Computing Structure and Motion of General 3D Rigid Curves from Monocular Sequences of Perspective Images

Théodore Papadopoulo, Olivier Faugeras

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Computing Structure and Motion of General 3D Rigid Curves from Monocular Sequences of Perspective Images

Théo Papadopoulo, Olivier Faugeras

N° 2765

Décembre 1995

PROGRAMME 4

Rapport de recherche

ISSN 0249-6399

Computing Structure and Motion of General 3D Rigid Curves from Monocular Sequences of Perspective Images

Théo Papadopoulo, Olivier Faugeras

Programme 4 — Robotique, image et vision
Projet Robotvis

Rapport de recherche n°2765 — Décembre 1995 — 27 pages

Abstract: This report discusses the problem of recovering 3D motion and structure for rigid curves. For this purpose, we use long monocular sequences of images of the curve and compute some set of derivatives (up to the second order) that are defined on the so-called spatio-temporal surface generated by the curve. For general 3D rigid curves, there is exactly one constraint for each image point that relates these derivatives to the kinematic screw and its first order time derivative. These equations derive directly from the perspective equations. However, this theory ignores (as does the projective model of cameras) the fact that the viewed object must be entirely in front of the focal plane.
In this paper, we provide a complete description (up to second order) of these conditions and of their geometrical consequences for 3D reconstruction. Moreover, the relations between the geometry of the image curve and the kinematic screw are investigated. We show that ignoring these constraints leads to erroneous solutions with a high probability even when using perfect data. Then, an algorithm which takes advantage of these constraints is proposed. The accuracy of the obtained results is demonstrated on both synthetic and real image sequences.

Key-words: Motion analysis, Rigid curves, Structure and motion, Monocular Vision, Optical Flow, Spatio-Temporal Surfaces, Kinematic Screw

(Résumé : tsvp)

Unité de recherche INRIA Sophia-Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex (France)
Téléphone : (33) 93 65 77 77 – Télécopie : (33) 93 65 77 65

Calcul du mouvement et de la structure de courbe rigides 3D générales à partir de séquences d'images perspectives monoculaires.

Résumé : Ce rapport traite du problème de la détermination du mouvement et de la structure d'une courbe rigide 3D à partir d'une longue séquence d'images monoculaires de celle-ci. Cette séquence permet de calculer une certain ensemble de grandeurs différentielles (jusqu'à l'ordre 2) qui sont définies sur la *surface spatio-temporelle* engendrée par la courbe. Pour des courbe 3D rigides générales, il existe en chaque point de la courbe image exactement une contrainte reliant ces dérivées au torseur cinématique associé au mouvement et à la dérivée première de ce torseur. Cette relation provient directement des equations de projection perspective. Cependant, cette théorie ne tient pas compte (tout comme le modèle de caméra utilisé) du fait que l'objet observé doit se trouver entièrement devant le plan focal pour pouvoir être observable.
Dans ce rapport, nous donnons une description complète (jusqu'à l'ordre 2) des conditions qui doivent être réalisée pour assurer l'observabilité de l'objet et des conséquences géométriques de ces conditions sur la reconstruction 3D. Une analyse sommaire du nombre de ces contraintes en fonction de la géométrie de la courbe image et de la position du torseur cinématique est également proposée. Nous montrons tout d'abord qu'ignorer ces contraintes conduit avec une forte probabilité à des solutions erronées même lorsque on utilise des données parfaites (c'est à dire non bruitées). Puis, nous proposons un algorithme prenant en compte ces contraintes. La précision des résultats obtenus est analysée à l'aide d'images synthétiques et réelles.

Mots-clé : Analyse du mouvement, Courbes rigides, Détermination du mouvement et de la structure, Vision monoculaire, Flot optique, Surfaces spatio-temporelles, Torseur cinématique

1 Introduction

Among the methods which attempt to recover three-dimensional information about an observed scene from images, those involving only monocular sequences are some of the most difficult to bring to life. Many of the methods proposed up to now are either point or line based (see [ZHAH88, WHA92] for example). Most of the general methods for edges are based on the point equations and it is necessary, at least in theory, to compute the full motion field along the edges; in general this is known to be theoretically impossible. To solve this problem, researchers have developed a number of techniques:

- Imposing some additional properties on the geometry of the 3D curve. For example, the fact that the 3D curve is planar [WWB91, PF94b], or assuming that the depth is known;
- Imposing some knowledge about the motion. For example, techniques have been defined for pure rotation, pure translation or known rotation [HW88];
- "Inventing" a full motion field compatible with the normal one and that obeys some smoothness constraint [Hil83b, WW88].

