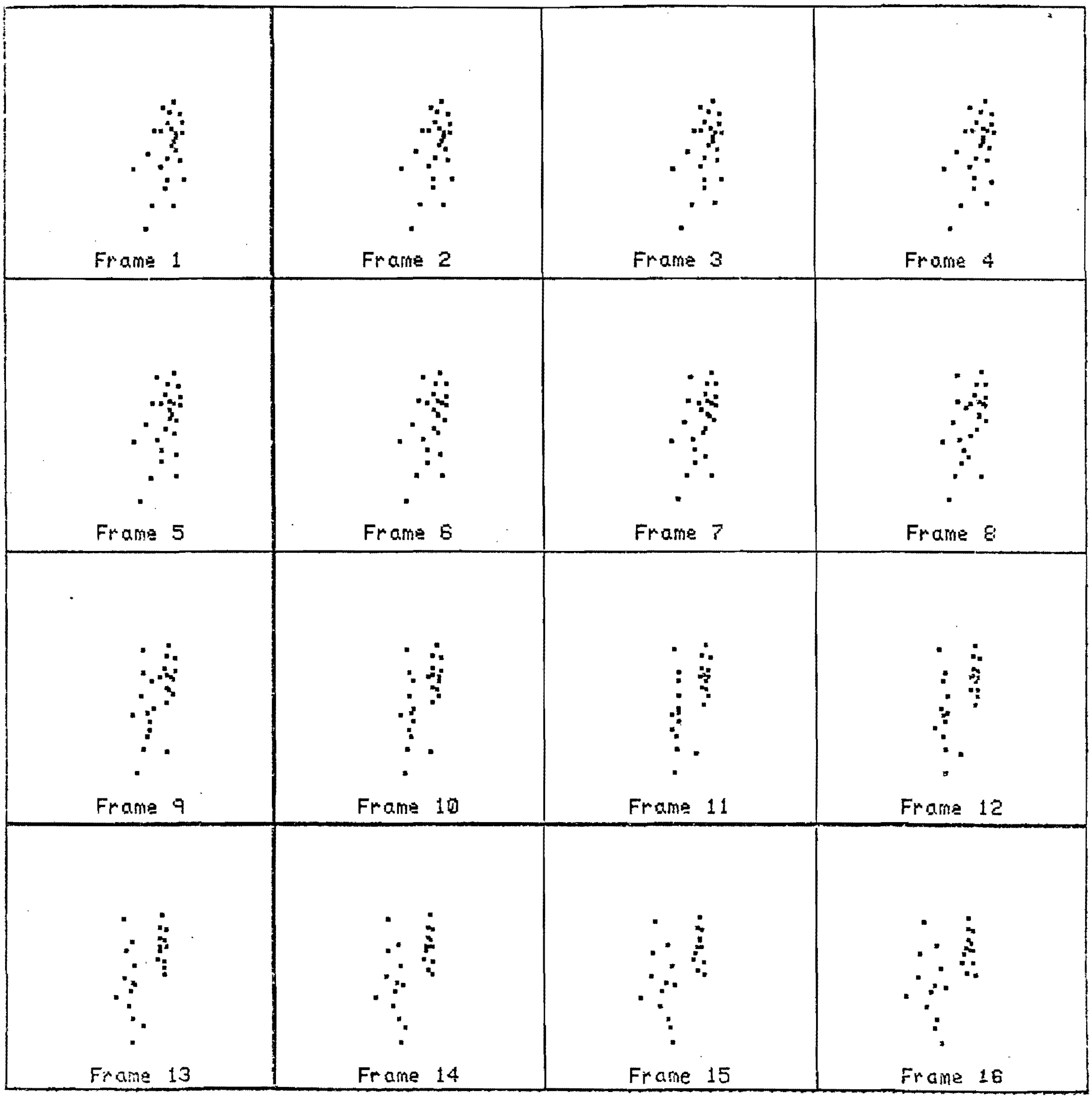


Figure 1
TriangleCircle.Lights
"Two men walking" - Sample Frames

