Real-time Time-To-Collision from variation of Intrinsic Scale

Amaury Nègre, Christophe Braillon, Jim Crowley, Christian Laugier

▶ To cite this version:

HAL Id: inria-00182028

https://hal.inria.fr/inria-00182028

Submitted on 24 Oct 2007

Real-time Time-To-Collision from variation of Intrinsic Scale

Amaury Nègre[1], Christophe Braillon[1], James L. Crowley[1], and Christian Laugier[1]

INRIA, Grenoble, France
`firstname.lastname@inrialpes.fr`

Summary. Time-to-collision can be directly measured from a spatio-temporal image sequence obtained from an uncalibrated camera. This it would appear to offer a simple, elegant measurement for use in obstacle avoidance. However, previous techniques for computing time to collision from an optical flow have proven impractical for real applications.

This paper present a new approach for computing time to collision (TTC) based on the idea of measuring the rate of change of the "intrinsic scale". Intrinsic scale is a geometric invariant that is valid at most points in an image, and can be rapidly determined using a multi-resolution pyramid.

In this paper we develop the approach and demonstrate its feasibility by comparing the results with range measurements obtained from a laser ranging device on a moving vehicle. Experimental results show that this is a simple method to obtain reliable TTC with a low computational cost.

1 Introduction

This article takes place in the obstacle avoidance context for autonomous vehicle in dynamic environment. In this domain, the estimation of time-to-collision (TTC) is crucial. This variable τ has been defined by Lee [11] as the ratio of the distance z of the object and the relative speed of approach between the robot and an obstacle. Estimation of TTC has also been proposed as a biological model for insect navigation [7], bird flight [8] as well as the human "looming" reflex [17].

Most of vision based methods to compute TTC use the optical flow. In a first method [3], a focus of expansion (FOE) is estimated and τ is computed from the pixel velocity and the FOE. But this method cannot be applied if many objects move independently (several FOE). It has been also shown that τ can be approximated by the inverse of the optical flow divergence [16], and in [18] the authors show that τ can be directly computed from the optical flow and its derivatives when using a log-polar mapping. Unfortunately, those methods require a first step of reliable dense optical flow estimation, which has proved difficult.

In this paper, we propose a simple new approach to direct measurement of time to collision. Our approach is based on the observation that the local "size" of features in

an image may be directly measured from the scale of extremal points in a Laplacian scale space. This invariant provide a simple, robust measurement of local size that can be obtained at most points in an image.

A key problem when using intrinsic scale for time to collision is image registration. We describe such a registration for natural interest points as defined in [13], as well as for natural interest lines [9]. We experimentally measure the precision of TTC obtained from an image sequence taken from a moving vehicle using the distance to targets as provided by a laser scanner.

2 Computation of TTC

2.1 TTC from variation of size

Time-to-collision is easily derived from geometrical considerations : let P represents an object in front of a camera such that P travels toward the camera. We note F the focal length of the camera, S the real size of P, s the size of P in the image and Z the distance of the object. (see fig. 1).
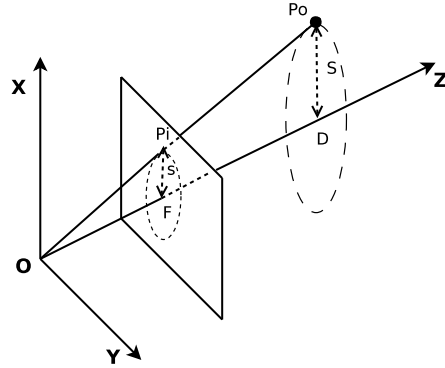

Fig. 1. Model of the camera projection. The object P_o with the size S at the distance D is projected in the focal plane F to obtain P_i with the size s

We can show that

$$\frac{1}{s} = \frac{Z}{S \cdot F}$$

So, the TTC can be expressed only with s and its derivatives :

$$\tau = -\frac{Z}{\dot{Z}} = \frac{s}{\dot{s}}$$

The main difficulty is then to compute the object size in the image. Our approach consists in approximating the size by computing the intrinsic scale from a Laplacian pyramid.

2.2 Intrinsic-scale-based size estimation

To estimate a characteristic size of an object in the image, we used a Scale Space representation, introduced by Witkin [19] and Koenderink [10] as a continuous extension of pyramid-based multi-resolution descriptions of images [4], [2].