None of these techniques are really satisfying from a theoretical point of view, and moreover some have never been tested on real images. The work described in this paper attempts to solve the same kind of problem but does not suppose that point matches are available, and furthermore does not compute an estimate of the full motion field prior to the 3D motion computation (it corresponds essentially to the implementation of the theory described in [Fau93, FP93]). From this point of view, we show how crucial it is to take into account the visibility constraints described in [Car94] for the motion computation. Moreover, it is demonstrated on examples that these are still not sufficient if an accurate 3D structure has to be computed from the motion since small errors in the motion parameters induce discontinuities in the reconstructions. More recently, [CÅG95, JAP95] have presented a generalization of the work of [Car94] to compute the motion of dynamic silhouettes.

This paper is organised into four parts: the first section introduces some notation and terms, and gives a brief reminder of the underlying theory. The second section focusses on some singular situations resulting from this theory whereas the third and fourth sections describe respectively an algorithm to solve the considered problem and the results which have been obtained with it.

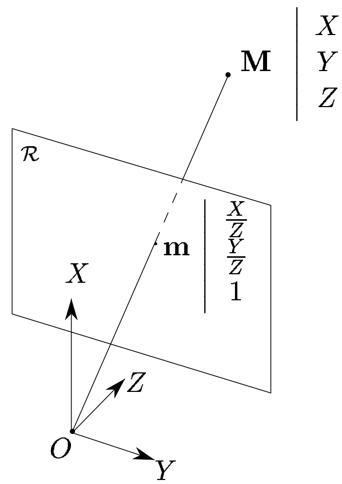

Figure 1: The pinhole model of a camera

2 The Theory of Motion Fields for Curves

The goal of this section is to recall briefly some results about motion fields of curves and to introduce notation which will be used in the remainder of this paper. Full descriptions of the used quantities can be found in [Fau93, FP93]. Notice that although the presentation adopted here makes an important use of some concepts introduced in these papers, the overall complexity of the theory is much lower as some equations given previously have been shown to be useless and are thus no longer considered [PF94b, Pap95].

2.1 The Camera Model

We assume that the camera obeys the pinhole model with unit focal length as shown in Fig. 1. We denote O as the focal center point and $\mathcal{R}$ as the retinal plane. A frame (O, X, Y, Z) is attached naturally to the camera by supposing that the plane (O, X, Y) is parallel to $\mathcal{R}$. As such, all equations involving 3D parameters will be written in this frame.

Given a 3D point $\mathbf{M} = (X, Y, Z)$ and its 2D perspective projection $\mathbf{m} = (x, y, 1)$ on $\mathcal{R}$, we express their relationship by the perspective equation:

$$\mathbf{M} = Z\mathbf{m} \ . \tag{1}$$

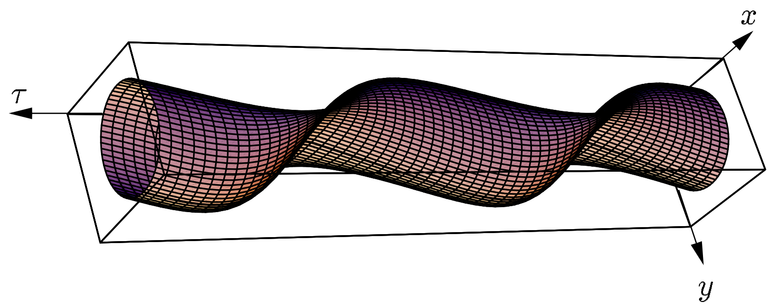

Figure 2: The spatio-temporal surface generated by a circle rotating in front of the camera.

This equation is fundamental in that all of the constraints we will present here are derived directly from it. The concept of temporal changes can be incorporated with the introduction of a time parameter τ.

2.2 Definitions

We now assume that we observe in a sequence of images a family (c_τ) of curves, where τ denotes time. For each time instant τ, c_τ is supposed to be the perspective projection in the retina of a 3D curve (C) that moves in space. Furthermore, for each time instant τ, c_τ is assumed to be a regular curve. This family of curves sweeps out a surface (Σ) in the three-dimensional space (x, y, τ). Figure 2 illustrates an example of such a spatio-temporal surface which has been generated by a circle rotating around one of its diameters in front of the camera.

