

An Introduction to Model-Based Pose Estimation and 3-D Tracking Techniques

Christophe Doignon
*LSIIT - University of Strasbourg
 France*

1. Chapter overview

The aim of this chapter is to present a general overview of the feature-based 3-D pose estimation and tracking techniques. Principles, classical techniques and recent advances are presented and discussed in the context of a monocular camera. The objective is to focus on techniques employed within both the visual servoing and registration fields for the wideclass of rigid objects. The main assumption to this problem rely on the availability of a 3-D model of the object to track.

2. Introduction: Model-based tracking and pose estimation

The recovery of the 3-D geometric information from 2-D images is a fundamental problem in computer vision. When only one view is available, the appearance or the relative arrangement of the object features of interest should be modelled in a symbolic description so as to be compared with the image descriptors thanks to a similarity criterion. Geometric-based approaches restrict the search for correspondence to a sparse set of geometrical features. They use numerical and symbolic properties of entities available. To automatically compute a rigid-body transformation (the pose), it is necessary to match a 3-D model features with part of the visible 2-D image features, a process referred to as the correspondence problem, and for the past four decades, the model-based pose estimation of objects with a simple geometry has been intensively studied.

The major goal is tracking at camera frame rate the pose parameters in the world space. Therefore, features such as points, lines, ellipses are not only extracted from 2-D images, but the 3-D model and the pose of the object has to be also exploited.

2.1 Related work on model-based 3-D tracking

The modelbased 3-D tracking is closely related to the pose estimation problem. It can cope with abrupt motions and it is generally more efficient to deal with partial occlusions of the object of interest than 2-D tracking. However, it needs the correspondence problem to be solve at least once. The definition of object tracking algorithms in image sequences is an important issue for research and applications related to robot vision. A robust extraction and realtime spatiotemporal tracking of measurements is one of the keys to a successful visual servoingbased tracking, in particular for positionbased visual servoing approaches

(PBVS) (Hutchinson, 1996), where the tracking error is computed in the task space and thereafter used by the robot control system.

The reference work on visionbased navigation is the dynamic tracking approach introduced by Dickmanns and Graefe (Dickmanns, 1988). In this work, the steering of cars, vehicle docking and aircraft landing are performed by dynamic modelling and edge feature extraction and processing.

Feature extraction and matching play an important role, and restriction of matching is commonly done by windowing technique. Harris and Stennet (Harris, 1990), Papanikolopoulos et al. (Papanikolopoulos, 1993) and Hager and Toyama (Hager, 1998) with the XVision system restrict the search space of matching to a neighbourhood of the searched features. This reduces computational costs because only small parts of the image are processed. More recently, Thompson et al. (Thompson, 2001) use robust methods for line fitting and pose estimation by applying iterative median filters in order to detect outliers. In (Marchand, 2004), Marchand & Chaumette described both feature tracking issues and modelbased tracking for visual servoing techniques. Compared to Thompson's work, robust methods are used for tracking visual 2-D features which are related in turn to the 3-D motion (the motion field applied to imagebased visual servoing purposes). In this work, the pose estimation problem is seen as the dual problem of 2-D visual servoing: for a given pose, a set of virtual 2-D measurements are computed (the virtual camera). The approach consists of estimating the real pose by minimizing visual measurements errors (thus, by moving the virtual pose to the real one) with a robust control law. This approach can be thought as a minimization technique driven by the the 2-D/3-D mapping (perspective law herein) at each iteration, and which restricts the path from a guess solution to the desired one.

2.2 The pose determination versus camera calibration

Generally, the identification of some of the camera parameters - the intrinsic parameters - is needed to achieve the pose. This process named as the camera calibration can be done as a preliminary step or it can be achieved in some situations simultaneously with the estimation of the pose (Sturm, 1997). The extrinsic parameters of the pose relate the world coordinates to the camera orientation and position and there is a lot of related works in the computer vision and photogrammetry literatures both on the pose estimation and camera calibration (Horn, 1986; Tsai, 1987; Haralick, 1992; Hartley, 2000) as the two problems are strongly connected. With a single view, the accuracy of the intrinsic and extrinsic parameters estimates relies on the ability to extract the perspective effect in a reliable fashion from the imaged 3-D objects or from a couple of sets of planar patterns (Doignon, 1999; Zhang, 1999). Non linear effects and electronic synchronizations has been modelled by (Tsai, 1987) twenty years ago and are significant disturbing factors in the captured perspective image. The radial alignment constraint (RAS) described by Tsai alleviates some of the above mentioned problems for parts of extrinsic parameters, mainly the orientation and the direction of the position vector.

The practical use of a camera calibration procedure is strongly limited by the socalled correspondence problem since many methods used points (Faugeras, 1987 ; Tsai, 1987) or lines features (Liu, 1990) for that purpose. Recently, Meng and Hu have proposed a calibration based on circular points which does not need any correspondences (Meng, 2003) and nor computation of extrinsic parameters. A complete review of most of calibration techniques is reported in (Salvi, 2002).

2.3 The importance of feature grouping

The pose determination from a single perspective view requires a local geometric description of objects and image features of interest. The space search for the one-to-one correspondence may be very large when no constraint is used to associate them. Various recognition schemes have been proposed in the past to solve the search of the correspondence problem like the interpretation tree (Grimson, 1990), geometric hashing (Lamdan, 1988), aspect graphs (Ikeuchi, 1987; Hansen, 1989), focal features (Bolles, 1982 and 1986), pose clustering (Olson, 2001), the soft assignment (Gold, 1998; David, 2002) and alignment techniques (Huttenlocher, 1990; Torr, 1999; Bartoli, 2001).

Many authors have presented solutions to this problem in the context of the registration with alignment techniques. The goal is then to match two subsets (a subset of image features with a subset of features of the 3-D model) corresponding to a geometric transformation consistent with the data [2] (Haralick, 1984; Lowe, 1987; Ikeuchi, 1987; Thompson, 1987; Sugihara, 1988; Grimson, 1991; Jurie, 1999; David, 2002). For linear primitives such as points (vertices of a polyhedra, corners, zero-curvature points, intersections of lines,...) or straight lines (edges, tangent lines at zero-curvature points, polar lines,...), the search space of a consistent viewpoint¹ is large as several features are needed to constrain it (typically 3 to 5 matched linear features are required).

Quadratic primitives constrain more the viewpoint. In many situations, only one quadratic feature is needed to find a small set of consistent pose parameters. In counterpart, quadratic primitives are more tricky to detect with a sufficient level of reliability (see (Weiss, 1993; Werghi, 1995; Pilu, 1996)).

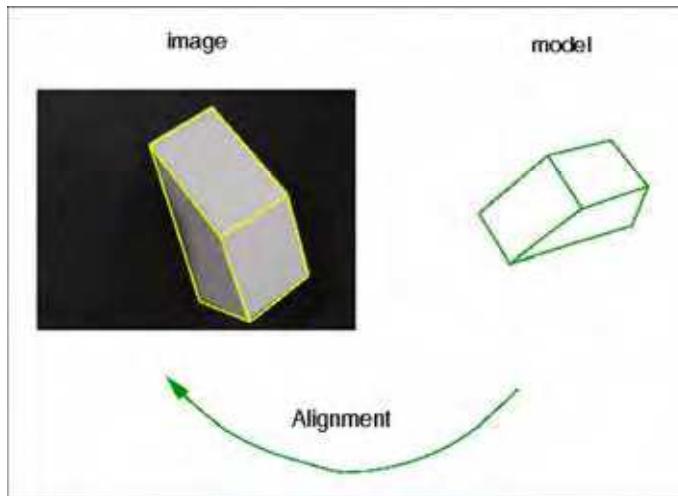


Figure 1. The 2D/3D rigid registration as an alignment technique from a subset of model features (in green) to a subset of visible image features

¹The parameters of the viewpoint are corresponding to the determination of the pose or attitude of the viewed object. It is a imagedToWorld (2-D/ 3-D) rigid registration. For a camera-based vision sensor, it is equivalent to the recovery of extrinsic parameters of a calibrated camera.

The size of these subsets are depending both on the geometric transformation and the number of dof which are constrained by the features used. Features grouping and classification are then important stages, to speed up the correspondence problem by rejecting *a priori* several very probable inconsistent subsets. For instance, the determination of a onedimensional homography between two sets of points requires the grouping of 3 collinear points in both sets, since a onedimensional projective basis is defined with three points on the same line².

2.4 An introduction to the correspondence problem

We introduce the correspondence problem as it is an essential vision component for automate the pose and to detect outliers. A set of n_m 3-D model features and a set of n_l 2-D image features are used through a matching process to determine the "best" Euclidean transformation. Usually n_m and n_l are different for many reasons. A model feature may be occluded or outofthe camera field of view or some irrelevant feature may be detected in the image, because of the acquisition and segmentation processes. These artefacts act as outliers and they must be discarded. Anyway, even if $n_l = n_m$, a featurebased pose determination algorithm needs the onetoone correspondence between subsets of all available features. A naive approach could be done by computing the large set of geometric transformations with all combinations of n_m ordered model features and n_m unordered image points, and to select the transformation with the lowest target registration error (the "best" transformation). However, this is unpractical since the number of combinations is increasing very quickly with the number of features and lead to a cumbersome computing time.

