

Using a spatiotemporal plane to recover a moving object's shape using Spatiotemporal Boundary Formation

Douglas W. Cunningham
BTU Cottbus-Senftenberg
Konrad-Wachsmann-Allee 5,
03046 Cottbus, Germany
Douglas.Cunningham@b-tu.de

Mikhail Ashkerov
BTU Cottbus-Senftenberg
Konrad-Wachsmann-Allee 5,
03046 Cottbus, Germany
Mikhail.Ashkerov@b-tu.de

ABSTRACT

When ever an object moves, it successively covers and uncovers surfaces that are farther away. This occlusion and dis-occlusion always occurs precisely at the boundaries of the moving object and as such provide information not only about the shape of the object but also about its velocity, transparency, and relative depth. Humans can and do use this information, and the process has come to be called Spatiotemporal Boundary Formation (SBF). Previous authors have used the wealth of experimental investigations into SBF to create a mathematical model of the process. In this article we proposed a novel method to recover the orientation and velocity the local edge segments of the moving objects which is more flexible, more robust, more compact, and allows the recovery of edges that do not have a constant velocity. The method can be used in object segmentation algorithms or as a pre-filter for machine-learning-based recognition algorithms in order to improve the overall result.

Keywords

Motion perception, Dynamic occlusion, SBF

1 INTRODUCTION

The human visual system is amazingly flexible and powerful. It is not surprising, then, that many computer vision algorithms strive to mimic human capabilities in order to automatically accomplish a wide variety of importance, practical applications in engineering. Many of the human visual system's abilities require that the visual scene first be segmented into component objects. Although most studies in both psychology and computer science focus on the information available from static images, there is a considerable amount of information available over time for object segmentation.

One of the most promising human abilities for object segmentation using dynamic information is Spatiotemporal Boundary Formation (SBF) [1], which is based on the dynamic occlusion of texture caused by the relative motion of two overlapping surfaces. Since our everyday environment is generally cluttered [2], the motion of any given object will successively cover and uncover the background. Interestingly, the background texture will always disappear at the front of the moving objects and will reappear at the back of the moving object. In fact, the texture will appear and disappear precisely at the edges of the object, and therefore provide direct information about many different object properties.

Several researchers have shown that humans can use dynamic occlusion to segment an object from its background, as well as perceive the object's exact shape, velocity, relative depth, and transparency [3, 1, 4, 5, 6,

7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]. A review of this and related phenomena can be found in [21]. In addition to the examination of dynamic boundaries in humans, several researchers have examined the impact of dynamic figure information on the control of fly behavior[22, 23].

One traditional demonstration of dynamically defined edges comes from Michotte's classic tunneling effect [24, 25], where a luminance defined surface is progressively deleted (see Figure 1 for a sketch). As the black circle is displaced upwards, less of it is visible. Perceptually, this is often seen as a complete circle that is being progressively hidden by something that is the same color as the background. While this effect is most convincing in its traditional dynamic form, it can even be seen, to some degree, statically.

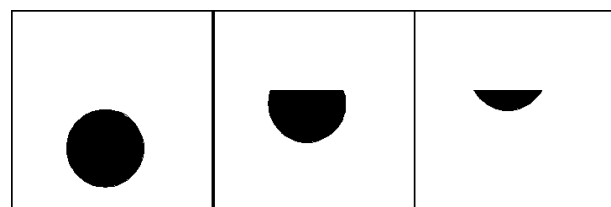


Figure 1: Three frames illustrating Michotte's tunneling effect.

