Interpretation of Motion Trajectories Using Focus of Expansion

Krishnan Rangarajan and Mubarak Shah

Abstract—The focus of expansion (FOE) of a group of motion trajectories is defined to be a point in the image plane at which the trajectories intersect when they are extended. The FOE observed over a time sequence defines the locus of FOE. We present an analytical approach for the study of dynamic events as they project on the image plane by analyzing the locus of FOE. We have found that the locus of FOE can be used to make qualitative assertions regarding the type of motion. An interesting behavior of the locus of FOE for various types of motion is observed. The cases include a single point, a horizontal, a vertical, and a slanted straight line. We can also determine whether the object has approaching and receding motion or when the object changes its direction of motion. This inference may be used in qualitative computer vision.

Index Terms—Dynamic scene analysis, focus of expansion, motion trajectories.

I. INTRODUCTION

Given n frames taken at different time instants and m points in each frame, the motion correspondence establishes a mapping of a point in one frame to another point in the next frame. This correspondence can be used to generate a path followed by a point lying on an object. A path can be generated by starting from a point in the first frame and ending at some point in the last frame, touching each frame at not more than one point, and by joining a point in a frame by a straight line with its corresponding point in the next frame. We call such a path a trajectory. Each trajectory is identified by a point in the first frame. A set of nonintersecting paths, which together involve all points in all frames, is a trajectory set.

Under perspective projection, lines that are parallel in space may not remain parallel in the image plane. With the camera pointing in the Z axis,1 lines parallel in space and having components in the Z axis do not project to be parallel lines in the image plane. These lines meet at a common point in the x-y plane known as the vanishing point. Lines that are parallel in space and that do not have a component in the Z axis project to the parallel lines in the image plane, and hence, their vanishing point is at infinity.

The motion trajectories are representative of the motion of the underlying points. If we assume that the image plane x-y is parallel to the X-Y plane, the trajectories that have a nonzero motion component in the Z direction and are parallel in space will exhibit the same vanishing point. The trajectories of points lying on an object undergoing translation remain parallel in space. Hence, depending on the vanishing point, we can segment the trajectories belonging to the same object. Fig. 1 shows two cubes moving with different velocities in space. The vanishing point of trajectories belonging to one cube

1 We will be using uppercase X, Y, and Z to denote the world coordinates, whereas we will use lowercase x and y to denote the image coordinates.

is point P (=100, -100), and the vanishing point of trajectories of the other cube is point Q (=300, -150). If the motions of objects are assumed to be independent, each vanishing point will represent one object.

The focus of expansion (FOE) is very similar to the vanishing point of the motion trajectories. The FOE is defined to be a point in the image from which, for a given forward direction of translating motion and direction of gaze, all image features seem to diverge radially. The FOE can be used to determine, for example, the direction of vehicle heading. Since the FOE has previously been used in motion research, we will also continue to use the FOE for the vanishing point of motion trajectories. Previous related research has been limited to computation of the FOE. In this correspondence, we will assume that there exists a reasonably good method for computing the FOE, and we will focus on the use of the FOE for interpretation of motion trajectories. Precisely, we will be dealing with an extended sequence of frames, which gives rise to a sequence of trajectories between two frames and, hence, a sequence of FOE's. We term this sequence of FOE's as the locus of FOE. The aim of this correspondence is to show that the locus of the FOE is closely related to the motion of the underlying object.

We present an analytical approach for the study of dynamic events as they project on the image plane by analyzing the locus of the FOE. We have found that the locus of the FOE can be used to make qualitative assertions regarding the type of motion. We can also determine whether the object has approaching and receding motion or when the object changes its direction of motion. This inference may be used in qualitative computer vision.