At a given time instant τ, let s be the Euclidean arclength of (c_τ) and S the Euclidean arclength of (C). We further suppose that S is not a function of time (i.e. the motion is isometric). To define the speed of a point of the image curve, we have to specify a way to define the path followed by that point over time. Two natural strategies can be adopted toward this goal: firstly, such a path can be defined by considering the points for which S is constant over time, or alternatively we can keep s constant. It is important to notice that keeping S constant has a physical meaning: a point is fixed on the 3D curve. Of course, this choice is impossible both in practice and in general. On the contrary, keeping s constant over time is quite easy to realize, but this choice has no obvious physical meaning. Momentarily disregarding this difficulty, we have thus defined two different motion fields:

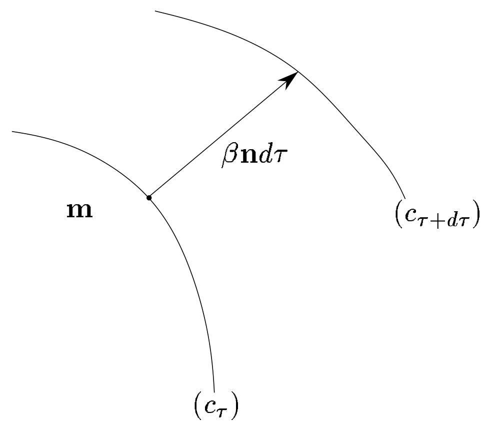

Figure 3: A geometric interpretation of β

- The *real motion field* $\mathbf{v}^r_{\mathbf{m}}$ (r denotes *real*) is the partial derivative of $\mathbf{m}(s,\tau)$ with respect to time when S is kept constant, or its total time derivative $\dot{\mathbf{m}}$. This field is the projection of the 3D velocity field in the retina;

- The *apparent motion field* $\mathbf{v}^a_{\mathbf{m}}$ (a for *apparent*) is the partial derivative of $\mathbf{m}(s,\tau)$ with respect to time when s is kept constant, $\frac{\partial \mathbf{m}}{\partial \tau} = \mathbf{m}_\tau$.

Moreover, introducing the Frenet frame $(\mathbf{t},\mathbf{n})$, where $\mathbf{t}$ and $\mathbf{n}$ are respectively the unit tangent and normal vectors to (c_τ) at $\mathbf{m}$, we have:

$$\mathbf{v}^a_{\mathbf{m}} = \alpha\mathbf{t} + \beta\mathbf{n} \; ,$$
$$\mathbf{v}^r_{\mathbf{m}} = w\mathbf{t} + \beta\mathbf{n} \; ,$$

where α is the apparent tangential motion field, w is the real tangential motion field. β, the normal motion field is a common component to both of these motion fields. Note that β has a nice geometric interpretation (as shown in Fig. 3). We consider the normal $\mathbf{n}$ at point $\mathbf{m}$ of curve (c_τ): at time $\tau + d\tau$, the curve $(c_{\tau+d\tau})$ is intersected by the line defined by $\mathbf{m}$ and $\mathbf{n}$ at a point represented by $\mathbf{m} + \beta\mathbf{n}d\tau$.

With this notation, under the weak assumption of *isometric* motion, the following conclusions can be reached from the study of the spatio-temporal surface [Fau93, FP93]:

1. The normal motion field β can be recovered from the normal to the spatio-temporal surface;

2. The tangential apparent motion field can be recovered from the normal motion field;

3. The tangential real motion field *cannot* be recovered from the spatio-temporal surface.

Therefore, the full real motion field is not computable from the observation of the image of a moving curve under the isometric assumption. This can be considered as a new statement of the so-called *aperture* problem [Hil83a]. In order to solve it we *must* add more constraints, for example that the 3D motion is rigid.