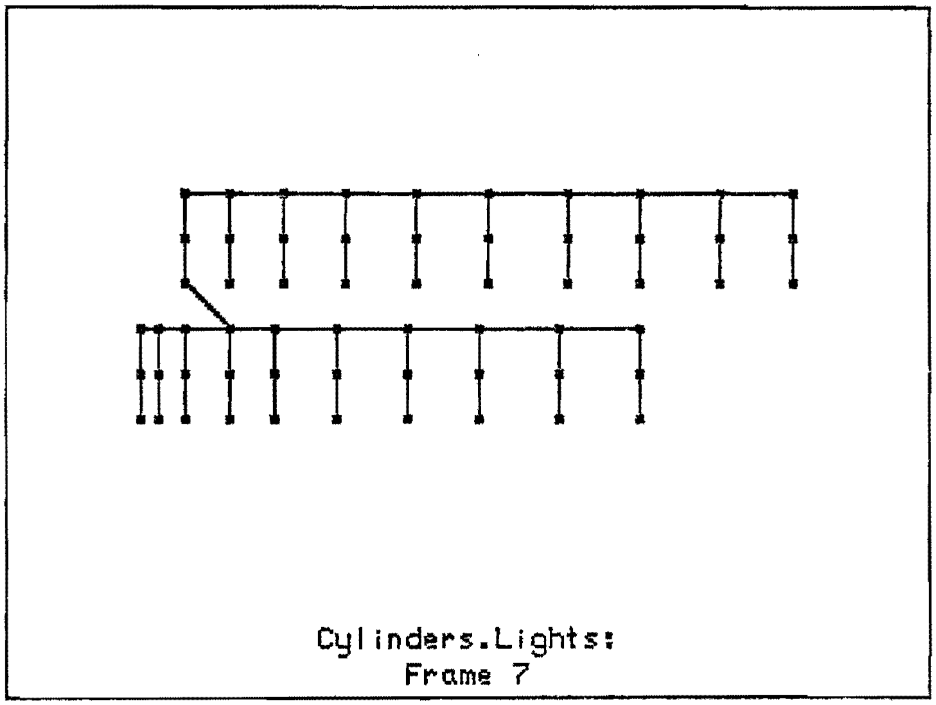


Figure 2
MST of points on two rotating co-axial cylinders