The organization of the rest of the correspondence is as follows. The next section deals with a survey of methods for computing the FOE. The projection model that will be used throughout this correspondence is described in Section III. In Section IV, we analyze the locus of the FOE for various types of motion. Finally, Section V deals with interpretation of motion trajectories using the locus of the FOE.
II. RELATED WORK

Earlier researchers have developed methods for computing the FOE. The FOE is defined only for pure translation. The work on FOE can be broadly categorized into two categories based on the density of the optical flow used: dense flow methods and sparse flow methods. The methods proposed by Ballard et al. [3], Negaharian and Horn [9], Prazdny [10], Dutta et al. [5], and Burger et al. [4] are dense flow methods, whereas the methods proposed by Jain [6] and Lawton [7] are sparse flow methods based on a small number of point correspondences. These methods make use of the Hough transform or employ an error surface for a restricted region in the image plane and estimate the FOE by hill climbing. Each of these categories can again be divided into subcategories based on whether they address a multiple independently moving objects or not.

Ballard et al. [3] use the fact that all the flow vectors converge at the FOE, and they estimate the FOE by a Hough formulation. Collinear flow vectors are detected by considering the \((r, \theta)\) space. A point \((x, y)\) in the image that has a flow vector \((u, v)\) votes for \((r, \theta)\), where \(r = x \cos \theta + y \sin \theta\), and \(tan \theta = \frac{v}{u}\). The points in \((r, \theta)\) space of radial flow lines form a circle, each circle corresponds to one rigid object, and each circle is associated with a FOE. From \((r, \theta)\) space, the FOE is identified by the Hough method. Each \(r, \theta\) votes for cells \((a, b)\), obeying the constraint \(\frac{a}{b} = a \cos \theta + b \sin \theta\). The FOE is given by \((2a, 2b)\). They concentrate on scenes having a single moving object, but they suggest how it can be expanded to scenes with multiple moving objects.

Dutta et al. [5] address determination of the FOE under ego motion. Pure translation of the camera is idealistic, and in the sequences of real scenes, there is always a small rotation of the camera, which introduces error into the FOE determination. They consider the error due to rotation as a systematic error, develop expressions for the error, and undo the rotation. They show that error in the FOE estimation due to rotation is proportional to the depth of the points whose flow vectors are considered. The FOE estimated by using large flow vectors of the nearby points is closer to the actual FOE than the one found by using small flow vectors belonging to distant points. They verify their claims by comparing their method with the one that uses Anandan’s method [2] for estimating the flow vectors and Advic’s method [1] for estimating motion parameters and, hence, the FOE. The method has been designed for sensor motion as in autonomous navigation and is not suitable for scenes with multiple independently moving objects.

The method proposed by Burger et al. [4] also takes a similar approach to estimating the rotation and correcting it to find the correct FOE. They define a fuzzy FOE that has an associated area around it and assume that the FOE falls inside it. This area is grown by a connected component algorithm around a point. This method has also been designed for sensor motion as in autonomous navigation and is not suitable for scenes with multiple independently moving objects.

Jain [6] has proposed a method for estimating the FOE without optical flow. He uses a simple function based on the geometry that, when maximized, leads to the FOE. The function that is maximized is the sum of the distance from an arbitrary point to the points in the two consecutive frames—the sum of the distances between points in frames. This is a continuous function, and a gradient method is used to find the minimum. There is no direct extension of this method to scenes having multiple independently moving objects.

Lawton has proposed a method [7] that can estimate the FOE given a sparse flow field. The main difference between this method and that of the method proposed by Jain is that it combines the token identification with the FOE estimation. The method, as it is, is applicable to images generated by moving sensor, and it cannot be extended to scenes with multiple independently moving objects in a straightforward manner.

There has been some work done in finding parallel lines in a perspective image using vanishing points. There is a similarity between vanishing points and the FOE. The vanishing point is the point of intersection of line segments in the scene that are the projection of parallel lines in 3-D space, and the FOE is the point of intersection of flow vectors in the image that are the projection of parallel 3-D flow vectors. Magee and Aggarwal [8] have proposed a method for finding the vanishing point. They consider end points of lines in the image and fit a plane with the focal point. The line of intersection between two such planes points toward the vanishing point of the two lines in the image. They map the unit vector of this line of intersection onto a point on the surface of the Gaussian sphere and cluster the points on the Gaussian surface based on distance. The lines corresponding to these clustered points are parallel in 3-D space and have a common vanishing point in the image plane.