The dynamic transformation of the statically visible surface provides information about both its continued existence and the presence of an occluding form. Thus, two surfaces are seen, one of which has the same color

as the background and is slowly covering the other one. Gibson and colleagues [2] asserted that the changes over time at the boundaries of an object are sufficient to define the shape of those boundaries. Gibson's (1968) film [26], constructed to support this claim, depicts a square with a speckled texture moving over a similarly textured surface. Since there is no static information for the form or its boundaries (e.g., no global difference in luminance, hue, texture, or depth between the inside and outside of the square), neither the surface nor the boundaries are visible when the figure is stationary. When the figure moves, however, a well-defined surface bounded by clear edges is seen moving in front of another surface with similar texture. The displays is reminiscent of naturally camouflaged animals, which can become nearly invisible when stationary in front of the proper background but are nearly always visible when moving. The edges seen in Gibson's displays, as well those seen in Michotte's demonstration, have a phenomenal quality similar to the edges of illusory figures such as the Kanizsa triangle (see Figure 2), they are strong, sharp, and clearly visible just as normal, luminance defined edges [27].

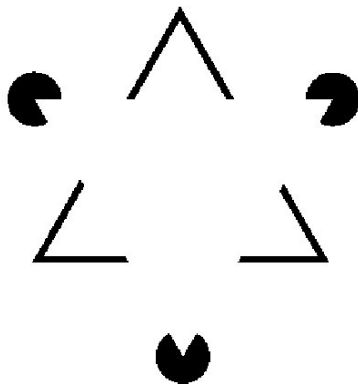


Figure 2: The Kanizsa triangle.

Shipley and Kellman [28] provided a mathematical proof that when the spatial and temporal location of at least 3 elements transformations (e.g., occluded or dis-occluded background elements) are known, we can recover the orientation and velocity of the edge that transformed those elements. Briefly, the spatial and temporal separation between pairs of texture element transformations is represented by local motion vectors. Vector subtraction of the local motion vectors resulting from three non-(spatially)-collinear texture elements defines the orientation of the edge that transformed them. Cunningham and colleagues subsequently modified and extended this model to use a relative encoding scheme (avoiding the explicit requirement for motion vectors) and then modeled the extraction of the occluding form's global shape and its velocity (a preliminary version can be found in [13]). Computational simulations of the model show that it captures the major psychophysical aspects of SBF, including a

dependency on the spatiotemporal density of element changes and a sensitivity to spurious changes [11].

Most modern object recognition tasks either require that (or strongly benefit when) the input images are segmented into discrete objects. This is especially true of machine learning algorithms in general and neural networks in particular. Thus, we propose that prefiltering of the input data with SBF can not only segment and make explicit the information that the object recognition algorithms need (i.e., objects) but also significantly decrease the dimensionality of the input and thus will make machine learning algorithms more robust.

2 MATERIAL AND METHODS

The starting point of our algorithm is the same as the previous models of SBF [28]. Specifically, the dataset consists of the set of (background) texture elements $p_i = (x_i, y_i, t_i)$ that have been covered or uncovered by a moving shape. The spatial location of the element (x_i, y_i) is a projection of the elements true three dimensional location onto the observer's fronto-parallel plane (e.g., a camera's sensor array or a human's retina). The time t_i is the time when the point was transformed (either removed or revealed). As with the previous works, we will begin with the following assumptions, derived from human psychophysical experiments:

1. The local edge can be approximated with a straight line.
2. The velocity of the line is constant.
3. The line does not change orientation.
4. The points are not collinear.

It is critical to note that these assumptions are local in both space and time and are flexible. For example, an circle can be approximated by a series of straight edges, and the smaller the straight edges are, the better the approximation will be. Likewise, the velocity and edge orientation only need to be constant for at most 150 ms [21].

Here, we provide a new approach to extracting the local segment's orientation and velocity. This new approach represents the motion of the moving segment as a plane in space-time, with the occluded points lying on – and defining – this plane. This method is more flexible than previous versions, can be extended to use progressive shape segmentation methods as well as to recover shapes with non-constant velocities. The input to this stage and the output from it will be the same as for the two previous models. This means, among other things, that this step can replace the same stage in existing SBF implementations to recover the full shape and global velocity.