The scale-space representation of an image, $I(x, y)$, is a continuous space of images $L(x, y, \sigma)$ obtained by convolution with a variable-scale Gaussian, $G(\sigma)$:

$$L(x, y, \sigma) = G(\sigma) * I(x, y)$$

where

$$G(\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

The Laplacian of an image is a circularly symmetric second derivative computed from the convolution of derivatives in orthogonal directions. The Laplacian of an image can be computed by convolution with sampled Gaussian derivatives.

$$Lap(x, y, \sigma) = \sigma^2 \left(\frac{\partial^2 G}{\partial x^2} * I(x, y, \sigma) + \frac{\partial^2 G}{\partial y^2} * I(x, y, \sigma) \right) \tag{1}$$

As with the Gaussian function, the Laplacian can be used to define a continuous Laplacian scale space by varying scale (sigma) over a range of values. Extremal points in a Laplacian scale space have long been known to provide scale invariant feature points [5] for matching or registering images. Such points, known as "natural interest points", have recently become popular for use as key-points for matching and recognition [13]. Experimental comparison with other key-points has demonstrated that natural interest points provide reliable scale invariant key-points for registration and recognition [15].

Except degenerate points, one or more intrinsic scales may be measured at any position within an image from the values of sigma at which the Laplacian exhibits a local extrema [12]. This property has been used to define image indexing and recognition processes that are invariant to camera position [6].

For any image position, the intrinsic scale is proportional of the characteristic size of the local image feature. This suggests that Time-To-Collision can be estimated with an intrinsic scale. However, for such an approach to work, images must be registered. Such registration can be obtained by detecting and tracking interest points and lines.

2.3 Object detection and tracking

2.4 Natural interest points

As described in the previous section, extremal points in the Laplacian scale space can be used for matching. To detect such points, the method consists in computing a Laplacian sampled scale space. Then, for each level of scale, each pixel is compared with its eight space neighbors and its two scale neighbors. A point is accepted if it the Laplacian is greater or less than all of its neighbors. To compute an accurate

estimation of the intrinsic scale, we can approximate the Laplacian profile over a scale with a cubic spline. The zero of the derivate gives us the intrinsic scale. It is interesting to characterize significance in order to filter such points. To do so, it is possible to consider the Laplacian value and the curvature of the Laplacian profile. The Laplacian value describes the contrast of the points in the image, and the curvature of the Laplacian profile characterizes the contrast over scales.

Fig. 2. Interest point detection. The size of the circles correspond to the intrinsic scale. Circular objects as the walker's body are well detected whereas stretched objects give unstable results.

The figure 2 shows the interest points detected in an urban environment. We can see that interest points are obtained for circular objects as the walker and some part of cars. Nevertheless, stretched objects are not adequately described. The pole at the right is not detected whereas many points are detected on the wall. To resolve this problem, Lowe [14] proposes to eliminate all points that correspond to boundaries. In our case, this solution is not acceptable because most of objects in urban environment are stretched and would not be detected.

2.5 Extension to lines

As seen in the previous section, natural interest points are local extrema of the Laplacian over all directions (x, y and scale). If we relax one spatial constraint, it is possible to detect elongated feature.

Three steps are needed to detect such feature and to compute their characteristics. The first step consists in detecting ridge points in the image. Next, a connexity analysis is computed to separate each feature. The last step computes the first and second order moments.

Ridge detection

To extract ridge points at each level, a method has been developed in [9]. For each pixel of the Gaussian image, the principal curvature direction is computed, and if the pixel is an extremum of the Laplacian image in this direction, this is a ridge.

The principal curvatures and their directions are respectively the eigen values and eigen vectors of the Hessian matrix :

$$\mathcal{H} = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial xy} \\ \frac{\partial^2 f}{\partial xy} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}$$

Where $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial xy}$, $\frac{\partial^2 f}{\partial y^2}$ are the second derivates of the Image.

The principal curvature direction correspond to the eigen vector associated to the biggest eigen value. If the Laplacian assumes a local extremum in this direction, this is a ridge point.

(a)   (b)

Fig. 3. Gaussian image 3(a) and detected ridges 3(b) at the scale level $\sigma = 4$. As we can see, the elongated shape, as the legs or the ground lines are well detected.

The result of this step is a binary image for each scale level, which represents the ridges (cf. fig.3).

Connexity analysis

Ridge structure is obtained by connected component analysis. Two neighbor points (over scale and space) have the same label if they are identified as ridge and if the difference of two local ridge directions at these two points is inferior to a threshold.

Computing moments

To obtain the mean position, the mean scale and the orientation of ridge features, we compute first and second order moments.

Let N be the number of pixels of the feature and $X_{n=1..N} = \begin{pmatrix} x_n \\ y_n \\ s_n \end{pmatrix}$ the points

which form the feature, where s_n is the intrinsic scale.