The matching process must be fast and as few timesensitive as possible to a small variation of the number of features involved. That's why, it is of prime importance to design pose algorithms with few features even if the final registration computation takes advantage of all available inliers.

For instance, when one solves the solutions for the pose by means of the wellknown *perspective 3-points problem* (Horaud, 1987; Haralick 1989; DeMenthon 1992; Alter 1994), n_m model points should be matched with n_l image points, considering every arrangement of triples of feature points. Hence, for n_m 3-D model points and n_l image points, there are $A_{n_m}^3 C_{n_l}^3$ candidates (couples of triples) (Huttenlocher, 1990). To register the polyhedra with respect to the camera frame in Figure 1, there are typically $n_m = 8$ vertices and at most $n_l = 7$ visible points in the image. This lead to a large search space since 11760 putative couples of triples should be scanned so as to find the right alignment³. One more image point issuing from artefacts extends the number of candidates up to 18 816, two more image points up to 28 224.

To provide a practical solution, some authors proposed to turn towards the estimation theory. There are many existing robust estimation algorithms and we relate one the most popular family of techniques referred to as the *hypothesizeandtest* approach (Grimson, 1990) where a small set of correspondences are first hypothesized, and the corresponding

² If a fourth point (collinear to the three others) is included in the subset, a wellknown projective invariant can be computed, the crossratio, which is equivalent (but not equal) to the projective coordinate of this extra point in the projective basis

³Considering a calibrated camera, every alignment provides 4 solutions for the pose following the perspective 3points approach, only two with the paraperspective approximation.

transformation is computed. The best known example of this approach is the RANdom SAmple Consensus (RANSAC) algorithm of Fischler and Bolles (Fischler, 1981) which is able to cope with a large amount of outliers and to automatically compute the a geometrical transformation. Another close algorithm is that of Rousseeuw & Leroy (Rousseeuw, 1987) applied by Rosin (Rosin, 1999) for the pose determination.

3. Some pose estimation problems

We begin this section with a short review of wellknown techniques for the pose with feature points and lines in general configuration, so as to orient the discussion toward less wellknown techniques like collinear points, spheres, cylinders and multiple heterogeneous features.

3.1 Pose recovery from points

This problem has been addressed for many years and we suggest the interested reader to look for the following related work. The literature is vaste and the proposed solutions differ from the relative configurations of points (collinear points (Haralick, 1992), planar points (Tsai 1987) or scattered 3-D data (Triggs, 1999)), the number of feature points used (Haralick, 1991; Triggs 99, Nister, 2003), the computational approach (closedform (Dhome, 1989; Haralick, 1989; Horaud, 1989) or numerical iterative approaches (Horaud, 1987; Lowe, 1987; Yuan, 1989; DeMenthon, 1995; Rosin, 1999; David, 2002), concurrent processing (Linnainmaa, 1988), solutions which take advantages of data redundancy (Quan, 1999), the imaging model used (perspective model (Triggs, 1999; Hartley, 2000), weak perspective (Huttenlocher, 1990), paraperspective (Aloimonos, 1990) or orthoperspective (Alter, 1994)).

3.2 Pose recovery from lines

Usually, the extension of the previous related methods to the case of lines is straightforward for some specific arrangements as lines and points are dual entities in a projective plane. For instance, with a non parametric planar curve, it is possible to extract zero curvature points (which are invariants by perspective projection in most cases), but since this kind of points are very unlocalized, one may preferably used tangent lines instead of zero curvature points to match the 2D image features with the object model (Mokhtarian, 1986; Richetin, 1991).

In the case of a polyhedra (see Figure 1 and Figure 3), a 3-D model can be built with a set of 3-D straight lines which are generally not in the same plane. The linepoint duality does not hold any more and a specific method with 3-D lines (4 dof) should be investigated. It's a more difficult problem than that of the pose from points, since the equations which relate the 3-D line representation (see Figure 2) and its perspective projection in the image are quadratic functions of the pose parameters (Hartley, 2000). Dhome et al. (Dhome89), Chen (Chen, 1991b), Liu et al. (Liu, 1990}, Navab & Faugeras. (Navab, 1993), Andreff et al. (Andreff, 2000), Bartoli & Sturm (Bartoli, 2001) and Ansar & Daniilidis (Ansar, 2003) have proposed some pose and displacement algorithms from lines. In (Dhome, 1989), the solution is given by solving a polynomial equation of degree 8 with at least three lines. Chen (Chen, 1991) points out some particular and important arrangements of 3 lines (concurrent lines, three lines with two parallel lines, coplanar lines, perpendicular lines,...) in order to reduce the degree of the polynomial equation, and consequently to reduce the number of solutions. The various approaches are differing from the geometric constraint used (Liu, 1990 and

Chen, 1991), and by the type of representation used like the Plückerian representation for the pose and motion analysis (Navab, 1993; Mitiche, 1995), for large displacements (Bartoli, 2001), with the normalized version of the Plückerian representation (Andreff, 2000) or to get a linear algorithm with large data redundancy (Ansar, 2003).

The perspective projection of a 3-D line Δ represented with the Plücker matrix L (whose components are described with $\mathcal{L} = (r, w)$ with $r^T w = 0$ in the object frame) is the line I^c such that

$$[I^c]_{\times} = P^c L (P^c)^T \quad (1)$$

with

$$L = \begin{bmatrix} [w]_{\times} & -r \\ -r^T & 0 \end{bmatrix} \quad (2)$$

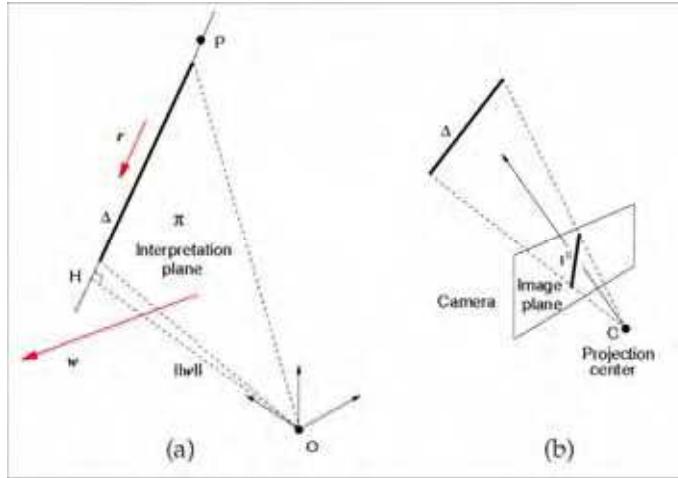


Figure 2. (a) The geometric representation of a 3-D line with the Plücker coordinates $\mathcal{L} = (r, w)$. (b) A backprojection of an image line does not completely defines the 3-D line from which it came from

w is a vector with a direction perpendicular to the interpretation plane (also called pullback plane) which contains the straight line and the origin of the frame (the projection centre C in Figure 2b). It is a suitable representation since one may easily deal with geometrical transformations (Bartoli, 2001) including the perspective projection. The dual of the Plücker matrix, L^* , may also represent the 3-D straight line, with the intersection of two planes. L and L^* are related with a simple rule ($L L^* = 0$) and both representations are commonly used. These two (4×4) matrices are defined up to a scale, skewsymmetric and singular. The rank value (2) is expressing the orthogonality constraint between the two vectors r and w . Moreover, with the derivation of the characteristic polynomial of the Plücker matrix, one can easily show that eigenvalues are complex conjugates scalars ($i \mu, -i \mu, 0, 0$) with the expression $\mu^2 = \|r\|^2 + \|w\|^2$. μ can be arbitrarily set to any nonzero value in order to normalize L .