For the sake of simplicity, we will start by defining the algorithm using the smallest possible dataset: three spatiotemporal occlusion events:

$$P = \{p_1, p_2, p_3\} \quad (1)$$

although all the texture appearances and disappearances in the full dataset feed into the SBF algorithm. Indeed, increasing the number of points used to calculate the spatiotemporal edge plane will increase the spatial and temporal accuracy of the results [21].

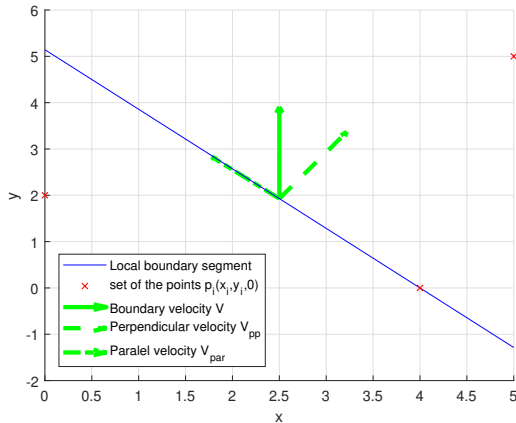


Figure 3: Input data for the method and task definition in projection on XY plane.

Figure 3 depicts the set of points P , the local segment S of the moving edge, and the velocity vector \vec{V} . As stated in assumption 1 above, S is represented as an infinite line. Obviously, the edge is a mere segment, but we cannot at any given time directly see the ends of the line. Indeed, at no time do we even see the line, only its effects (occlusion and dis-occlusion). One consequence of this is that additional information is needed to recover the true length of the segment. We will follow the length calculations presented in [11] (essentially, the points in P are projected along the velocity vector of the edge). Another, more critical, consequence of not being able to see the edge segment is that it is not possible *in principle* to recover the component of the velocity that is parallel to the line. This is because it is not possible to uniquely determine any single point on the line at two different points in time. This problem is called the "aperture problem" and was first described in 1935 by Hans Wallach [29]. Thus, we will decompose the velocity vector \vec{V} of the edge into a vector along the boundary \vec{V}_{\parallel} and one perpendicular to the boundary \vec{V}_{\perp} (see figure 3). Our task is to find the velocity perpendicular to the boundary line \vec{V}_{\perp} and the orientation of the local segment of the boundary. The recovery of at least two different edges is sufficient to restore the true velocity \vec{V} of the object [12].

The first step in recovering the orientation of the edge is to realize that each point in P lies on the edge at the in-

stant that is occluded or revealed. Imagine the simplest case where all three points disappear at the same time. Since the all have the same time $t_i = 0$, all the points will be on the fronto-parallel (or XY) plane at the same time, and as such the edge segment will be uniquely defined (as the connection of the three points) and will also lie on the XY plane. Note that technically these events are "collinear in time", and as such we cannot recover the velocity component.

Using a somewhat more complicated case, let's assume that two of three elements changing state at the same time t_1 , and the third changes at some later time t_2 . We can still plot all three points on the XY plane (that is, show each point where it is on the sensor array, regardless of the time at which it is revealed or deleted). As with the previous case, the local edge segment can be recovered – by definition – as a line connecting the two dots that change together (see Figure 4). As the local edge segment moves, it will occlude the final point at time t_2 and yet will still be parallel to its previous orientation at time t_1 . If we were to project the location (x_i, y_i) of the points that disappear at time t_1 along velocity vector perpendicular to the edge \vec{V}_{\perp} to time t_2 , it is clear that the projected points will still lie on the edge segment. In other words, if we can "discount" the motion of the edge (such as by projecting the position of the element changes along the velocity vector), we can recover the orientation of the edge segment. Without loss of generality, then, we can allow the three points to change at any location in space and time and they will still lie on a unique plane in space-time (as long as the points are not collinear). The general case, where the three element changes occur at different points in time is shown in Figure 5.