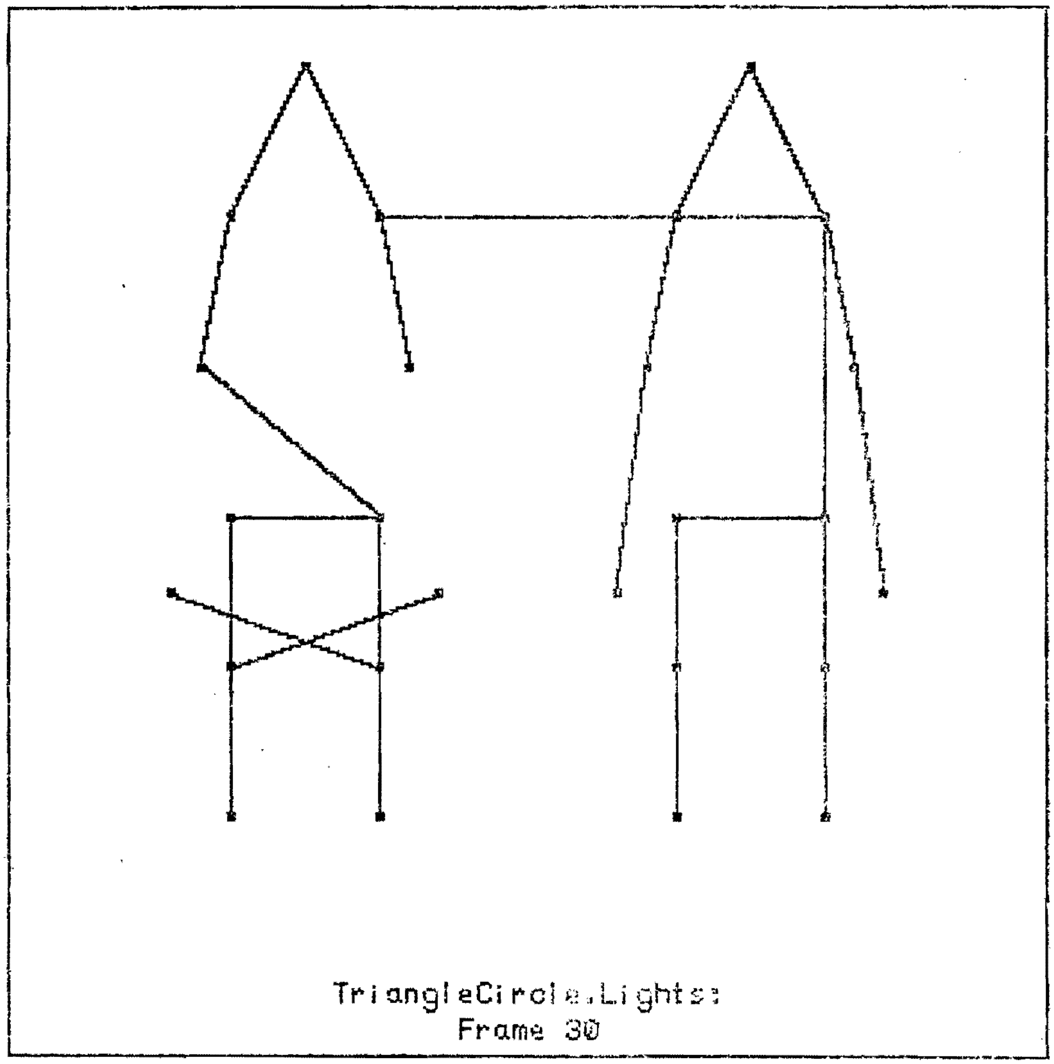


Figure 3
MST of two men walking, one around
circle, other along triangular path