P^c is the perspectivity, a (3×4) real matrix, composed of the camera parameters matrix K^c , the 3-D rotation R and the position vector t of the object frame w.r.t. the camera frame:

$$P^c = K^c [R \ t] = K^c [(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3), \mathbf{t}] = \begin{bmatrix} \mathbf{p}_1^T \\ \mathbf{p}_2^T \\ \mathbf{p}_3^T \end{bmatrix} \quad (3)$$

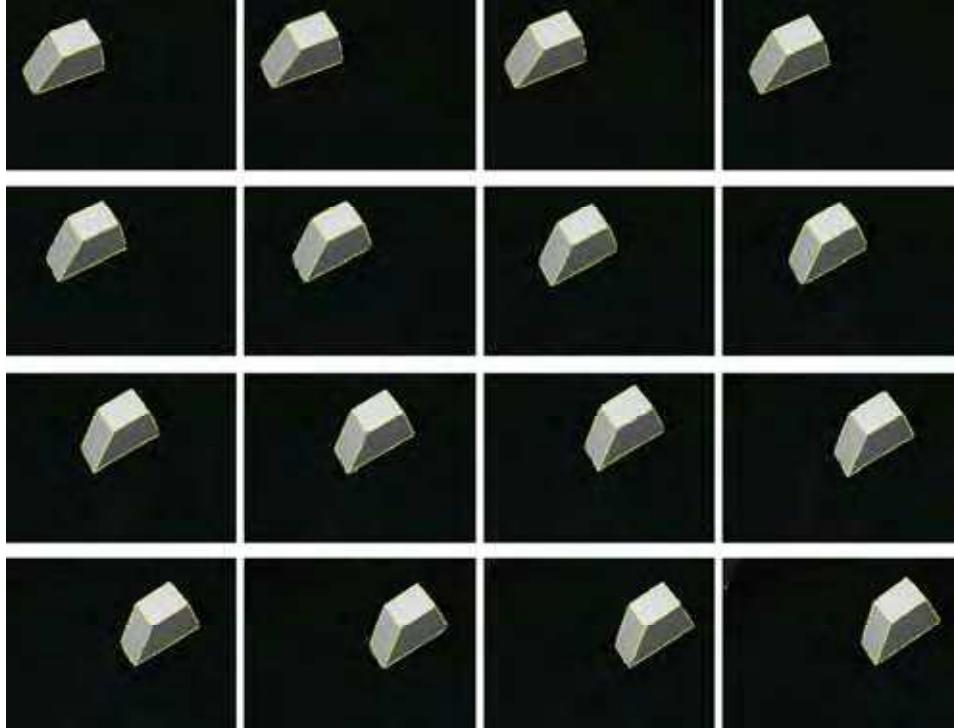


Figure 3. Tracking simple shape in a structured environment. The modelbased polyedral object pose estimation is used to compute the camera displacements during the image sequence

It is now clear that (1) is non linear with respect to the pose parameters $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ and \mathbf{t} . The pose computation needs to solve these parameters given a set of n 3-D lines Δ_i and the corresponding imaged lines \mathbf{l}_i^c . It may be shown from equation (1) applied to the i^{th} line, that we have

$$\left([\mathbf{l}_i^c]_{\times} \otimes (\mathbf{r}_i^T, \mathbf{w}_i^T) \right) \begin{bmatrix} \mathbf{r}_3 \times \mathbf{t} \\ -\mathbf{r}_2 \times \mathbf{t} \\ \mathbf{r}_1 \\ \mathbf{r}_1 \times \mathbf{t} \\ \mathbf{r}_2 \\ -\mathbf{r}_3 \end{bmatrix} = \mathbf{0} \quad (4)$$

where \otimes is the matrix Kronecker product. This is a linear system with respect to the (18×1)

vector of algebraically dependent unknowns, and it can be solved with at least $n = 3$ straight lines.

3.3 Pose recovery from collinear points

In this part, we discuss on a very particular case of pose from points, especially when points are all lying on a common line as it is with fiducial markers or for patterns with structured lighting.

Recovering the relative orientation (2 dof - a unit vector \mathbf{r}) and the position (a 3-vector \mathbf{t}) of a set of n collinear points such as the markers in Figure 4, with respect to the camera frame has been previously investigated by Haralick fifteen years ago (Haralick, 1992). The interpoint distances and the focal length f of the camera are assumed to be known. Haralick solved this problem with a linear algorithm.



Figure 4. Collinear points (centroids of blue markers) stuck on a metallic and cylindrical surface

Let $P_0 = \mathbf{t}$, $P_1 = \mathbf{t} + \lambda_1 \mathbf{r}$, $P_2 = \mathbf{t} + \lambda_2 \mathbf{r}$, ..., $P_{n-1} = \mathbf{t} + \lambda_{n-1} \mathbf{r}$ be n distinct points where λ_i represents the distance between the $i+1^{\text{th}}$ and i^{th} points. The first point P_0 is arbitrarily chosen as the origin ($\lambda_0=0$), hence the perspective projection Q_i of the i^{th} point with homogeneous coordinates $(u_i, v_i, 1)$ is given by

$$\begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \mathbf{t} + \lambda_i \mathbf{r} \end{pmatrix} \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = K^c \begin{pmatrix} \mathbf{t} + \lambda_i \mathbf{r} \end{pmatrix} \quad (5)$$

where K^c is a (3×3) upper diagonal matrix containing the parameters of the camera. From the above equation, Haralick built an homogeneous linear system with a univariate matrix $K^c = \text{diag}(f, f, 1)$ and vectors \mathbf{t} and \mathbf{r} as unknowns

$$\begin{bmatrix} A_r & A_t \end{bmatrix} \begin{bmatrix} r \\ t \end{bmatrix} = 0 \quad (6)$$

A_r and by A_t are two $(2n \times 3)$ real matrices whose components are functions of the camera parameters and the λ 's. A closedform solution can be found with $n > 2$ distinct points. This system may be reformulated as a classical optimization problem with an equality constraint $\|r\|=1$. The solution for r is given by the eigenvector associated with the smallest eigenvalue of the following symmetric matrix $E = A_r^T (I - A_t (A_t^T A_t)^{-1} A_t^T) A_r$ and the position vector t is straightforwardly given by the expression $t = -A_t (A_t^T A_t)^{-1} A_t^T A_r r$. We end up with two estimates for r (a twofold ambiguity in the sign). However, for real objects placed in front of the camera, the third component of vector t must be strictly positive assuming that the camera zaxis (usually, the optical axis) is pointed towards the scene. This leads to the uniqueness of the solution for the pose.

It worth pointing out that in presence of both noisy data and close points in the object pattern, matrices A_r and A_t are illconditioned, which may introduce a significant bias in the results. The use of the least mean squares for $n > 3$ and the lack of data normalization in the original algorithm tend the solution to be sensitive to the matrix condition number. One has to pay attention to data normalization (Hartley, 1997) since the pose estimation may be computed with points not always well scattered. This may also lead to numerical problems. To lower the condition number, it seems advisable to normalize data coordinates with an affine transformation (Trucco, 1998).

3.4 Pose recovery from spheres

To deal with quadratic primitives, we begin with the pose from spheres in this paragraph. The projection of a sphere surface through the central projection is a cone with the vertex at the projection centre (see Figure 5a). The intersection of that cone with the sphere surface is called the contour generator (Γ) whereas intersections with the image plane provide the apparent contour (γ), both are elliptic planar curves in general.

To our knowledge, the mathematical formulation and a solution to the 3-D pose of spherical objects has been firstly proposed by Shin and Ahmad (Shin, 1989). The solution was based on 3-D analytical geometry and a closedform solution is given. SafaeRad et al. (SafaeRad, 1992) have also studied spherical objects pose in the context of mobile robotics and they have pointed out some major practical limitations in the pose accuracy, like the location of edges used in the ellipse parameters fitting, the uncertainty of intrinsic scale factors (due to timing mismatches that occur between camera scanning hardware and image acquisition hardware) and also the radial distortion of the lens. Ferri et al. (Ferri, 1993) present some algorithms for linear and quadratic primitives which include the computation of the 3-D pose and in particular a quadric of revolution. They provide a simple pose recovery procedure from the eigendecomposition of the matrix representation and they mentioned that in the case of a sphere, it must have two equal eigenvalues. Pose determination and camera calibration by means of images of spherical objects has also been studied, in particular by Teramoto and Xu (Teramoto, 2002), by Agrawal and Davis (Agrawal, 2003), by Dhome et al. (Dhome, 1990) and by Daucher et al. (Daucher, 1994).

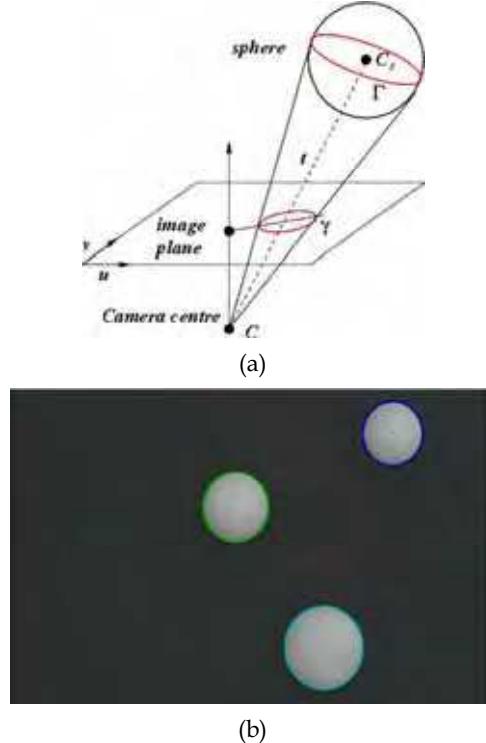


Figure 5. (a) A sphere, its contour generator (Γ) and its apparent contour (γ) in image plane is a conic. (b) Each detected ellipse is represented by a matrix E^* related to the 3-D position of the sphere centre

A sphere S is a quadric, and with the homogeneous coordinates of any point M on the sphere surface, it can be represented by a (4×4) symmetrical matrix S as $M^T S M = 0$. The matrix S depends on the radius r_s and the sphere position vector $t = (t_x, t_y, t_z)^T$ from its centre to the world reference frame. When the camera reference frame coincides with the world reference frame, t_z is chosen in the direction perpendicular to the image plane, and pointing towards the scene. When expressed in that frame, it is provided by

$$S(r_s, t) = \begin{bmatrix} I & -t \\ -t^T & \|t\|^2 - r_s^2 \end{bmatrix} \quad (7)$$

where I is the identity matrix. Since $\|t\|$ is the distance between the sphere centre and the projection centre, C , the scalar $\|t\|^2 - r_s^2$ must be positive as it is for a sphere placed in front of the camera, taken into account its own size. The dual of the sphere S is a sphere represented by the adjoint S^* of S since it is a symmetric matrix. It is given by $S^* = S^{-1}$. Both matrices S and S^* will be used hereinafter since a sphere and its image are easily related by dual matrices.