2.3 The Case of a Rigid 3D Curve

Assuming now that the curve (C) is moving rigidly, let $(\mathbf{\Omega}, \mathbf{V})$ be its kinematic screw at the optical center O of the camera. Additionally, recall that the velocity $\mathbf{V_M} = \dot{\mathbf{M}}$ of any point $\mathbf{M}$ attached to the rigid curve is given by $\mathbf{V_M} = \mathbf{V} + \mathbf{\Omega} \times \mathbf{M}$. Taking the total derivative of the equation $\mathbf{M} = Z\mathbf{m}$ with respect to time, we get:

$$\dot{Z}\mathbf{m} + Z\dot{\mathbf{m}} = \mathbf{V_M} \ ,$$
$$\mathbf{V}_{\mathbf{M}_Z}\mathbf{m} + Z(w\mathbf{t} + \beta\mathbf{n}) = \mathbf{V_M} \ ,$$

where $\mathbf{V}_{\mathbf{M}_Z}$ is the component of $\mathbf{V_M}$ along the Z axis. Projecting this vector equation onto $\mathbf{t}$ and $\mathbf{n}$ yields two scalar equations:

$$Z(w + \mathbf{\Omega}\cdot\mathbf{b}) = -\mathbf{U_n}\cdot\mathbf{V} \ , \tag{2}$$
$$Z(\beta - \mathbf{\Omega}\cdot\mathbf{a}) = \mathbf{U_t}\cdot\mathbf{V} \ , \tag{3}$$

where $\mathbf{U_t}$ denotes $\mathbf{m}\times\mathbf{t}$ and $\mathbf{U_n}$ denotes $\mathbf{m}\times\mathbf{n}$ and $\mathbf{a}$ and $\mathbf{b}$ are given by $\mathbf{a} = \mathbf{m}\times\mathbf{U_t}$ and $\mathbf{b} = \mathbf{m}\times\mathbf{U_n}$.

These equations are fundamental in the sense that they express the relationship between the unknown 3D motion of a point and the real motion field of its image. Notice that these equations are not new, they are just a reformulation (in the basis $\mathbf{t}, \mathbf{n}$) of the standard *image field* equations derived by Longuet-Higgins and Prazdny [LHP80]. Now, equation (2) can be considered as the *definition* of w given Z, $\mathbf{\Omega}$ and $\mathbf{V}$, whereas equation (3) can be considered as a constraint equation should the unknown Z not be present and thus varying with the considered point. The solution is thus to find a new estimate for Z as a function of measurable values and of the unknown quantities which are constant along the curve. One way to obtain such an estimate without adding new constraints is to differentiate (3) with respect to time:

$$\mathbf{V}_{\mathbf{M}_Z}(\beta - \mathbf{\Omega}\cdot\mathbf{a}) + Z\left(\dot{\beta} - \dot{\mathbf{\Omega}}\cdot\mathbf{a} - \mathbf{\Omega}\cdot\dot{\mathbf{a}}\right) = \dot{\mathbf{U}}_{\mathbf{t}}\cdot\mathbf{V} + \mathbf{U}_{\mathbf{t}}\cdot\dot{\mathbf{V}}\ . \tag{4}$$

This operation introduces, of course, the new unknowns $\dot{\mathbf{\Omega}}$ and $\dot{\mathbf{V}}$. All the other new quantities which appear in this formula ($\mathbf{V}_{\mathbf{M}_Z}$, $\dot{\beta}$, $\dot{\mathbf{U}}_{\mathbf{t}}$) can be expressed as functions of the unknowns $\mathbf{\Omega}$, $\mathbf{V}$, $\dot{\mathbf{\Omega}}$, $\dot{\mathbf{V}}$, and of the values $\mathbf{m}$, $\mathbf{n}$, β, $\frac{\partial\beta}{\partial s}$, κ (the curvature at point $\mathbf{m}$), and $\partial_{\mathbf{n}_\beta}\beta$ (the normal acceleration) which can be recovered from the spatio-temporal surface (Σ). The resulting equation can be written as:

$$(\mathbf{V}_Z + Z(\mathbf{\Omega}\times\mathbf{m})\cdot\mathbf{k}))(\beta - \mathbf{\Omega}\cdot\mathbf{a}) + Z\left(w\frac{\partial\beta}{\partial s} + \partial_{\mathbf{n}_\beta}\beta - \dot{\mathbf{\Omega}}\cdot\mathbf{a} - \mathbf{\Omega}\cdot\dot{\mathbf{a}}\right) = \dot{\mathbf{U}}_{\mathbf{t}}\cdot\mathbf{V} + \mathbf{U}_{\mathbf{t}}\cdot\dot{\mathbf{V}}\ . \tag{5}$$

Replacing w and $\dot{\mathbf{U}}_{\mathbf{t}}$ by their expressions in terms of the other quantities yields a linear equation in Z; it is thus easy to eliminate this quantity between this new equation and (3). This can be done *iff.* the coefficient of Z in these equations is non-zero. Points of the image curve for which this condition is true are called *non-degenerate.* We can thus state the following theorem.