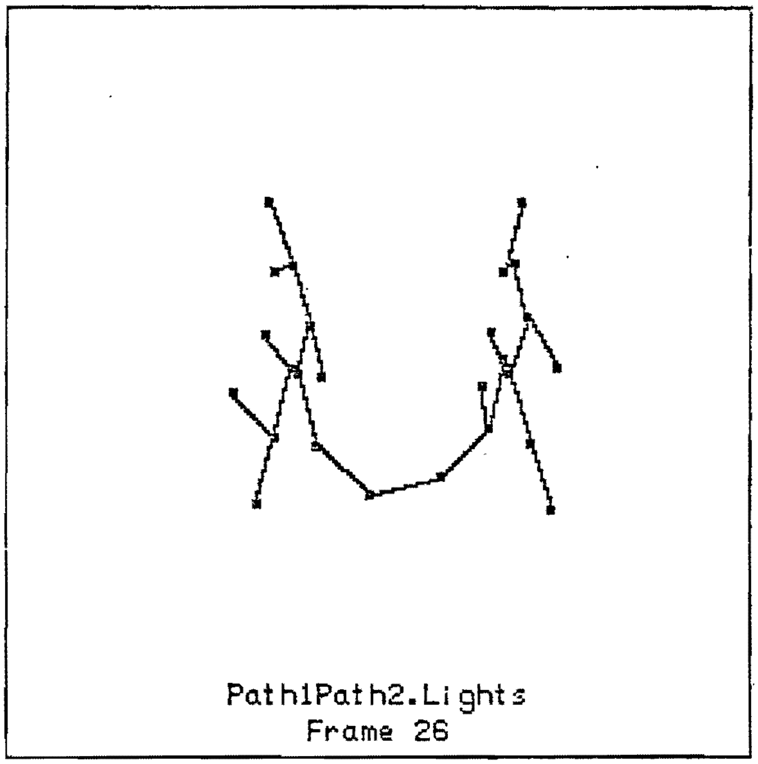
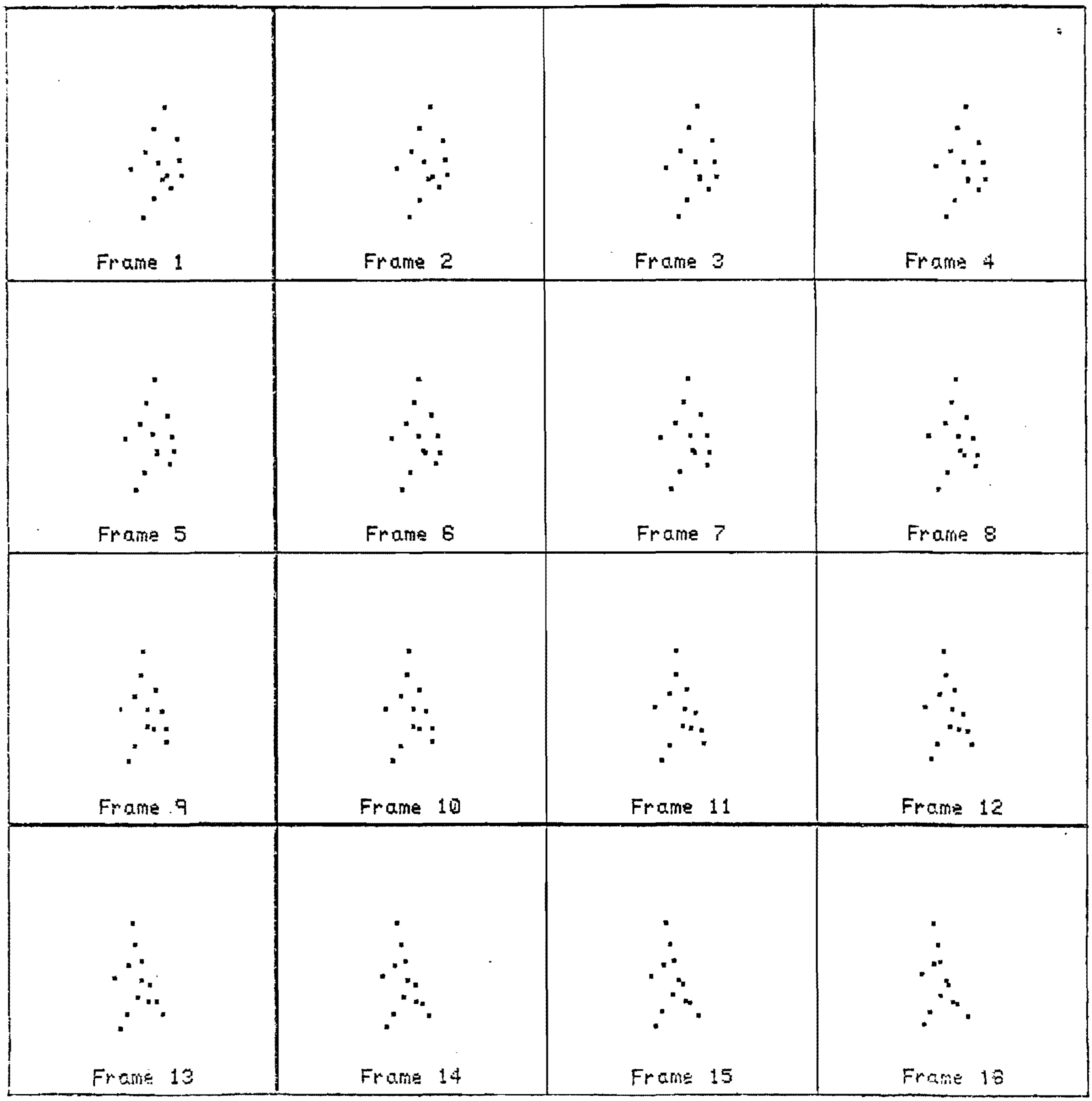
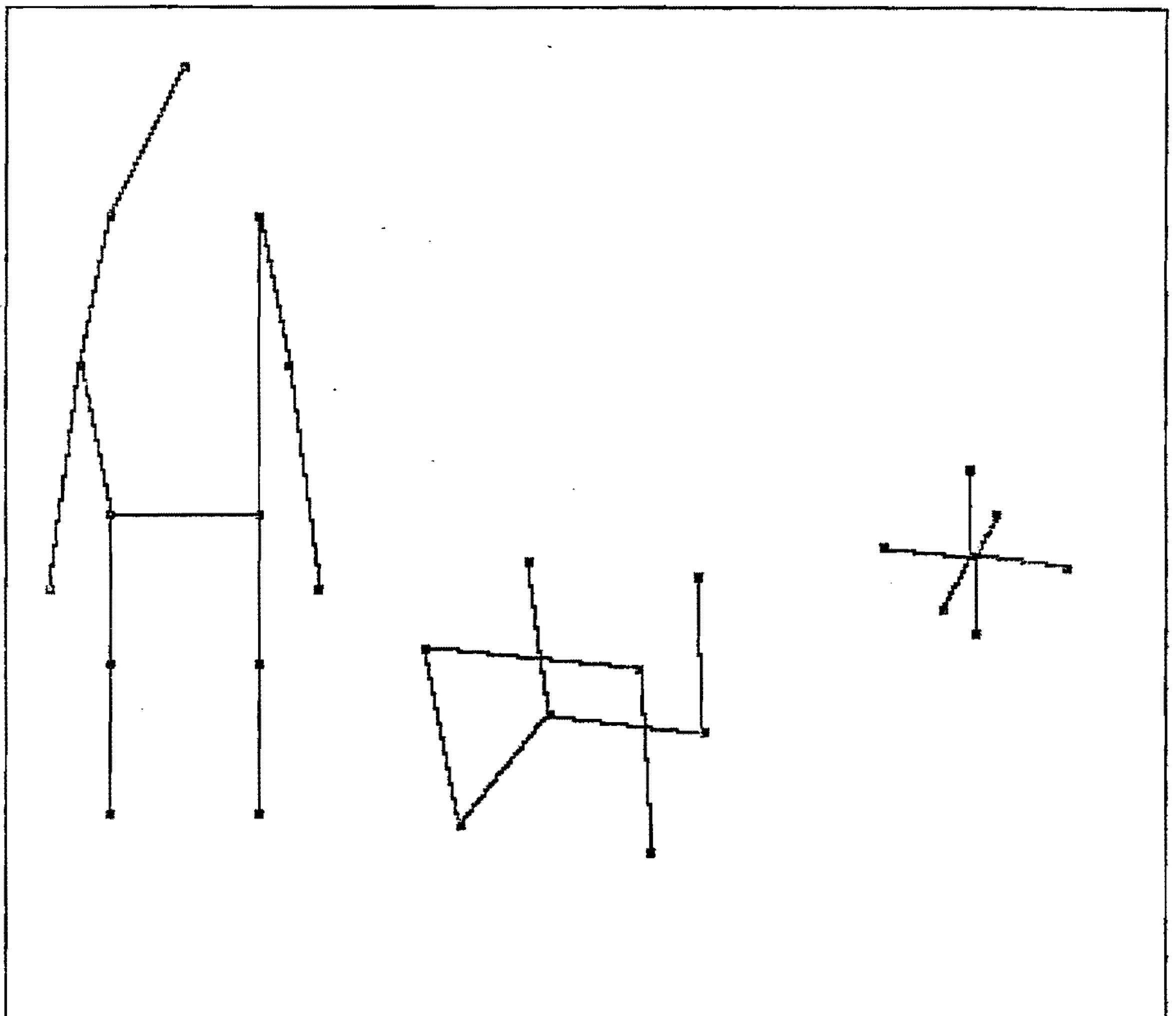