With the pinhole camera model and homogeneous coordinates, a 3-D point M is projected onto the image plane in a 2-D point m such that

$$\mathbf{m} = \mathbf{P}^c \mathbf{M} \quad (8)$$

with

$$\mathbf{P}^c = \mathbf{K}^c \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (9)$$

where the (3×4) camera matrix \mathbf{P}^c is a perspectivity (or camera matrix) and λ is a non null scalar. When expressed in the camera frame with the projection centre \mathbf{C} as the origin, the camera matrix is of the form given by (8). Under the camera matrix \mathbf{P}^c , the outline of the sphere S is an ellipse E in the image which can be represented with a (3×3) symmetric matrix E . The dual of E is the adjoint E^* which can be related to the adjoint of S as follow

$$E^* = \mathbf{P}^c S^* (\mathbf{P}^c)^T \quad (10)$$

By substituting equations (7) and (9) in (10), it has been shown by Agrawal (Agrawal, 2003) that

$$(\mathbf{K}^c)^{-1} E^* (\mathbf{K}^c)^{-T} = I - \frac{1}{r_s^2} \mathbf{t} \mathbf{t}^T, \quad \mu \in \mathbb{R}^* \quad (11)$$

The work of Teramoto et al. (Teramoto, 2002) is a pose determination method in which the position direction and the size of a ball is derived in the case of known intrinsic parameters. It is extended here to the recovery of the full 3-D position with a slightly modification of the original work which will serve to analyze the eigendecomposition in presence of noisy data. Given the dual matrix E^* , let us denote with $\mathbf{t}_u = \mathbf{t}/s$, a unit vector with a nonnull scalar ($s = \pm \|\mathbf{t}\|$). Starting from equation (7), we have:

$$Q^* = I - \frac{s^2}{r_s^2} \mathbf{t}_u \mathbf{t}_u^T \quad (12)$$

with $Q^* = (\mathbf{K}^c)^{-1} E^* (\mathbf{K}^c)^{-T}$. Since s/r_s is always greater 1, the righthandside of the above equation is a rank-3 matrix, it can be written as:

$$[\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_u] \text{diag} \left(1, 1, 1 - \frac{s^2}{r_s^2} \right) [\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_u]^T \quad (13)$$

and the lefthandside can be decomposed as $\mathbf{U} \text{diag}(\lambda_1, \lambda_2, \lambda_3) \mathbf{U}^T$ with $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3]$ is an orthonormal matrix. It's clear that we have

$$1 - \frac{s^2}{r_s^2} = \lambda_3 / \lambda_2 \quad (14)$$

and

$$\mathbf{t} = s \mathbf{U}_3 \quad (15)$$

The solution is unique since the sign for t_z ($t_z > 0$) reveals the sign for the scalar s . It is worth pointing out that in one hand the symmetrical matrix E^* has 5 independent components as it represents an ellipse in the image. In the other hand, the symmetrical matrix $I - \mathbf{t} \mathbf{t}^T / r_s^2$ has two eigenvalues equal to one (clearly $\mathbf{t} \mathbf{t}^T$ is a rank-1 matrix), hence representing a

"calibrated ellipse". So, it's clear that the geometric information issued from equation (11) has not been fully exploited. This means that either the position vector \mathbf{t} can be solved with an overdetermined system (with data redundancy) or other parameters (like intrinsic parameters) may be determined from a unique sphere and its corresponding ellipse in a single image. Let us now consider the former case. Following this remark, a simple improvement of the Teramoto's method consists of a slightly modification of equation (14) since the two singular values λ_1 and λ_2 must be equal with uncorrupted data. Thus, we propose to replace equation (14) to by

$$1 - \frac{s^2}{r_s^2} = 2\lambda_3 / (\lambda_1 + \lambda_2) \quad (16)$$

that is λ_1 is replaced by the midvalue of the first two singular values in presence of noise. The resulting matrix Q_m^* is then the closest symmetrical matrix to Q^* with the Frobenius norm and is given by $Q_m^* = U \text{diag}([\lambda_1 + \lambda_2]/2, [\lambda_1 + \lambda_2]/2, \lambda_3] U^T$. Once the matrix Q_m^* has been derived, the (4×4) matrix \hat{S} may be computed with the estimated position vector $\hat{\mathbf{t}}$

$$Q_m^* = I - \frac{1}{r_s^2} \hat{\mathbf{t}} \hat{\mathbf{t}}^T$$

thanks to the Teramoto's method, that is with

Simulation results reported in Figure 6 show a better behaviour with respect to noise for the modified version we propose compared to the original algorithm. It is simple to implement and it does not require more computations. It has been used as an efficient tool while tracking the 3-D position of a moving camera mounted on a wheeled robot for robotic competition (see Figure 7).

Windowing techniques are used to define the search space of the area of interest (the closest red ball). The diagonal of the inner blue square that area is corresponding to the estimated depth, whereas the blue cross are corresponding to the position of the centre. (add at the end of figure caption) of the centre.

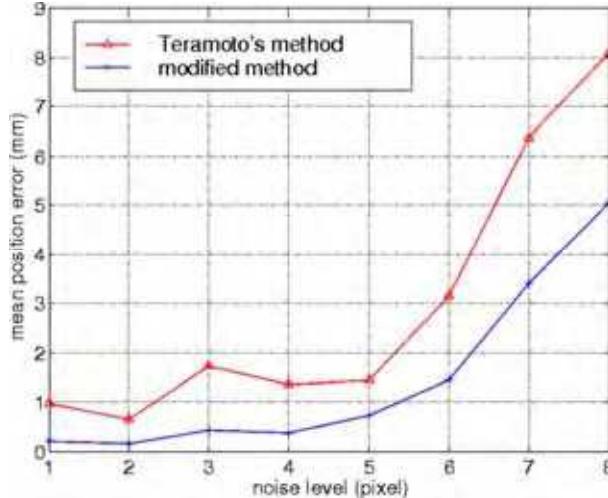


Figure 6. Mean position errors with the Teramoto's method (red) and with the modified version (blue) with respect to varying noise level



Figure 7. 3-D tracking of red balls. A camera is mounted on a wheeled robot which is moving in the ground plane (basketball robot players). Given the radius, red balls should be tracked and picked up

3.5 Pose estimation from cylinders

To go ahead with quadratic primitives, we now discuss on the pose from cylinders, especially with straight homogeneous circular cylinders (SHCC), that is the class of cylinders with a straight axis and a circular section with constant radius (see Figure 8).

In the late 80s and early 90s, shape from contour approaches have been developed in an attempt to determine constraints on a threedimensional scene structure based on different assumptions about the shape. The understanding of the relations between image contours geometry and the shape of the observed object and the viewing parameters is still a challenging problem and it is essential that special shapes are not represented by freeform surfaces without regard to their special properties, but treated in a way more appropriate to their simple nature. Explicit relations from occluding contours to the model of a curved threedimensional object have been presented for objects with geometrical properties such as generalized cylinders or surfaces of revolution (Dhome, 1992; Ferri, 1993; Kriegman, 1990 ; Ponce, 1989).

More recent works are based on the image contour of a cylinder crosssection when it is visible. Puech et al. (Puech, 1997) used the image of two crosssections to locate a straight uniform generalized cylinder in 3-D space and Shiu and Huang (Shiu, 1991) solve the problem for a finite and known cylinder height, that is a 3-D pose determination for 5 degrees of freedom. SafaeRad et al. (SafaeRad, 1992) estimate the 3-D circle centre and orientation from the projection of one of the two circles on the cylinder ends. However, ellipse fitting generally becomes inaccurate when the cylinder radius is small with respect to the cylinder height and also since both circles on the cylinder are not completely visible. Huang et al. (Huang, 1996) solve the pose determination of a cylinder through a reprojection transformation which may be thought as a rectification transformation. The computed transformation is applied to the image of the cylinder and brings the camera optical axis to perpendicularly intersect the cylinder axis, which is then parallel to one of the two image axes. The new image (called "canonical" image) is a symmetrical pattern which simplify the computation of the pose. It is an interesting method which provides an analytical solution of the problem, including the recovery of the height of the cylinder. However, it's requiring an image transformation and errors for estimating the reprojection transformation may lead to significant bias in the contours location of the resulting canonical image and consequently to the pose parameters. In a similar way, Wong et al. (Wong, 2004) take advantage of the invariance of surface of revolution (SOR) to harmonic homology and have proposed to recover the depth and the focal length (by assuming that the principal point is located at the

image center and that the camera has unit aspect ratio) from the resulting silhouette which exhibits a bilateral symmetry. It is also a rectification which brings the revolution axis to coincide with the y -axis of the image. If the image of a latitude circle in the SOR can be located, the orientation of the revolution axis can also be estimated.