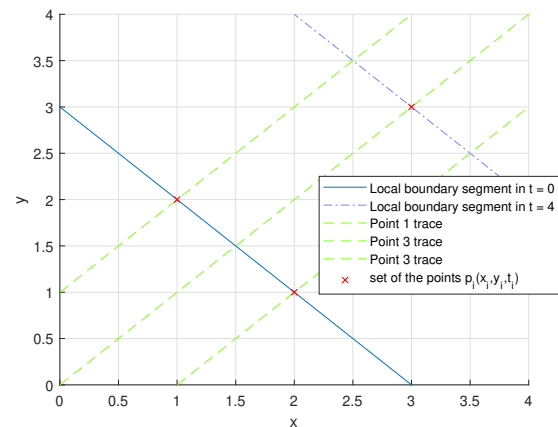


Figure 4: Case with simplifications.

In the next step, we place set of points in a space-time plot by plotting the time t along the z axis (and thus orthogonal to the XY plane). Since the edge will have a spatial extent at any given instant, it clearly must be parallel to the XY plane at every given instant. Like-

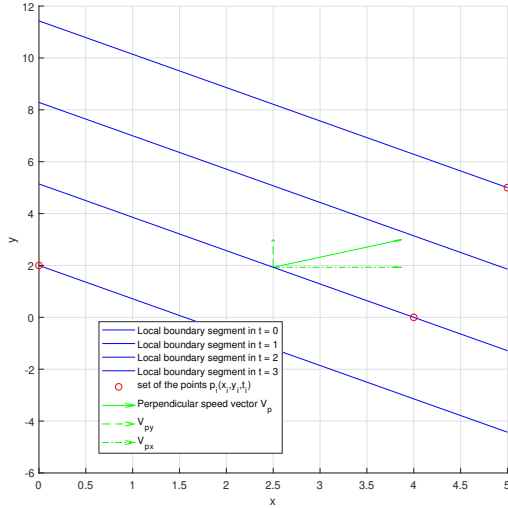


Figure 5: Representation of local borders in time.

wise, the edge segment will intersect each point p_i at the correct time t_i . Thus, we can construct through these 3 space-time points a plane P_{set} (from the definition of the plane it can be done uniquely) and the edge segment will lie on this plane. In fact, the edge segment will be the intersection of the 3D space-time plane and the XY plane for any given time t_i (see Figure 6).

To recover the edge orientation, we first write the equation of the plane in the canonical form 2

$$Ax + By + Ct + D = 0 \quad (2)$$

If we solve plane equation with our set of points 1 we obtain the coefficients A, B, C, D for constructing the plane 2 in 3D space. The line of intersection of this plane with the time plane corresponds to the local boundary at the time t_i and its projection onto the XY plane corresponds to the orientation of the segment in two-dimensional space.

Changing the time and finding the intersection of the new P_{time} we get a family of parallel lines on the plane XY that correspond to the local boundary at different times, see (Figure 5).

In the final step, we use the recovered orientation of the line at different times to extract the perpendicular component of the velocity vector \overline{V}_\perp . Geometrically, the perpendicular velocity will be equal to the slope of the plane in the direction of the perpendicular to the local segment. This can easily be calculated from the equation of the plane and the local segment. For convenience, we decompose the speed \overline{V}_\perp to the velocity along the axis X and the axis Y:

$$\overline{V}_\perp = \left\{ \frac{-AC}{A^2 + B^2}; \frac{-BC}{A^2 + B^2} \right\} \quad (3)$$