Our correspondence deals with the analysis and use of the FOE, and we assume there exists some reasonably good algorithm for computing the FOE. The experiments reported in the correspondence use the FOE computed by finding the intersection of two straight line segments of trajectories analytically.

III. PROJECTION MODEL

Let us assume that the origin of the world coordinate system coincides with the origin of the image coordinate system. In addition, let \((0, 0, 0)\) be the position of the camera in the world coordinates. Let \((X_i^j, Y_i^j, Z_i^j)\) be the world coordinates of point \(i\) at time \(t_j\), and the image plane coordinates \((x_i^j, y_i^j)\) of point \(i\) at time \(t_j\) under perspective projection are given by

\[
x_i^j = -\frac{f.X_i^j}{Z_i^j}\\
y_i^j = -\frac{f.Y_i^j}{Z_i^j}
\]

where \(f\) is camera focal length. The FOE of trajectories belonging to an object is the point of intersection between the 2-D projection of the trajectories in the image plane. It is assumed that when the time interval between subsequent frames is small, these trajectories can be considered to be straight lines. Let \((\phi_i^j, \theta_i^j)\) denote the FOE of trajectories belonging to the same object between times \(t_j\) and \(t_{j+1}\). Let \(v_i^j, v_{\phi_i}^j,\) and \(v_{\theta_i}^j\) denote the velocity of the object in the \(X, Y,\) and \(Z\) directions, respectively, in this time interval. The FOE of an object between two time frames is determined by finding the point of intersection of the projection of a pair of 2-D trajectories of points belonging to this object and is given by

\[
\phi_{\phi_i}^j = \frac{-f.v_{\phi_i}^j}{v_{\theta_i}^j} \\
\phi_{\theta_i}^j = \frac{-f.v_{\theta_i}^j}{v_{\phi_i}^j}
\]

These equations are derived in the Appendix. These equations imply that the FOE is the perspective projection of the point in space whose coordinates are given by \((v_{\phi_i}^j, v_{\theta_i}^j, v_{\phi_i}^j)\).

IV. LOCUS OF FOE

For a sequence of frames \(1\) to \(n\), the world coordinates of any point at time \(t_{j+1}\) belonging to an object is related to its coordinates at time \(t_j\) by its motion parameters at time \(t_j\). The path traced by \((\phi_i^j, \theta_i^j)\), which is a FOE of the object, as we vary \(j\) from \(1\) to \(n - 1\) defines the locus of the FOE of the object. The type of motion and the nature of motion like constant velocity, constant acceleration,
Fig. 2. (a) Sector $POQ$ is rotating about $O$ in the $YZ$ plane; (b) velocity profile along $OP$. The points on $OP$ are at different radii from the center of rotation $O$, but are in phase with each other; (c) velocity profile along $OQ$; (d) velocity profile along arc $PQ$, where the radius remains constant but phase angle is changing.

Fig. 3. (a) Isometric view of 3-D trajectories of a line with end points at $(20,20,30)$ and $(30,20,30)$ under translation with velocity components $v_x = 3$, $v_y = 4$, and $v_z = 5$; (b) perspective projection of trajectories in (a); (c) locus of FOE, which is a single point.

eetc. affect the locus of the FOE. In fact, the locus of FOE is the perspective projection of the space curve given by the equations

$$
x = v_x t,
\quad y = v_y t,
\quad z = v_z t.
$$

In this section, we study the locus of the FOE under pure translation. For other types of motion, the FOE is not well defined.

For instance, under rotation, not all points belonging to the same object do not necessarily have the same FOE. Fig. 2(a) shows a sector $OPQ$ rotating about $O$ in the $YZ$ plane. Points $P$ and $Q$ have the same phase angle but different radii of rotation $OR$ and $OP$, respectively, whereas points $R$ and $Q$ have a phase difference and different radii of rotation. The velocity profile along $OP$, where the radius alone varies but not phase angle, is shown in Fig. 2(b). The velocity profile along $OQ$ is shown in Fig. 2(c), and the velocity profile along arc $PQ$, where the radius remains constant but the phase angle varies, is shown in Fig. 2(d). The points on the same object will have the same FOE if and only if they are located at the same distance from the axis of rotation and they are in phase with each other.