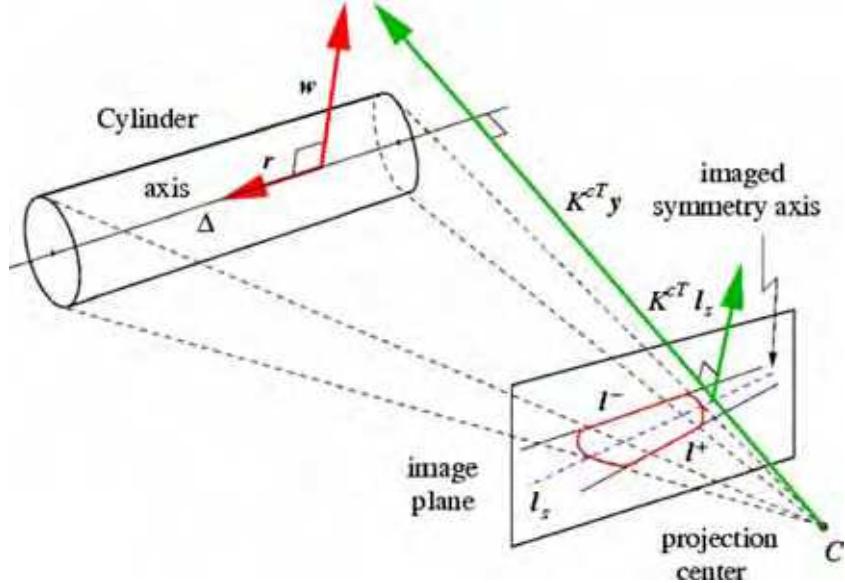


Figure 8. A straight homogeneous circular cylinder and its perspective projection. The backprojection of apparent lines (l^- , l^+) is a couple of planes $(P^c)^T l^-$ and $(P^c)^T l^+$ passing through the centre. The image of the cylinder axis Δ is the axis l_s of the harmonic homology H relating the two apparent lines

Some other approaches based on contours and shading have also been proposed. Asada et al. (Asada, 1992) provide a technique to recognize the shape of a cylinder crosssection and to determine the orientation axis under the weak Lambertian assumption for the reflectance of the object surface when the surface does not include specular component. Caglioti and Castelli (Caglioti, 1999) focus their work rather on metallic surfaces (cylinders and cones) and they recover the pose parameters by means of the axial symmetric reflection model. However, although methods which do not solely involve geometric features should be more investigated to achieve an accurate pose estimation, these two latter methods assume an orthographic projection for the camera model, a too strong approximation when cylinder orientation is partly embedded in the perspective effect. It has been shown in (Doignon, 2007) that the estimation of the Plücker coordinates (r, w) of the cylinder axis can be directly recovered from the degenerate conic $C = l^- (l^+)^T + l^+ (l^-)^T$ built with the two apparent lines (l^- , l^+) and the cylinder's radius r_c with a calibrated camera. In one hand, we have

$$(K^c)^T C K^c = U D U^T = U \text{ diag} \begin{pmatrix} \lambda_1 & \lambda_2 & 0 \end{pmatrix} U^T \quad (17)$$

and in the other hand (after some computations, see (Doignon, 2007) for details).

$$\begin{aligned} (\mathbf{K}^c)^T \mathbf{C} \mathbf{K}^c &\equiv (\mathbf{I} - \mathbf{r} \mathbf{r}^T - \mathbf{z} \mathbf{z}^T) = [\mathbf{a} \ \mathbf{z}_u \ \mathbf{r}] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 - \sigma^2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{a}^T \\ \mathbf{z}_u^T \\ \mathbf{r}^T \end{bmatrix} \end{aligned} \quad (18)$$

with $\mathbf{z} = (\sqrt{1 - \alpha^2}/\|\mathbf{w}\|) \mathbf{r} \times \mathbf{w}$, $\alpha = r_c/\sqrt{\|\mathbf{w}\|^2 - r_c^2}$ and the unit vector $\mathbf{z}_u = \mathbf{z}/\sigma$. It is easy to see that $\mathbf{U} = [\mathbf{a} \ \mathbf{z}_u \ \mathbf{r}]$, $\mathbf{w} \equiv \mathbf{z}_u \times \mathbf{r} = \mathbf{a}$ and finally $\|\mathbf{w}\| = r_c \sqrt{1 + \lambda_1/\lambda_2}$.

Some results, illustrated in Figure 9 and Figure 10, show the efficacy of the proposed fitting and pose determination. Firstly, the pose determination is performed in a simple and controlled environment (Figure 9). Second, in a complex environment with a moving background (the abdomen of a pig), a cylindricalshaped laparoscopic instrument is detected (thanks to a joint huesaturation colour attribute). Plücker coordinates are computed with the method described above. It is shown in (Doignon, 2007) that the direction of vector \mathbf{w} is directly related to the image of the cylinder's axis.



Figure 9. The 3-D tracking of the cylinder's axis through the detection of the conjugate apparent lines in a uniform background

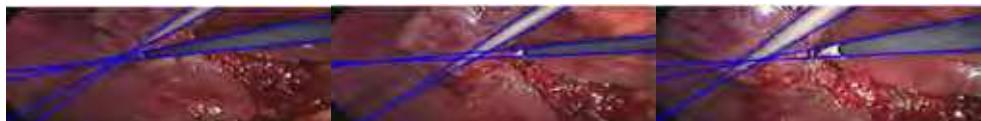


Figure 10. The 3-D of the cylinder's axis of cylindricalshaped laparoscopic instruments through the detection of the conjugate apparent lines in a complex environment (abdomen of a pig)

3.6 Pose estimation with multiple types of geometrical features

We now turn to a little bit towards more complex environments. As some features may be occluded during the tracking, it is necessary to estimate the pose with multiple features. We introduce this approach with an example in Figure 11. The virtual visual servoingbased pose estimation (see the end of paragraph 2.1) is carried out with three geometrical features : a cylinder, a circular needle and marker points. Only four degrees of freedom are necessary to estimate the attitude of the instrument axis. The 4 dof of the pose can be determined using the contour generator and its image (the apparent contour) of the cylinder (see paragraph 3.5). However, the positions of the marking spots not only define the proper rotations and translations, but also give information on the orientation and position of the axis of the

shaft. We then chose to estimate the 6 dofs of the instrument with all the available features. This can be done with analytical methods using both the apparent contours and one known point at the cylinder's surface (Nageotte, 2006).

The full pose estimation is interesting for robustness considerations only if all the available information given by the apparent lines and all the spots is used. To this purpose, the Virtual Visual Servoing (VVS) due to Sundareswaran (Sundareswaran, 1998) and also by Marchand (Marchand, 2002) and may handle the information redundancy. VVS is a numerical iterative method for minimizing the error between the extracted features and the forward projection of the object in the images and based on the imagebased visual servoing (IBVS) schemes. This process needs the computation of an interaction matrix which relates the variations of each image feature and the camera velocity screw τ . With these image features, the full interaction matrix L_s has the following generic form:

$$\begin{pmatrix} \dot{\mathbf{l}}^+ \\ \dot{\mathbf{l}}^- \\ \dot{p}_1 \\ \dot{p}_2 \\ \vdots \\ \dot{p}_n \\ x_e \\ y_e \\ r_{min} \\ r_{max} \\ \alpha_e \end{pmatrix} = \begin{pmatrix} L_{line}(\mathbf{l}+) \\ L_{line}(\mathbf{l}-) \\ L_{pt}(p_1) \\ L_{pt}(p_2) \\ \vdots \\ L_{pt}(p_n) \\ L_{ellipse} \end{pmatrix} \tau = L_s \tau \quad (19)$$

The interaction matrices associated to a point p , L_{pt} , to a line \mathbf{l} , L_{line} , and associated to an ellipse $E=(x_e, y_e, r_{min}, r_{max}, \alpha_e)$, $L_{ellipse}$, can be found in Espiau et al. (Espiau, 1992) or in Chaumette et al. (Chaumette, 1993). In order to guarantee a fast convergence and a good stability of the VVS, it is useful to initialize the algorithm close enough to the real pose. For this purpose, we use the modified version of the Haralick's method and the DeMenthon iterative method (DeMenthon, 1995) for points, that of Dhome (Dhome, 1992) to get the 4 solutions for the pose of a circle and the pose determination of the axis of a circular cylinder described at the paragraph 3.5. With this initial pose parameters estimates for the attitude of the camera with respect to the object of interest (a laparoscopic surgical instrument, see Figure 11), the following control law is applied to the virtual camera

$$\tau = -G \widehat{L}_s^+ (\mathbf{s} - \mathbf{s}^*) \quad (20)$$

until the control vector becomes smaller than a specified value. The process converges quickly towards the real pose of the camera (see Figure 11).