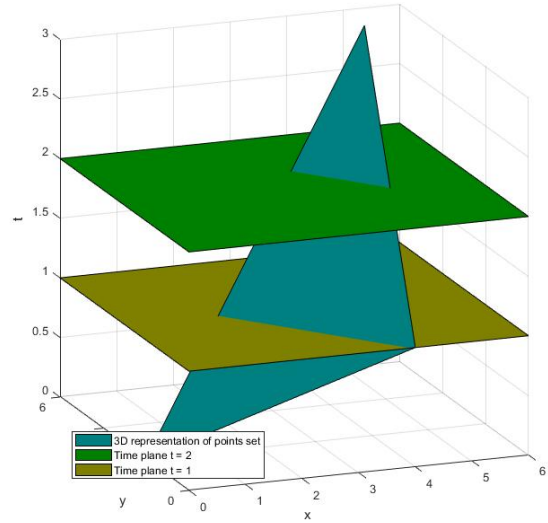


Figure 6: 3D representation of local border.

Numerically, this method gives us exactly the same solution as the vector and trigonometric methods. Since it takes the same input and provides the same form of output, it can be inserted in the an existing SBF pipeline to recover the gloval shape and velocity. One of the main advantages of this new approach is we can change P_{set} to a quadratic function. We will still be able to recover the orientation and speed, but no longer need to assumption a constant velocity.

3 RESULTS

To test the quality of the edge reconstruction, we created a test film similar to that used in previous works. The film was 300 frames long (at 640x480 pixels) and consisted of a black rhombus moving over a field of white dots on a black background. At no given instant can the edges of the rhombus be seen (i.e., the figure is not defined in the static luminance domain). The Rombus moved from left to the right with a constant velocity. In the figure 7 you can see frames from the film and the object shape recovered with our method. Notice that since there is no surface texture and nothing in the luminance domain actually moves, optic flow algorithms will have extreme difficulties with this display. When compared to the results of previous work (such as [11]), we see that the current method produces slightly cleaner shapes, but has the same problems in corners (see Figure 8).

In sum, we present a method for representing the shape and velocity of a moving shape with a space-time plane. The plane is recovered from the position in space and time of background elements transformed by the edges of the moving shape. The new method produces results that are more accurate than previous models and

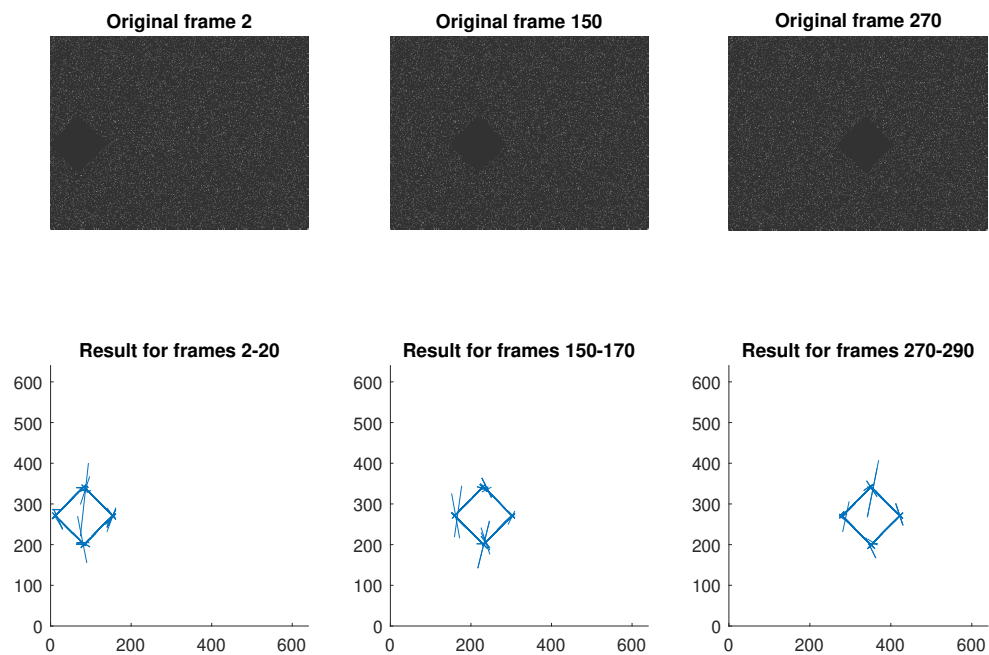


Figure 7: Reconstruction of movement.