Theorem 1 *At each non-degenerate point of an observed curve* (c_τ) *evolving during time, it is possible to write one polynomial equation in the coordinates* $\mathbf{\Omega}$, $\mathbf{V}$, $\dot{\mathbf{\Omega}}$ *and* $\dot{\mathbf{V}}$*. The coefficients of this equation are polynomials in the quantities*

$$\begin{array}{ccc} \mathbf{m}\,, & \mathbf{n}\,, & \kappa\,, \\ \beta\,, & \frac{\partial\beta}{\partial s}\,, & \partial_{\mathbf{n}_\beta}\beta\ . \end{array}$$

which can be measured from the spatio-temporal surface (Σ)*.*

This equation is denoted by L_1 hereafter (it is too big an equation to be given here). It can be proved that it is the simplest equation relating the motion to some spatio-temporal parameters for general rigid curves. All the other equations can be obtained by differentiating L_1 with respect to s (in which case higher order spatial derivatives appear), or with respect to time (in which case both new unknowns and new spatio-temporal derivatives of higher order in space and time appear). For this reason, only L_1 is considered in the remainder of this paper. Notice that once the motion has been computed, Eq. (3) provides a way to compute the depth Z along the curve. Equation (1) then yields the structure.

Remark 1 *Actually, it is easy to see that L_1 corresponds to the numerator of the fraction appearing when forming the difference of the two depths computed respectively from Eq. (3) and Eq. (4). So, asserting that this quantity is zero is equivalent to saying that the two depths are equal.*

3 Visibility Conditions

Recovering the 3D motion by minimizing the set of equations L_1 expressed at many different points of a curve is very difficult even with perfect data. Basically, the optimizing process almost always becomes lost in a local minimum which is far from the real solution. Carefully analysing the problem shows that these solutions correspond to 3D reconstructions for which differents parts of the curves lie on the two different sides of the focal plane. This is not so strange since the pinhole camera model makes no distinction between the part of the space which is in front of the camera and the one that is behind! Moreover, looking at reconstructions obtained with an algorithm which makes the assumption that the observed curve is planar (this case is much simpler than the one considered here) reveals that there are sometimes "singular" points on the curve (see figure 4). These two facts are actually correlated and correspond to the degenerate points of the previous section. The purpose of this part is to investigate further the implications of these singular points.

From a mathematical point of view, there are two cases for which a degenerate point may appear: these cases correspond to the places on the curve for which the coefficients of the depth Z in either (3) or (4) is zero. The next two paragraphs study respectively the constraints arising in these cases. The main conclusion of this section is that at such points (when they exist), there is a local loss of depth information, however they yield an important constraint on the 3D motion.

3.1 The First Order Visibility Constraints

The first case arises from the situation where the coefficient in Z in (3) is zero: $\beta - \mathbf{\Omega}\cdot\mathbf{a} = 0$. From Eq. (3), this implies that the quantity $\mathbf{U_t}\cdot\mathbf{V}$ must also vanish at the same point. Physically (but not mathematically), the converse is also true since if $\mathbf{U_t}\cdot\mathbf{V} = 0$, then either $Z = 0$ or $\beta - \mathbf{\Omega}\cdot\mathbf{a} = 0$. The first case is impossible (all the viewed point must have $Z > 0$), this implies that $\beta - \mathbf{\Omega}\cdot\mathbf{a} = 0$.

There is another way to analyze the special case $\beta - \mathbf{\Omega}\cdot\mathbf{a} = \mathbf{U_t}\cdot\mathbf{V} = 0$. This is by considering the sign of Z along the curve at a fixed time instant. If only