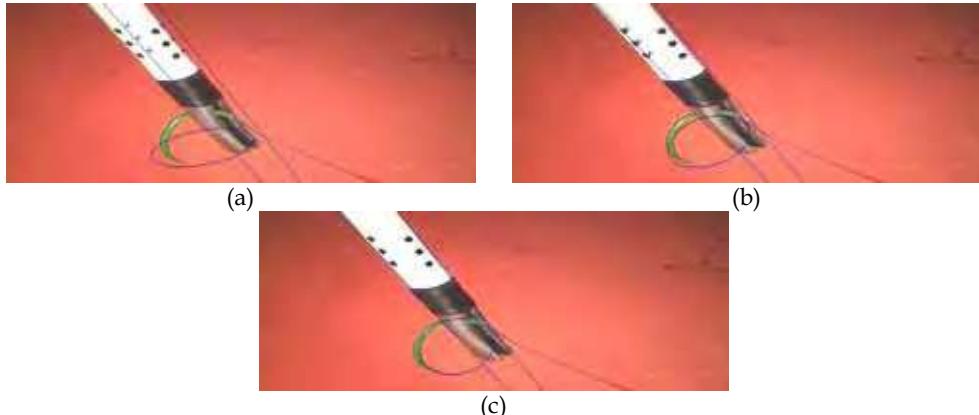


Figure 11. The pose estimation as a Virtual Visual Servoing process with multiple geometric features (apparent lines, markers and a circular needle). (a) The blue lines are the projections with the initial virtual camera position. (bc). The projections when the error vector ($s - s^*$) tends to 0

4. Further Readings : Pose estimation with multiple cues

We close the chapter by touching upon the difficult problem of the pose estimation and tracking in a complex and unknown environment. In many situations, like for assistance domestic environments, outdoor navigation or tracking inside the human body, the observed scene is unstructured and the background is not uniform nor constant. It is then not a trivial task to detect object of interests or patterns, since brightness and colour are changing and the background moves due to human displacements, the wind or the breathing or heart beating in the third case, leading to shadows, occlusions or specularities. Some rather recent works integrate multiple visual cues like colour (Vincze, 2005), texture (Pressigout, 2006) or global&local descriptions (Kragic, 2002) or a learning stage (Vacchetti, 2004) to improve the tracking process.

The Vision for Robotics (V4R) software package proposed by Vincze et al. (Vincze, 2005) integrates multiple cues like edge gradient, colour, intensity, topological interrelations among features and pose from preceding frames to provide an efficient visual modelbased tracking tool in realistic and unconstrained environments. These cues are derived not only from the images but also from object parameters (the model) and pose information stored within previous tracking cycles. A fourstage tracking scheme is designed, from the 2-D feature extraction to the pose computation and validation.

For domestic environment (living room), Kragic & Christensen propose to use both the appearance and geometrical models to estimate the position and orientation of the object relative to the camera/robot coordinate system. Following a threestage strategy (initialization, pose estimation, tracking), this system may be seen as a coarsesetofine tracking. The initialization step provides an approximation to the current object pose by means of the Principle Components Analysis. Once the model is found, a local fitting is used to extract linear primitives from which the pose is computed. Finally, if the object or the eyeinhand

robot start to move, the Drummond's method (Drummond & Cipolla, 2000) is adopted to provide a realtime estimate of the object pose.

Vacchetti et al. (Vacchetti, 2004) have formulated the tracking problem using a single camera in terms of local bundle adjustment and have developed an image correspondences method that can handle short and widebaseline matching. The video information of a very limited number of reference images created during a training stage is merged with that of preceding frames during the tracking. The tracking process needs a 3-D model of any object that can be represented by a 3-D mesh. Thus, keyframes serve to register an incoming image by means of a set of extracted corners thanks to the minimization of the reprojection error with the Tukey Mestimator. The reported results have demonstrated a very good robustness with respect to aspect changes, model inaccuracies, partial occlusions, focal length and scale changes, and illumination changes.

Pressigout and Marchand (Pressigout, 2006) propose a realtime hybrid 3-D tracking with the integration of the texture information in a nonlinear edgebased pose estimation algorithm. Pose and camera displacements are formulated in terms of a full scale nonlinear optimization instead of a point of interestbased pose estimation. In particular, the camera displacement estimation is based on a two images intensity matching. A non linear criterion based on intensity mapping error is defined to that purpose from which an interaction matrix is derived.

5. Conclusion

With this article, we have addressed some issues in pose estimation with geometrical features and a modelbased approach in the context of monocular vision. While the 3-D tracking/estimation may be performed with optimal estimators (with the Kalman filter and its extended/nonlinear versions or with the particle filter also referred to as the sequential Monte Carlo method), system models and state vectors need the pose parameters recovery or the 3-D motion recovery from the motion field. The pose determination is needed for applications with high accurate 3-D positioning requirements, when occlusions, shadows or abrupt motions have to be handle. To this purpose, several geometrical featurebased approaches have been reviewed for solving the pose with various constrained degrees of freedom.

Cue integration leads to robustness and automatic measurement of scene complexity. This is of prime importance to exploit the video information captured in a complex environment with dynamical changes. Composite features, colour, edges, texture integration are some additional data which have brought significant improvements of the tracker's behaviour thanks to robust and Mestimators. This is a key factor of success for a visionbased module with some autonomous capabilities, hence for the achievement of some visionbased (semi) autonomous tasks.

6. References

- Agrawal, M. & Davis, L.S. (2003). Camera calibration using spheres: A semidefinite programming approach, *Proceedings of the IEEE Int'l Conf. on Computer Vision*, Vol. 2, p. 782789, Nice, France
- Aloimonos, Y. (1990). Perspective approximations, *Image and Vision Computing*, Vol. 8, No 3, p. 177192, Elsevier Science

- Alter, T.D. (1994). 3-D Pose from 3 Points Using WeakPerspective, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, No 8, p. 802808, IEEE Computer Society Press
- Andreff, N.; Espiau, B. & Horaud, R. P. (2000). Visual Servoing from Lines, *Proceedings of the IEEE Int'l Conf. on Robotics and Automation*, p. 20702075
- Ansar, A. & Daniilidis, K. (2003). Linear Pose Estimation from Points or Lines. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No 5, p. 578589, IEEE Computer Society Press
- Asada, M.; Nakamura, T. & Shirai, Y. (1992). Weak Lambertian Assumption for Determining Cylindrical Shape and Pose from Shading and Contour, *Proceedings of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, p. 726729, IEEE Computer Society
- Bartoli, A. & Sturm, P. (2001). The 3-D line motion matrix and alignment of line reconstructions, *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, Hawaii, USA. p. 287292, IEEE Computer Society Press NY
- Bolles, R.C. & Cain, R.A. (1982). Recognizing and locating partially visible objects, the localfeaturefocus method, *Int'l Journal of Robotics Research*, Vol. 1, No 3, p. 5782, SAGE Publications, Thousand Oaks
- Bolles, R.C. & Horaud, R.P. (1986). 3-DPO: A threedimensional part orientation system, *Int'l Journal of Robotics Research*, Vol. 5, No 3, p. 326, SAGE Publications, Thousand Oaks
- Caglioti, V. & Castelli, E. (1999). Recovering cylindric and conic surfaces from contours and reflections, *Pattern Recognition Letters*, Vol. 20, p. 367382, Elsevier Science
- Chaumette, F.; Rives, P. & Espiau, B. (1993). Classification and realization of the different visionbased tasks. In *Visual Servoing*, K. Hashimoto (ed.), Vol. 7, p. 199228, World Scientific Series in Robotics and Automated Systems, Singapore
- Chen, H. (1991). Pose Determination from linetoplane correspondences: Existence of solutions and closedform solutions, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No 6, p. 530541, IEEE Computer Society Press
- Daucher, N.; Dhome, M. & Lapresté, J.-T. (1994). Camera Calibration From Spheres Images, *Proceedings of the third European Conference on Computer Vision*, Vol. 1, p. 449454, Stockholm, Sweden
- David, P.; DeMenthon, D. & Duraiswami, R. & Samet, H. (2002). SoftPOSIT: Simultaneous Pose and Correspondence Determination, *Proceedings of the European Conference on Computer Vision*, Copenhagen, Denmark, SpringerVerlag
- DeMenthon, D. & L.S. Davis, L.S. (1992). Exact and Approximate Solutions of the PerspectiveThreePoint Problem, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No 11, p. 11001105, IEEE Computer Society Press
- Dementhon, D. & Davis, L.-S. (1995). ModelBased Object Pose in 25 lines of Code, *Int'l Journal of Computer Vision*, Vol. 15, No 1, p. 123141, Kluwer Academic Publishers
- Dhome, M.; Richetin, M. & Lapresté, J.-T. & Rives, G. (1989). Determination of the attitude of 3-D objects from a single perspective view, *IEEE Transactions. on Pattern Analysis and Machine Intelligence*, Vol. 11, No 12, p. 12651278, IEEE Computer Society Press
- Dhome, M; Lapresté, J.-T. & Rives, G. & Richetin, M. (1992). Spatial localization of modelled objects of revolution in monocular perspective vision, *Proceedings of the European Conference on Computer Vision*, Antibes, France
- Dickmanns, E. D. & Graefe, V. (1988). Dynamic monocular machine vision, *Machine Vision Applications*, Vol. 1, p. 223240