The gravity center is :

$$\mu = \frac{1}{N} \sum_{n=1}^{N} X_n \tag{2}$$

To give more weight to hight contrast points, we have to weight the position by the absolute value of the Laplacian at this position. So the formula becomes :

$$\mu = \frac{\sum_{n=1}^{N} |lap(X_n)| X_n}{\sum_{n=1}^{N} |lap(X_n)|} \tag{3}$$

Because we work with a pyramid, each stage k of this pyramid is down-sampled, so a ridge point in the stage k is equivalent to two points at the stage $k-1$. The formula becomes :

$$\mu = \frac{\sum_{n=1}^{N} 2^{k_n} |lap(X_n)| X_n}{\sum_{n=1}^{N} 2^{k_n} |lap(X_n)|} \tag{4}$$

The second order moments are needed to evaluate the size and the orientation of the ridge. The covariance matrix C is :

$$C_{ij} = \frac{\sum_{n=1}^{N} 2^{k_n} |lap(X_n)| (X_{ni} - \mu_i)(X_{nj} - \mu_j)}{\sum_{n=1}^{N} 2^{k_n} |lap(X_n)|} \tag{5}$$

2.6 Tracking interest lines

Our method for tracking object in video is inspired by [1]. It begins with a detection process based on ridge detector explained previously, which gives a set of tracked objects.

Then for each frame, a correlation process is needed to associate new detected feature with tracked objects.

Ridge similarity

- *Spacial distance:* distance between two gravity centers

$$S_{prox} = \frac{\|X_A - X_B\|}{\overline{s_A}} \tag{6}$$

- *Scale distance :* distance between two average scales

$$S_{scale} = |\log(\frac{\overline{s_A}}{s_B})| \tag{7}$$

- *Significance:* compare the average Laplacian values

$$S_{lap} = |\log(\frac{\overline{lap_A}}{\overline{lap_B}})| \tag{8}$$

- *Second order moments :*

$$S_{mom} = \sqrt{\frac{\sum_{i,j}((C_A)_{ij} - (C_B)_{ij})^2}{\sum_{i,j}(C_A)_{ij}^2}} \tag{9}$$

The combined similarity is the weighted sum of (6), (7), (8) and (9)

$$S_{comb} = c_{pos}S_{pos} + c_{scale}S_{scale} + c_{lap}S_{lap} + c_{mom}S_{mom} \tag{10}$$

Where c_{pos}, c_{scale}, c_{lap} and c_{mom} are negative.

Fig. 4. Test sequence for ridge tracking : the licence plate of the car is tracked with our method. The rectangle represents the region of interest and the segment at center represents the position and the direction of the tracked ridge

For each tracked object, a region of interest is computed using the position and the size of the object. A similarity score is then computed for each detected object

in the new frame, the object which obtains the highest score is associated with the tracked object and it is updated with the new properties (position, scale, etc.). In the figure 4, we track the licence plate of the car. We can notice that both the camera and the car move.

3 Experimental results

To verify the precision of the TTC obtained by our method, we propose to capture images from a moving vehicle equipped with a laser range finder and a camera. In order to make the comparison possible, the camera is put on the laser.

In the test case (cf. fig.5(a)), the target is a white reflective mark placed in the camera axe. This mark can be detected in the laser scan using the intensity of the reflected signal. The laser provides precise distances to the obstacle and the Time-To-Collision can be easily computed with the formula : $\tau(t) = -\frac{z(t)}{\frac{\partial z}{\partial t}(t)}$

To detect the mark in the image (of size 640x480 pixels), we use our ridge tracker described in the previous section. The tracker is manually initialized, selecting the region of interest. At each frame, the object is tracked and its properties (position, orientation, scale) are computed. We use a Kalman Filter to improve these values. The time-to-collision is obtained with a windowed linear regression. The global computation time, including Laplacian scale space calculation, ridge extraction and tracking and the computation of time-to-collision is about 50 milliseconds in a 3GHz Pentium 4 processor.

In the test, the approach speed is constant and the distance to the obstacle, $z(t)$, is linear (cf. fig.5(b)). The inverse of the object size in the image is proportional to the distance so that the function $\frac{1}{s(t)}$ is linear (cf. fig.5(c)). The figure 5(d) shows the difference between the real TTC obtained by the laser and the TTC computed by our method. This experiment shows that, after a short time of instability due to the lower size of the seen object in the image, the TTC is well estimated.

4 Conclusion

The method described in this paper is a new visual approach to compute the time to contact from variation of intrinsic scale. To solve the problem of image registration we use a method based on natural interest points and an extension for oblong objects which requires an extraction and a tracking of ridges. Such an approach gives good results on contrasted features. In the shown test, only one object is tracked for more clarity, but several objects can be tracked at the same time. This approach is interesting as it uses advantages of cameras (low cost, wide field of view, etc.) and offers low computation time, thanks to a pyramidal algorithm.