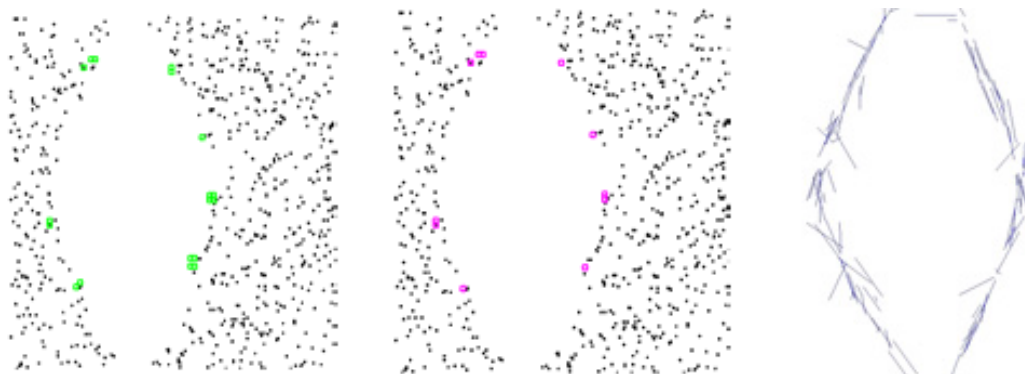


Figure 8: Reconstruction of object shape from poster T.Cooke [11].

provides a few additional advantages. The new approach produces an infinite family of lines, with line corresponding to an instant in time t from the interval $t \in [t_1 \cdot t_3]$. If we take the time outside of the specified area, we get a linear extrapolation of the boundary motion. This form of the solution is convenient for cluster analysis, allowing us to find similar segments of the object. Likewise, the method is convenient for applying machine learning on sets of initial points, which would improve their grouping in a segment. Finally, the method is easy to modify to remove the assumption that the segment velocity is linear.

4 ACKNOWLEDGEMENTS

The authors would like to thank Stefan Guthe for his suggestions and valuable comments.

5 REFERENCES

- [1] T. F. Shipley and P. J. Kellman, "Optical tearing in spatiotemporal boundary formation: When do local element motions produce boundaries, form, and global motion?," *Spatial Vision*, vol. 7, no. 4, pp. 323–339, 1993.
- [2] J. J. Gibson, *The ecological approach to visual perception*. Psychology Press, 2013.
- [3] G. Erlikhman and P. J. Kellman, "Modeling spatiotemporal boundary formation," *Vision research*, vol. 126, pp. 131–142, 2016.
- [4] G. J. Andersen and J. M. Cortese, "2-d contour perception resulting from kinetic occlusion," *Perception & Psychophysics*, vol. 46, no. 1, pp. 49–55, 1989.