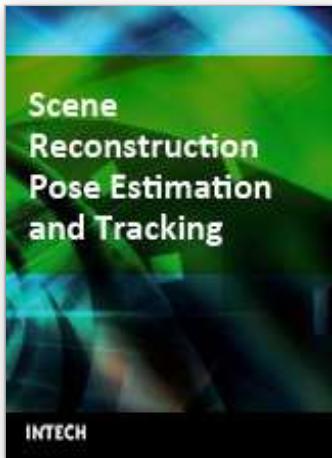
- Doignon C. & Abba G. (1999). A practical multiplane method for a lowcost camera calibration technique, *Proceedings of the fifth European Control Conference*, Karlsruhe, Germany, SpringerVerlag
- Doignon, C. & De Mathelin, M. (2007). A Degenerate ConicBased Method for a Direct Fitting and 3-D Pose of Cylinders with a Single Perspective View, To appear in the proceedings of the IEEE Int'l. Conf. on Robotics and Automation, Roma, Italy
- Drummond, T. & Cipolla, R. (2000). RealTime Tracking of Multiple Articulated Structures in Multiple Views. In *proceedings of the European Conference on Computer Vision*, Vol. 2, p. 2036, SpringerVerlag, Dublin, Ireland
- Espiau, B.; Chaumette, F. & Rives, P. (1992). A new approach to visual servoing in robotics. *IEEE Transactions on Robotics and Automation*, Vol. 8, No 3, p. 313326
- Faugeras, O.D. & Toscani, G. (1987). Camera Calibration for 3-D Computer Vision, *Proceedings of the Int'l Workshop on Machine Vision and Machine Intelligence*, Tokyo, Japan
- Ferri, M.; Mangili, F. & Viano, G. (1993). Projective Pose Estimation of Linear and Quadratic Primitives in Monocular Computer Vision, *Computer Vision, Graphics and Image Processing: Image understanding*, Vol. 58, No 1, p. 6684, Elsevier Science
- Fischler, M.A. & Bolles, R.C. (1981). Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography, *Communications of the ACM*, Vol. 24, No 6, p. 381395
- Gold, S.; Rangarajan, A. & Lu, C.P. & Pappu, S. & Mjolsness, E. (1998). New Algorithms for 2-D and 3-D Point Matching: Pose Estimation and Correspondence, *Pattern Recognition*, Vol. 31, p. 10191031, Elsevier Science
- Grimson, W. E. L. (1990). *Object Recognition by Computer: The Role of Geometric Constraint*, MIT Press, Cambridge, MA. ISBN 0262071304
- Grimson, W.E.L. & Huttenlocher, D.P. (1991). On the Verification of Hypothesized Matches in ModelBased Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No 12, p. 12011213, IEEE Computer Society Press
- Hager, G. D. & Toyama, K. (1998). The Xvision system: A portable substrate for realtime vision applications, *Computer Vision and Image Understanding*, Vol. 69, No 1, p. 2337, Elsevier Science
- Hansen, C. & Henderson, T.C. (1989). CAGDbased computer vision, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No 11, p. 11811193.
- Haralick, R. M. & Chu, Y.H. & Watson, L.T. & Shapiro, L.G. (1984). Matching wire frame objects from their 2-D perspective projection, *Pattern Recognition*, Vol. 17, No 6, p. 607619, Elsevier Science
- Haralick, R. M. (1989). Monocular vision using inverse perspective projection geometry: Analytic relations, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, p. 370378, IEEE Press
- Haralick, R. M.; Lee, C. & Ottenberg, K. & Nölle, M. (1991). Analysis and Solutions of the ThreePoint Perspective Pose Estimation Problem, *IEEE Conf. Computer Vision and Pattern Recognition*, p. 592598, Maui, Hawaï, USA, IEEE Computer Society Press
- Haralick, R. M. & Shapiro, L.G. (1992). *Computer and Robot Vision*, vol. 1, AddisonWesley Publishing, ISBN 0201108771
- Harris, C. G. & Stennet, C. (1990). RAPID A video rate object tracker, *Proceedings of the British Machine Vision Conf.*, p. 233236

- Hartley, R. (1997). In defense of the eightpoint algorithm, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No 6, p. 580593, IEEE Computer Society Press
- Hartley, R. & Zisserman, A. (2000). *Multiple view geometry in computer vision*, Cambridge University Press, UK. ISBN 0521623049
- Horaud, R.P. (1987). New Methods for Matching 3-D Objects with Single Perspective Views, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 9, No 3, p. 401412, IEEE Computer Society Press
- Horaud, R.P.; Conio, B. & Leboulleux, O. & Lacolle, B. (1989). An Analytic Solution for the Perspective 4Point Problem, *Computer Vision, Graphics, and Image Processing*, vol. 47, p. 3344, Elsevier Science
- Horn, B.K.P. (1986). *Robot Vision*, MIT Press, McGrawHill, Cambridge. ISBN 0262081598
- Huang,J.-B.; Chen, Z. & Chia, T.-L. (1996). Pose determination of a cylinder using reprojection transformation, Vol. 17, p. 10891099, *Pattern Recognition Letters*, Elsevier Science
- Hutchinson, S.; Hager, G. D. & Corke, P. I. (1996). A tutorial on visual servo control, *IEEE Transactions on Robotics and Automation*, Vol. 12, No 5, p. 651670
- Huttenlocher, D. P. & Ullman, S. (1990). Recognizing Solid Objects by Alignment with an Image, *Int'l. Journal of Computer Vision*, Vol. 5, No 2, p. 195212, Kluwer Academic Publishers
- Ikeuchi, K. (1987). Generating an interpretation tree from a CAD model for 3-D object recognition in binpicking tasks, *Int'l Journal of Computer Vision*, Vol. 1, No 2, p. 145165, Kluwer Academic Publishers
- Jurie, F. (1999). Solution of the Simultaneous Pose and Correspondence Problem Using Gaussian Error Model, *Computer Vision and Image Understanding*, Vol. 73, No 3, p. 357373, Elsevier Science
- Kragic, D & Christensen, H.I. (2002). Modelbased Techniques for Robotic Servoing and Grasping, *Proceedings of the IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems*, p. 299304, ISBN 0780373987, Lausanne, Switzerland.
- Kriegman, D.-J. & Ponce, J. (1990). On Recognizing and Positioning Curved 3-D Objects from Image Contours, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, No 12, p. 11271137, IEEE Computer Society Press
- Lamdan, Y. & Wolfson, H.J. (1988). Geometric hashing: A general and efficient modelbased recognition scheme, *Proceedings of the IEEE Int'l Conf. on Computer Vision*, p. 238249, Tampa, FL, USA
- Linnainmaa, S. ;Harwood, D. & Davis, L. S. (1988). Pose Determination of a ThreeDimensional Object Using Triangle Pairs, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10, No 5, p. 634647, IEEE Computer Society Press
- Liu, Y.; Huang, T.S. & Faugeras, O.D. (1990). Determination of camera location from 2-D to 3-D line and point, *IEEE Transactions on Pattern analysis and Machine Intelligence*, Vol. 12, No 1, p. 2837, IEEE Computer Society Press
- Lowe, D.G. (1987). Threedimensional object recognition from single twodimensional images, *Artificial Intelligence*, Vol. 31, No 3, p. 355395, Elsevier Science