Figure 4
MST drawn on frame data of two men walking along intersecting paths



Frame 1
Frame 2
Frame 3
Frame 4
Frame 5
Frame 6
Frame 7
Frame 8
Frame 9
Frame 10
Frame 11
Frame 12
Frame 13
Frame 14
Frame 15
Frame 16

Figure 5
"Mannequin on lazy susan" - Sample Frames



MST: Constant.Lights, Cube.Lights, Jack.Lights

Figure 6

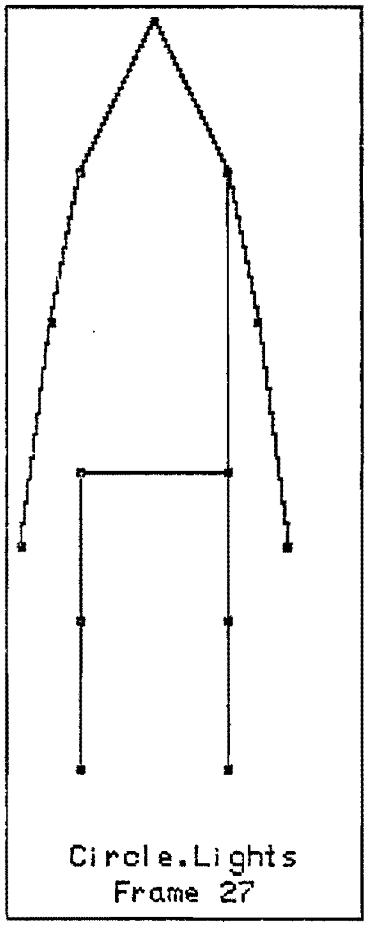


Figure 7

MST

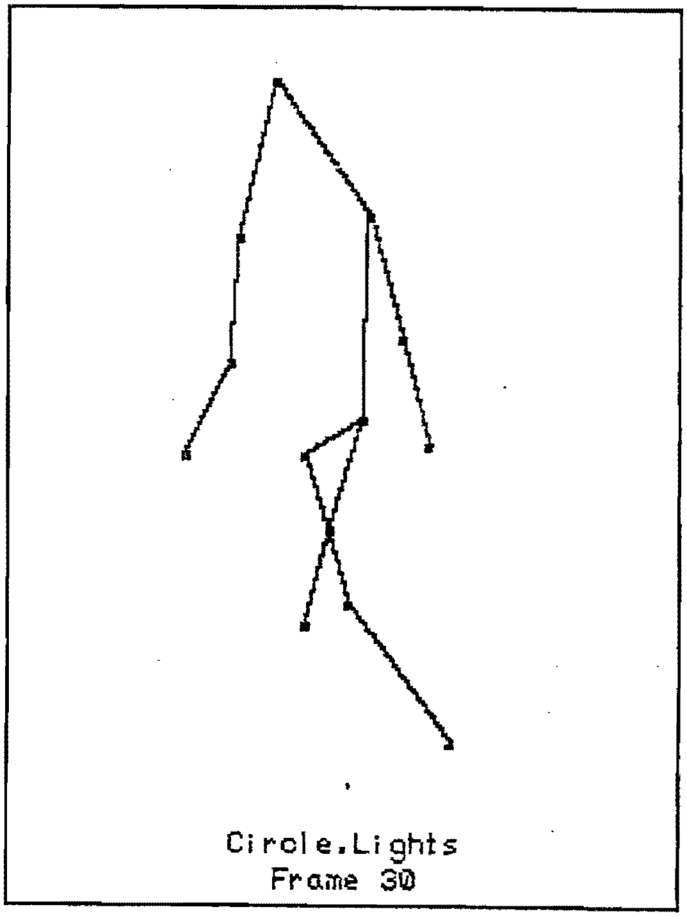


Figure 8

MST displayed on frame data

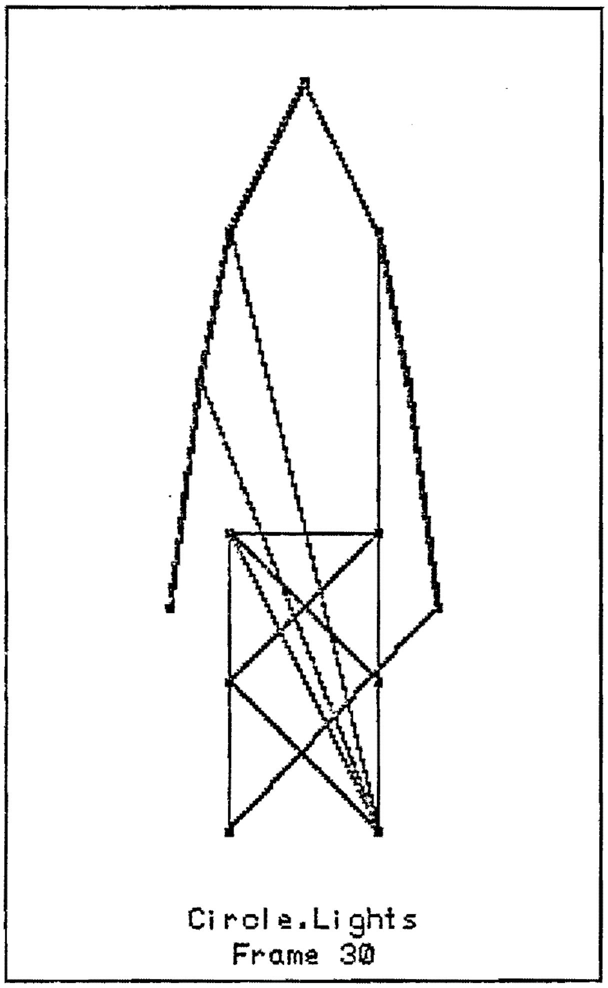


Figure 9

All edges within 30% of minimum
distance for a node are shown

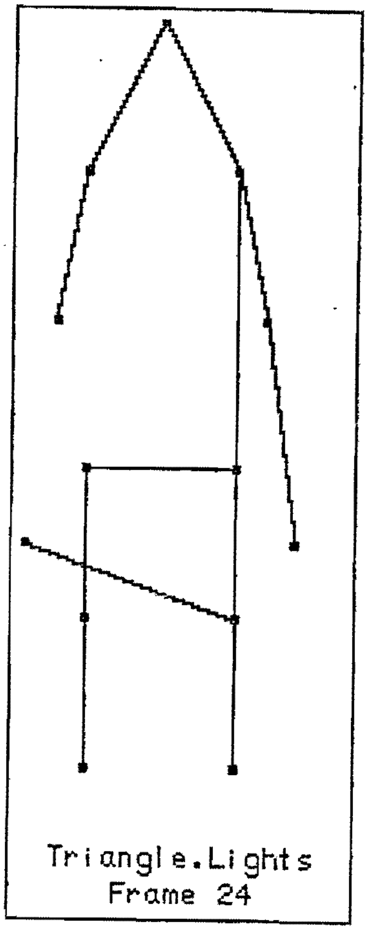


Figure 10

MST