The results in this section are reported with the aid of figures (Figs. 4–6) that have three parts. In (a), we plot the 3-D trajectories of the points in motion as in an isometric view; in (b), we plot the 2-D trajectories traced out in the image plane by the 3-D trajectories; and finally, in (c), we plot the locus of the FOE.

In this subsection, we will consider various cases related to translation.

1. When an object translates with a uniform velocity, the locus of the FOE is a point, that is, the FOE remains stationary. This follows from (1) and (2) as $v_x^j$, $v_y^j$, and $v_z^j$ remain constant for all $j$. Fig. 3 shows an instance of this case.

2. If the object translates in $X$, $Y$, and $Z$ but $X$ alone has a constant acceleration $a_x$, the velocities at time instant $j$ can be written as

$$
v_x^j = v_x^j + (j - 1)a_x,
\quad v_y^j = v_y^j,
\quad v_z^j = v_z^j.
$$

Since we are assuming acceleration to be constant, the locus of the FOE is given by the equation

$$
\theta_x^j = -f \frac{v_x^j}{v_z^j} + (j - 1)a_x,
\theta_y^j = -f \frac{v_y^j}{v_z^j}.
$$

The above equations represent a line parallel to the $x$ axis. Fig. 4 shows an instance of this case.

3. If a rigid line translates in $X$, $Y$, and $Z$ but $Y$ alone has acceleration, which is constant $a_y$, then the velocities at time instant $j$ can be written as

$$
v_x^j = v_x^j,
\quad v_y^j = v_y^j + (j - 1)a_y,
\quad v_z^j = v_z^j.
$$

Since we are assuming acceleration to be constant, the locus of the FOE is given by the equations

$$
\theta_x^j = -f \frac{v_x^j}{v_z^j},
\theta_y^j = -f \frac{v_y^j}{v_z^j} + (j - 1)a_y.
$$

The above equations represent a line parallel to the $y$ axis. Fig. 5 shows an instance of this case.
4. If a rigid object translates in $X$, $Y$, and $Z$ but $Z$ alone has constant acceleration, then the velocities at time instant $j$ can be written to be

$$v_x^j = v_x^1 + (j - 1) a_x$$
$$v_y^j = v_y^1 + (j - 1) a_y$$
$$v_z^j = v_z^1 + (j - 1) a_z$$

Since we are assuming acceleration $a_z$ to be constant, the locus of the FOE is given by the equations

$$\theta_x^j = -\frac{v_x^1}{v_x^1 + (j - 1) a_x}$$
$$\theta_y^j = -\frac{v_y^1}{v_y^1 + (j - 1) a_y}$$

The above equations represent an inclined line with slope $\frac{v_x^1}{v_y^1}$. Fig. 6 shows an instance of this case.

5. When an object translates in $X$, $Y$, and $Z$ with constant acceleration of $a_x, a_y, a_z$ in $X$, $Y$ and $Z$, the velocities at time instant $j$ can be written to be

$$v_x^j = v_x^1 + (j - 1) a_x$$
$$v_y^j = v_y^1 + (j - 1) a_y$$
$$v_z^j = v_z^1 + (j - 1) a_z$$

Since we are assuming acceleration to be constant, the locus of the FOE is given by the equations

$$\theta_x^j = -\frac{v_x^1 + (j - 1) a_x}{v_x^1 + (j - 1) a_y}$$
$$\theta_y^j = -\frac{v_y^1 + (j - 1) a_y}{v_y^1 + (j - 1) a_z}$$

The above equations represent an inclined line with slope $\frac{v_x^1 + (j - 1) a_x}{v_y^1 + (j - 1) a_y}$. From these equations, we can see that the locus of the FOE is an inclined line with slope $\frac{\dot{\theta}_x}{\dot{\theta}_y}$ when $a_z = 0$. When $a_x = 0, a_Y = 0$, and $a_z \neq 0$, the slope of this line is $\frac{\dot{v}_y}{\dot{v}_x}$, as in item 4 above.