- Marchand, E. & Chaumette, F. (2002). Virtual visual servoing: a framework for realtime augmented reality, *Proceedings of the EUROGRAPHICS Conference*, Vol. 21 (3 of Computer Graphics Forum), p. 289298, Saarbrücken, Germany
- Marchand, E & Chaumette, F. (2004). Feature tracking for visual servoing purpose. *Proceedings of the Int'l. Workshop on Advances in Robot Vision: From Domestic Environments to Medical Applications, in conjunction with the IEEE:RSJ Int'l. Conf. on Intelligent Robots and Systems*, p. 1020, Sendaï, Japan.
- Marchand, E.; Spindler, F. & Chaumette, F. (2005). ViSP for visual servoing: a generic software platform with a wide class of robot control skills. *IEEE Robotics and Automation Magazine*, Vol. 12, No 4, p. 4052, ISSN 10709932
- Meng, X & Hu, Z. (2003). A new easy camera calibration technique based on circular points, *Pattern Recognition*, Vol. 36, p. 11551164, Elsevier Science
- Mitiche, A. (1995). *Computational Analysis of Visual Motion*, Advances in Computer Vision and Machine Intelligence Series, ISBN 9780306447860, SpringerVerlag
- Mokhtarian, F. & Mackworth, A. (1986). Scalebased Description and Recognition of Planar Curves and TwoDimensional Shapes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, No 1, p. 3443, IEEE Computer Society Press
- Nageotte, F; Zanne, P. & Doignon, C. & de Mathelin, M. (2006). Visual ServoingBased Endoscopic Path Following for RobotAssisted Laparoscopic Surgery, *Proceedings of the IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems*, Beijin, China
- Navab, N. & Faugeras, O. D. (1993). Monocular Pose Determination from Lines : Critical sets and Maximum Number of Solutions, *IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, p. 254260, IEEE Computer Society Press
- Navab, N; Vieville, T. & Faugeras, O. D. (1993). The Critical Sets of Lines for Camera Displacement Estimation: A Mixed EuclideanProjective and Constructive Approach, *Proceedings of the IEEE Int'l Conf. on Computer Vision*, p. 713723, IEEE Computer Society Press
- Nister, D. (2003). An efficient solution to the fivepoint relative pose problem, *Proceedings of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, Vol. 2, p. 195202, IEEE Computer Society Press, Madison, WI, USA
- Olson, C.F. (2001). A General Method for Geometric Feature Matching and Model Extraction, *Int'l Journal of Computer Vision*, Vol. 45, No 1, p. 3954, Kluwer Academic Publishers
- Papanikolopoulos, N. P.; Khosla, P. K. & Kanade, T. (1993). Visual Tracking of a Moving Target by a Camera Mounted on a Robot: A Combination of Control and Vision, *IEEE Transactions on Robotics and Automation*, Vol. 9, No 1, p. 1435
- Pilu, M.; Fitzgibbon, A.W. & Fisher R.B. (1996). Ellipsespecific Direct leastsquare Fitting, *Proceedings of the IEEE Int'l Conf. on Image Processing*, Vol. 3, p. 599602, Lausanne, Switzerland, IEEE Press, Dublin
- Ponce, J.; Chelberg, D. & Mann, W. (1989). Invariant properties of straight homogeneous generalized cylinders and their contours", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 9, No 11, p. 951966, IEEE Computer Society Press
- Pressigout, M & Marchand, E. (2006). Realtime 3-D ModelBased Tracking: Combining Edge and Texture Information. *In proceedings of the IEEE Int. Conf. on Robotics and Automation*, p. 27262731, Orlando, Florida, USA

- Puech, W.; Chassery, J.-M. & Pitas, I. (1997). Cylindrical surface localization in monocular vision, *Pattern Recognition Letters*, Vol. 18, p. 711722, Elsevier Science,
- Quan, L. & Lan, Z. (1999). Linear NPoint Camera Pose Determination, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No 8, p. 774780, IEEE Computer Society Press
- Richetin, M.; Dhome, M. & Lapresté, J.-T. & Rives, G., Inverse Perspective Transform Using ZeroCurvature Contour Points: Application to the Localization of Some Generalized Cylinders from a Single View, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No 2, p. 185192, IEEE Computer Society Press
- Rosin, P.L. (1999). Robust Pose Estimation, *IEEE Transactions on Systems, Man, and Cybernetics (Part B)*, Vol. 29, No 2, IEEE Press
- Rousseeuw, P. & Leroy, A. (1987). *Robust Regression and Outlier Detection*, ISBN 9780471488552 (new edition), Wiley, NY SafaeRad,
- R; Tchoukanov, I. & Smith, K. -C. & Benhabib, B. (1992). ThreeDimensional Location Estimation of Circular Features for Machine Vision, *IEEE Transactions on Robotics and Automation*, Vol. 8, No 5, p. 624640
- Salvi, J.; Armangué, X. & Batalle, J. (2002). A Comparative Review of Camera Calibrating Methods with Accuracy Evaluation, *Pattern Recognition*, Vol. 35, No 7, p. 16171635, Elsevier Science
- Shin, Y. C. & Ahmad, S. (1989). 3-D location of circular and spherical features by monocular modelbased vision, *Proceedings of the IEEE International Conference on System, Man and Cybernetic*, p. 576581, Boston, USA
- Shiu, Y.-C. & Huang, C. (1991). Pose Determination of Circular Cylinders Using Elliptical and Side Projections, *Proceedings of the Int'l Conf. on Systems Engineering*, p. 265268, Dayton, OH, USA
- Sturm, P. (1997). Critical motion sequences for monocular selfcalibration and uncalibrated Euclidean reconstruction, *Proceedings of the IEEE Int'l. Conference on Computer Vision and Pattern Recognition*, Puerto Rico, p. 11001105
- Sugihara, K. (1988). Some location problems for robot navigation using a single camera, *journal of Computer Vision, Graphics and Image Processing*, Vol. 42, No 1, p. 112129, Elsevier Science
- Sundareswaran, V & Behringer, R. (1998). VisualServoingbased Augmented Reality, *Proceedings of the first IEEE Int'l. Workshop on Augmented Reality*, San Francisco, USA
- Teramoto, H. & Xu, G. (2002). Camera calibration by a Single Image of Balls: From Conics to the Absolute Conic, *Proceedings of the Asian Conf. on Computer Vision*, Vol. 2, p. 499506, Melbourne, Australia
- Thompson, D.W. & Mundy, J.L. (1987). Threedimensional model matching from an unconstrained viewpoint, *Proceedings of the IEEE Int'l Conf. on Robotics and Automation*, Raleigh, NC, USA, p. 208220, IEEE Press
- Thompson, R. L.; Reid, I. D. & Munoz, L. A. & Murray, D. W. (2001). Providing synthetic views for teleoperation using visual pose tracking in multiple cameras, *IEEE Transactions System, Man and Cybernetics*, Vol. 31, No 1, p. 4354, IEEE Press
- Torr, P. H. S. & Zisserman, A. (1999). Feature based methods for structure and motion estimation, In: *Vision Algorithms Workshop.: Theory and Practice*, p. 278294, ISBN 3540679731, Springer Verlag, London, UK

- Triggs, W. (1999). Camera Pose and Calibration from 4 or 5 Known 3-D Points, *Proceedings of the IEEE Int'l Conf. on Computer Vision*, Vol. 1, p. 278284
- Trucco, E. & Verri, A. (1998). *Introductory Techniques for 3-D Computer Vision*, ISBN 0132611082, Prentice Hall Upper Saddle River, NJ 07458
- Tsaï, R.Y. (1987). A versatile camera calibration technique for highaccuracy 3-D machine vision metrology using offthe shelf TV cameras and lenses, *IEEE Journal of Robotics and Automation*, Vol. 3, No 4, p. 323344, IEEE Press NY
- Vacchetti, L; Lepetit, V. & Fua, P. (2004) Stable RealTime 3-D Tracking Using Online and Offline information, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No 10, p. 13851391, IEEE Computer Society
- Vincze, M.; Schlemmer, M. & Gemeiner, P. & Ayromlou, M. (2005). Vision for Robotics: A tool for ModelBased Object Tracking, *IEEE Robotics and Autonomous Magazine*, Vol. 12, No 4, p. 5364.
- Weiss, I. (1993). NoiseResistant Invariants of Curves, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No 9, p. 943948, IEEE Computer Society Press
- Wong, K.-Y.; Mendonça, P.-R. -S. & Cipolla, R. (2004). Reconstruction of surfaces of revolution from single uncalibrated views, *Image and Vision Computing*, Vol. 22, p. 829836, Elsevier Science
- Werghi, N. & Doignon, C. (1995). Contour decomposition with wavelet transform and parametrisation of elliptic curves with an unbiased extended Kalman filter, *Proceedings of the Second Asian Conference on Computer Vision*, Vol. 3, p. 186190, Singapore
- Yuan, J.-C. (1989). A general photogrammetric method for determining object position and orientation, *IEEE Transactions on Robotics and Automation*, Vol. 5, No 2, p. 129142
- Zhang, Z. (1999). A flexible new technique for camera calibration, *Proceedings of the International Conference on Computer Vision*, p. 666673, Corfu, Greece, IEEE Computer Society Press



Scene Reconstruction Pose Estimation and Tracking

Edited by Rustam Stolkin

ISBN 978-3-902613-06-6

Hard cover, 530 pages

Publisher I-Tech Education and Publishing

Published online 01, June, 2007

Published in print edition June, 2007

This book reports recent advances in the use of pattern recognition techniques for computer and robot vision. The sciences of pattern recognition and computational vision have been inextricably intertwined since their early days, some four decades ago with the emergence of fast digital computing. All computer vision techniques could be regarded as a form of pattern recognition, in the broadest sense of the term. Conversely, if one looks through the contents of a typical international pattern recognition conference proceedings, it appears that the large majority (perhaps 70-80%) of all pattern recognition papers are concerned with the analysis of images. In particular, these sciences overlap in areas of low level vision such as segmentation, edge detection and other kinds of feature extraction and region identification, which are the focus of this book.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Christophe Doignon (2007). An Introduction to Model-Based Pose Estimation and 3-D Tracking Techniques, Scene Reconstruction Pose Estimation and Tracking, Rustam Stolkin (Ed.), ISBN: 978-3-902613-06-6, InTech, Available from:

http://www.intechopen.com/books/scene_reconstruction_pose_estimation_and_tracking/an_introduction_to_model-based_pose_estimation_and_3-d_tracking_techniques



InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大酒店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2007 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.