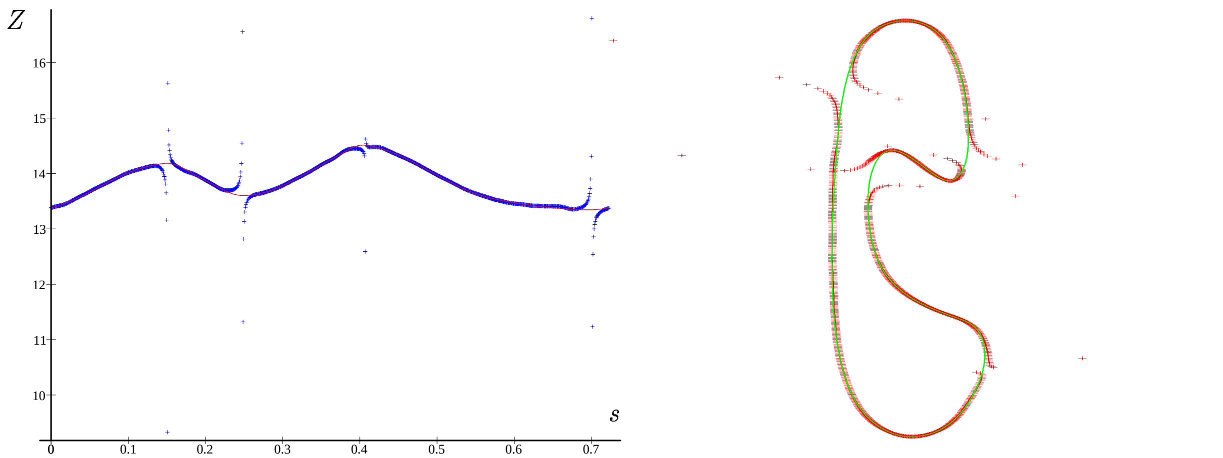

Figure 4: On the left: the Z value corresponding to the depth of the points of the curve is plotted as a function of the arclength of the curve. On the right: the corresponding 3D reconstruction. These results were obtained with a program recovering motion and structure for planar curves. This program simultaneously computes the equation of the plane and the kinematic screw. In each case the two plots superimposed correspond to the values obtained using either the plane equation or the kinematic screw and equation (3). Whereas better results can be obtained, these are making conspicuous the 4 unstable points that appear in this case.

one of these two quantities cancels out then the sign of Z has to change at that point. Since the curve must be observable it is not possible that one part has $Z > 0$ whereas the other one has $Z < 0$, thus the other quantity must be zero as well at the same point. Then, $\beta - \mathbf{\Omega}\cdot\mathbf{a}$ and $\mathbf{U_t}\cdot\mathbf{V}$ will change sign simultaneously and the sign of Z is constant (it can be indifferently positive or negative, the two situations being toggled by a change of sign of $\mathbf{V}$). Consequently, we can state the following proposition:

Proposition 1 *Ensuring that the conditions $\beta - \mathbf{\Omega}\cdot\mathbf{a} = 0$ and $\mathbf{U_t}\cdot\mathbf{V} = 0$ are true at the same point of the curve is equivalent to ensuring that the reconstructed curve will be entirely in front or behind the focal plane. For this reason, and because they come from the study of equation (4) which involves only first order derivatives, we will call these degeneracy conditions the* first order visibility conditions. *The points at which these conditions are verified are called* first order degenerate points.

Remark 2 *Trying to compute the depth at a degenerate point using Eq. (3) yields an underdeterminacy of type 0/0. Of course, in practice, this situation never occurs exactly due to the noise on the data. However, at these points the depth information is unstable and this explains the apparition of the unstable points of Fig. 4.*

The importance of these constraints for the motion analysis of curves has already been recognized in [Car94]. It is quite easy to show that degenerate points are generically isolated points on the curve, the only possible configurations of curve and motion for which non isolated points appear being[1]:

- The curve is reduced to a point.
- The curve is locally part of a straight line passing through the projective point $\mathbf{V}$ (the focus of expansion).
- $\mathbf{V} = \mathbf{0}$.

All these cases are non-generic, and so are not taken into consideration hereafter. Following [Car94], it is important to notice that the condition $\mathbf{U_t}\cdot\mathbf{V} = 0$ has a very nice geometrical interpretation corresponding to the fact that $\mathbf{V}$ is the continuous analog to the epipole for a camera in a stereo rig. Actually, $\mathbf{U_t}$ is the projective representative of the tangent line to the curve (c_τ) at point $\mathbf{m}$: thus the condition

[1] This result is obtained starting with $\beta - \mathbf{\Omega}\cdot\mathbf{a} = \mathbf{U_t}\cdot\mathbf{V} = 0$, and differentiating these equations with respect to s. This is possible as we now assume that the degenerate points are non isolated.