V. INTERPRETATION OF LOCUS OF FOE

In this section, we will summarize the observations regarding the interpretation of locus of the FOE under various types of motion. From the analysis, it will also be shown that the change of direction of motion and receding and approaching motion can be determined from the locus of the FOE.

1. If the locus of the FOE is a single point, then the object is undergoing translation without any acceleration.

2. If the FOE moves along a line, then the following holds:
   a. If the locus is a straight line parallel to the $x$ axis, then the motion is translation with acceleration along the $X$ direction only.
   b. If the locus is a straight line parallel to the $y$ axis, then the motion is translation with acceleration along the $Y$ direction only.
   c. If the locus is an inclined straight line, then the motion is translation with acceleration in the $X$ direction or along two or more of the $X$, $Y$, and $Z$ directions. This includes cases 4 and 5 discussed in Section IV.

3. We can identify the instants at which the object changes its direction of motion along the $Z$ axis. This is because the velocity along the $Z$ direction appears in the denominator of the $X$ and $Y$ coordinates of the FOE. As the velocity along the $Z$ direction changes sign, there is a discontinuity in the plot of $\theta_x$ and $\theta_y$ against time at the same time instant. This fact can be used in identifying instants at which the object changes its direction of motion along the $Z$ direction.

4. We can also identify instants at which the object changes its direction of motion along the $X$ and $Y$ axes. The change in direction of motion along the $X$ axis is seen as a smooth crossover from positive to negative values in the plot of the $X$ coordinates of the FOE against time. This is because $v_x$, which is the velocity along the $X$ direction, appears in the numerator of $\theta_x$. Similarly, the change in direction of motion along the $Y$ axis is seen as a smooth crossover from positive to negative values in the plot of the $y$ coordinates of the FOE against time.

5. We can determine whether an object is approaching or receding at any instant of time using the sign of the FOE and the sign of image coordinates of a point. Fig. 7 shows the $X-Z$ plane and a reference grid at the point of interest at time instant $j$. In each quadrant, a ‘+’ or ‘−’ sign is marked, and it denotes the sign of the FOE if the points move into that quadrant in time instant $j + 1$. These signs can easily be determined by using the fact that $\dot{\theta}_x = -\frac{\dot{v}_x}{v_x}$ for $a_x$. For instance, the point $i$ shown in Fig. 7 will only move to the area $A_1$ if its $v_x$ is negative and $v_z$ is positive; therefore, $\dot{\theta}_x$ will be positive. We will consider the following cases:
   a. If $\dot{\theta}_x > 0$, then the following hold:
Fig. 8. Rigid line with end points at 1, 2 was moved in space, and the $\partial_j$ was computed. Our decision procedure was used in deciding whether point 1 was approaching or receding. In this table, we show six cases listed in column 1. The columns 2-4 and 5-7, respectively, show the world coordinates of point 1 in frames 1 and 2, whereas columns 8 and 9 and 10 and 11 show their corresponding image coordinates. The $\theta_x$ of each case is given in column 12.

VI. CONCLUSION

In this correspondence, we have shown that the FOE’s of motion trajectories carry rich information related to the motion of the objects. The locus of the FOE also helps in identifying instances at which the object changes its direction of motion along any principle axes. Since under translation all points on the same object have the same velocity, their trajectories have the same FOE. This property of the FOE can be used in developing an algorithm for segmenting trajectories into groups belonging to the same object, which will make use of the information available in the whole span of the trajectories.

APPENDIX

In this Appendix, we will derive the expressions for the FOE (1) and (2). Let $(x_1, y_1)$ and $(x_{1+i}, y_{1+i})$ be the trajectory segment of point 1 in frames $j$ and $j+1$, and let $(x_2, y_2), (x_{2+i}, y_{2+i})$ be the trajectory segment of point 2 in frames $j$ and $j+1$. Further, assume that points 1 and 2 lie on the same object, which is translating in space with a velocity of $v_x$, $v_y$, and $v_z$ in the $X$, $Y$, and $Z$ directions, respectively. Let the 3-D coordinates of points 1 and 2 in frame $j$ be $(X_1, Y_1, Z_1)$ and $(X_2, Y_2, Z_2)$, and those of points 1 and 2 in frame $j+1$ be $(X_{1+i}, Y_{1+i}, Z_{1+i})$ and $(X_{2+i}, Y_{2+i}, Z_{2+i})$.