One possible direction to further research will be to generalize the computation of TTC on more image points in order to obtain more information about the environment, an ideal case would be to compute a continuous field of TTC. Another point in our

(a) Video sequence used to evaluate the precision.
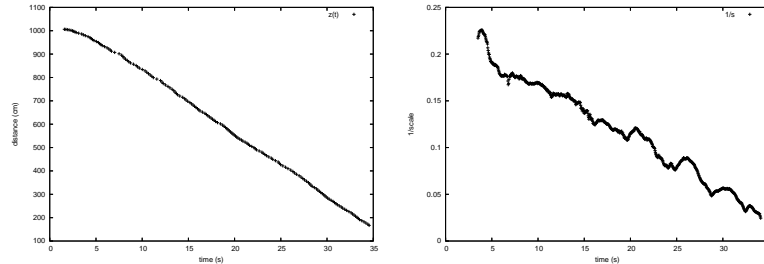
(b) Distance measured by the laser range finder

(c) Inverse of the scale evaluated by intrinsic scale
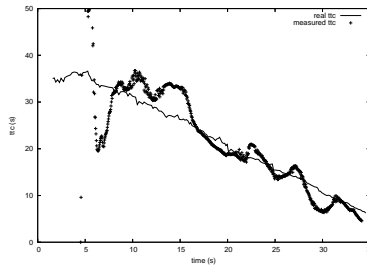
(d) TTC obtained by the laser range finder and computed with TTC from intrinsic scale.

Fig. 5. Test with a constant speed. The distance (5(b)) and the inverse size (5(c)) of the obstacle are linear. The time-to collision is shown on (5(d)), after a period of initialization, the two curve are fairly near.

approach is the hypothesis that objects move with constant speed. This work uses an approximation with a piecewise linear model, future work will explore more complexes models.

References

1. L. Bretzner and T. Lindeberg. Feature tracking with automatic selection of spatial scales. *Computer Vision and Image Understanding*, 71(3):385–392, 1998.

2. P. Burt and E. H Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 9(4):532–540, 1983.
3. T.A. Camus. *Real-time optical flow*. PhD thesis, Brown university, Providence, RI 02912, USA, September 1994.
4. J. L. Crowley. *A representation for visual information*. PhD thesis, Carnegie-Mellon University, 1981.
5. J. L. Crowley and A. C. Parker. A representation for shape based on peaks and ridges in the difference of low-pass transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(2):156–170, 1984.
6. J.L. Crowley D. Hall and V. Colin de Verdière. View invariant object recognition using coloured receptive fields. *Machine GRAPHICS and VISION*, 9(2):341–352, 2000.
7. N. Franceschini, J. M. Pichon, and C. Blanes. From insect vision to robot vision. *Philosophical Transactions : Biological Sciences*, 337:283–294, 1992.
8. J.J. Gibson. *The ecological approach to visual perception.* Houghton Mifflin, Boston, 1979.
9. TRAN Thi Thanh Hai and Augustin Lux. Extraction de caractéristiques locales : Crêtes et pics. In *RIVF*, pages 203–211, 2003.
10. Jan J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–396, 1984.
11. D.N. Lee. A theory of visual control of braking base on information about time-to-collision. *Perception*, 5:437–459, 1976.
12. T. Lindeberg. Feature detection with automatic selection of spatial scales. In *IJCV*, volume 30, pages 79–116, 1998.
13. D. G. Lowe. Object recognition from local scale-invariant feature. In *International Conference on Computer Vision*, pages 1150–1157, 1999.
14. D. G. Lowe. Distinctive image features from scale-invariant keypoints. In *IJCV*, volume 60, pages 91–110, 2004.
15. K. Mikolajezyk and C. Schmid. Indexing based on scale invariant interest points. In *Proceedind of the International Conference on Computer Vision*, pages 525–531, 2001.
16. R. C. Nelson and J. Aloimonos. Obstacle avoidance using flow field divergence. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(10):1102–1106, 1989.
17. W. Schiff, JA. Caviness, and JJ. Gibson. Persistent fear responses in rhesus monkeys to the optical stimulus of "looming". *Science*, 136:982–3, 1962.
18. M. Tistarelli and G. Sandini. On the advantages of polar and log-polar mapping for direct estimation of time-to-impact from optical flow. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(4):401–410, 1993.
19. Andrew P. Witkin. Scale-space filtering. *IJCAI*, pages 1019–1023, 1983.