- [5] D. R. Bradley and K. Lee, "Animated subjective contours," *Attention, Perception, & Psychophysics*, vol. 32, no. 4, pp. 393–395, 1982.
- [6] N. Bruno, M. Bertamini, and F. Domini, "Amodal completion of partly occluded surfaces: Is there a mosaic stage?," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 23, no. 5, p. 1412, 1997.
- [7] N. Bruno and M. Bertamini, "Identifying contours from occlusion events," *Perception & Psychophysics*, vol. 48, no. 4, pp. 331–342, 1990.
- [8] N. Bruno and W. Gerbino, "Illusory figures based on local kinematics," *Perception*, vol. 20, no. 2, pp. 259–274, 1991.
- [9] C. M. Cicerone and D. D. Hoffman, "Color from motion: dichoptic activation and a possible role in breaking camouflage," *Perception*, vol. 26, no. 11, pp. 1367–1380, 1997.
- [10] C. M. Cicerone, D. D. Hoffman, P. D. Gowdy, and J. S. Kim, "The perception of color from motion," *Perception & Psychophysics*, vol. 57, no. 6, pp. 761–777, 1995.
- [11] T. Cooke, D. W. Cunningham, C. Wallraven, and H. H. Bülthoff, "Local processing in spatiotemporal boundary formation," in *Proceedings of the 7th Tübingen Perception Conference*, p. 65, Knirsch Verlag Kirchentellinsfurt, 2004.
- [12] D. W. Cunningham, T. F. Shipley, and P. J. Kellman, "Interactions between spatial and spatiotemporal information in spatiotemporal boundary formation," *Perception & psychophysics*, vol. 60, no. 5, pp. 839–851, 1998.
- [13] D. Cunningham, A. Graf, and H. Bülthoff, "A relative encoding model of spatiotemporal boundary formation," in *5. Tübinger Wahrnehmungskonferenz (TWK 2002)*, 2002.
- [14] T. Hine, "Subjective contours produced purely by dynamic occlusion of sparse-points array," *Bulletin of the Psychonomic Society*, vol. 25, no. 3, pp. 182–184, 1987.
- [15] P. J. Kellman and M. H. Cohen, "Kinetic subjective contours," *Perception & Psychophysics*, vol. 35, no. 3, pp. 237–244, 1984.
- [16] K. Prazdny, "Illusory contours from inducers defined solely by spatiotemporal correlation," *Perception & Psychophysics*, vol. 39, no. 3, pp. 175–178, 1986.
- [17] I. Rock and F. Halper, "Form perception without a retinal image," *The American journal of psychology*, vol. 82, no. 4, pp. 425–440, 1969.
- [18] T. F. Shipley and P. J. Kellman, "Spatiotemporal boundary formation: Boundary, form, and motion perception from transformations of surface elements.," *Journal of Experimental Psychology: General*, vol. 123, no. 1, p. 3, 1994.
- [19] T. F. Shipley and P. J. Kellman, "Spatio-temporal boundary formation: The role of local motion signals in boundary perception," *Vision Research*, vol. 37, no. 10, pp. 1281–1293, 1997.
- [20] P. Stappers, "Forms can be recognized from dynamic occlusion alone," *Perceptual and Motor Skills*, vol. 68, no. 1, pp. 243–251, 1989.
- [21] T. F. Shipley and D. W. Cunningham, "Perception of occluding and occluded objects over time: Spatiotemporal segmentation and unit formation.," 2001.
- [22] H. Bülthoff, "Figure-ground discrimination in the visual system of *Drosophila melanogaster*," *Biological Cybernetics*, vol. 41, no. 2, pp. 139–145, 1981.
- [23] W. Reichardt and T. Poggio, "Figure-ground discrimination by relative movement in the visual system of the fly," *Biological Cybernetics*, vol. 35, no. 2, pp. 81–100, 1979.
- [24] A. Michotte, *On phenomenal permanence: facts and theories*, vol. 7. Acta Psychologica, 1950.
- [25] A. Michotte, G. Thine, G. Crabbé, et al., *Les compléments amodaux des structures perceptives*. Institut de psychologie de l'Université de Louvain, 1964.
- [26] J. Gibson, "The change from visible to invisible: A study of optical transitions [film]," *State College, PA: Psychological cinema register*, 1968.
- [27] G. Kanizsa, *Organization in vision: Essays on Gestalt perception*. Praeger Publishers, 1979.
- [28] T. F. Shipley and P. J. Kellman, "Spatio-temporal boundary formation: the role of local motion signals in boundary perception," *Vision Research*, vol. 37, no. 10, pp. 1281 – 1293, 1997.
- [29] H. Wallach, "Über visuell wahrgenommene Bewegungsrichtung.," *Psychologische Forschung*, pp. 325–380, 1935.