Clearly, the following equations hold:

\[
X_1^{i+1} = X_1 + i Z_1 \\
Y_1^{i+1} = Y_1 + i v_y \\
Z_1^{i+1} = Z_1 + i v_z
\]

In addition, applying the projection equations

\[
x_i = -\frac{fX_i}{Z_i} \\
y_i = -\frac{fY_i}{Z_i} \\
x_{i+1} = -\frac{fX_i + v_x}{Z_i + v_z} \\
y_{i+1} = -\frac{fY_i + v_y}{Z_i + v_z}
\]

The equations of the trajectory segments of points 1 and 2 are given by the following equations:

\[
x - x_i = x_{i+1} - x_i \\
y - y_i = y_{i+1} - y_i \\
x - x_i = x_{i+1} - x_i \\
y - y_i = y_{i+1} - y_i
\]

FOE $(\theta_x, \theta_y)$ is the point of intersection of these lines and is given by

\[
\theta_x = -\frac{f v_x}{v_z} \\
\theta_y = -\frac{f v_y}{v_z}
\]
REFERENCES


I. INTRODUCTION

The statistics of images and image sequences have been extensively studied for image coding and compression applications [1], [2] as well as for the development of models of biological image processing [3], [4]. An exponential autocorrelation function has been shown to be a good model for temporal frame-to-frame correlations of image sequences, e.g., [5]-[8], and for spatial correlations within each frame, e.g., [2], [3], [9].

This paper focuses on the separability of the spatiotemporal statistics of image sequences and on the validity of using a separable exponential autocorrelation model for the spatiotemporal statistics. The autocorrelation function is uniquely related to the power spectrum via a Fourier transform, and either is valid as a description of the statistics.

The spectra of image sequences were calculated. The sequences represented a small ensemble of possible motion activity. The sequences were selected for a range of motion activity. For example, a fast camera pan represents the maximum image motion activity, and a small moving object with a static background represents the least activity. Sequences with motion activity between these extremes had slight camera motion and some object motion.

II. CALCULATION OF IMAGE STATISTICS

We collected 14 image sequences (256 × 256 × 64 @ 8 b/pixel, 30 frames/s with no scene cuts) from a video disc that contained scenes from a broadcast TV source. Each frame was originally sampled at 512 × 512 pixels/screen, but adjacent pixels were averaged, and the image was subsampled to 256 × 256 pixels/screen. The sample mean of each sequence was removed to reduce low-frequency bias in the calculations.

The sample power spectrum \( P(k_1, k_2, f) \) of each sequence \( x(n_1, n_2, t) \) is the squared magnitude of the discrete Fourier transform calculated as

\[
P(k_1, k_2, f) = \frac{1}{256 \cdot 256 \cdot 64} \left( \sum_{n_1=0}^{255} \sum_{n_2=0}^{255} \sum_{t=0}^{63} x(n_1, n_2, t) e^{-j2\pi(k_1n_1 + k_2n_2 + f t)} \right)^2
\]

where \( k_1, k_2, \) and \( f \) are spatial frequencies, \( n_1, n_2 \) are spatial locations, and \( t \) is time measured in frame number.

We converted the two spatial frequency dimensions \( k_1 \) and \( k_2 \) into one radial frequency dimension \( k \) by averaging in 32 annuli around the spatial frequency origin as illustrated in Fig. 1. In this manner, the spatial frequency range of 0-127 cycles/screen of \( k_1 \) and \( k_2 \) is represented by 32 annuli in bands of 4 cycles/screen. Averaging the spatial spectra in annuli is equivalent to assuming a circularly symmetric spatial autocorrelation function. This autocorrelation function is not separable in the two spatial dimensions but is considered a better fit than the corresponding separable autocorrelation function for most images [9].

The average magnitude of the power spectrum in each annulus can be obtained by summing over the power spectrum \( P(k_1, k_2, f) \) in the annulus indexed by \( k \) and normalizing by the number of sample...