Determination of the Pose of an Articulated Object From a Single Perspective View

M. DHOME, A. YASSINE & J.M. LAVEST

Electronics Laboratory, URA 830 of the CNRS
Blaise PASCAL University, 63177 AUBIERE CEDEX (FRANCE)
Tel : 73.40.72.28, Fax : 73.40.72.62
email : dhome @ le-eva.univ-bpclermont.fr

Abstract - This paper presents a new method that permits to estimate, in the viewer coordinate system, the spatial attitude of an articulated object from a single perspective image. Its principle is based on the interpretation of some image lines as the perspective projection of linear ridges of the object model, and on an iterative search of the model attitude consistent with these projections. The presented method doesn't locate separately the different parts of the object by using for each of them a technics devoted to the localization of rigid object but computes a global attitude which respects the mechanical articulations of the objet. In fact, the geometrical transformations applied to the model to bring it into the correct attitude are obtained in two steps. The first one is devoted to the estimation of the attitude parameters corresponding to a rotation and involves an iterative process. The second step permits by the resolution of a linear system to estimate the translation parameters. The presented experiments correspond to the localization of robot arms from synthetical and real images. The former case permits to appreciate the accuracy of the method since the final result of the pose estimation can be compared with the attitude parameters used to create the synthetical image. The latter case presents an experiment made in an industrial environment and involves the estimation of twelve paramaters since the observed robot arm owns six inner degrees of freedom. The presented method can be useful in some operation driven by remote control.

1 Introduction

The topic of the presented method is the *localization of an articulated object from a monocular image* knowing its CAD model and using the perspective projection as image formation assumption. This is typically a model based approach which is not a great restriction in the industrial world where the CAD model of the environment can be generally known.
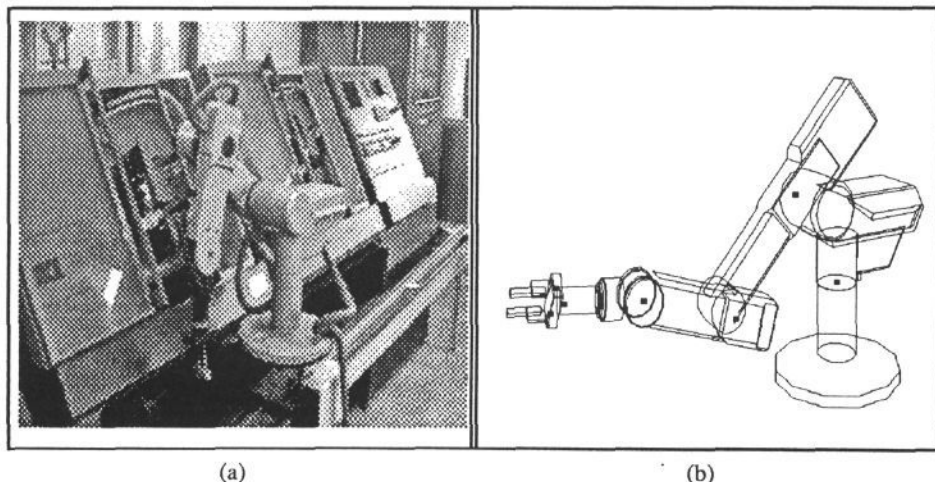

(a)   (b)

Figure 1 :   Presented method topic : pose estimation of an articulated object from one perspective grey-level image (a) knowing its CAD model (b).

In the last ten years, many researchers have tried to locate *rigid object* by monocular vision. All of proposed technics need some prerequisites which are the knowledge of the observed object model, matches between 2D image primitives and 3D model elements, and assumption about the projection of the real world on the image plane (orthographic or perspective). These methods can be classified in function of the manipulated image primitives and on the manner to reach the solution (closed form solution from lines ([KAN-81], [BAR-85], [SHA-86], [DHO-89]) or points ([RIV-81], [HUT-87], [HOR-89]), accumulation process [HOR-87] or iterative process [LOW-85]).

In fact, very few papers address the current problem which corresponds to the estimation, in the viewer coordinate system, of the *spatial attitude of an articulated object from a single perspective image*. In [MUL-89] Mulligan and al. present a technic to locate the boom, stick an bucket of an excavator. The object model used contains 3 inner rotations and these parameters are computed one by one.

The presented method doesn't locate separately the different parts of the object by using for each of them a technic devoted to the localization of rigid object (as presented previously), but computes a global attitude which respects the inner mechanical articulations of the objet. The presented method is based on the interpretation of *some image lines* as the perspective projection of *linear ridges of the object model*. No alignments point to point are needed between the extremities of the 2D and 3D matched primitives. We look for the object attitude such that the perspective projections of the selected model ridges coincide with the supporting lines of the image segments. We believe that it is an important advantage of our method because the extraction of the orientation and the distance from the origin of an image segment is more accurate than the location of its extremities. A similar approach can be found in [LOW-91] where Lowe proposes a technic to fit a parametrized 3D model to a perspective image. Nevertheless, our method differs from Lowe's one in two directions :

- the pose is computed in two steps. The first involves purely vector equations. The unknowns are all rotation angles and the devoted process is properly iterative. The second step is to compute the remaining unknown translations. It merely consists in the resolution of a linear least square problem.
- the minimized criterion differs from Lowe's one. Lowe uses a 2D criterion calculated in the image plane. Our criterion is 3D one which permits to greatly simplify the computations involved.

This paper is organized as follow; the second paragraph is devoted to the development of the attitude estimation method and the third paragraph presents experiments from synthetical and real images.

2 Overview of the method

2.1 General formulation

The estimation of the localization of an articulated object with respect to the inner mechanical articulations needs the selection of a *reference part*. This part can be independently located in the viewer coordinate system and the others parts are relatively positioned using the inner degrees of freedom. In practice, we select as reference part, the part having the highest number of ridges matched with image segments. Consequently, the pose estimation of an articulated object from a monocular perspective image depends on three kinds of parameters :

- *the intrinsic camera parameters*. These five parameters are (u_0, v_0) the coordinates of the intersection of the optical axis of the camera with the image plane, f the focal length and (dx, dy) the digitalization steps of the acquisition system. In fact, in this paper we will consider all the image primitives directly defined in the viewer coordinate system and only the focal length f will appear in our equations.

- *the extrinsic attitude parameters.* These parameters correspond to the three rotations and three translations around and along the world coordinate axis which permit to locate a rigid object. In our process, these parameters determine the pose of the *reference part* of the viewed articulated object. We will note $(\theta_1, \theta_2, \theta_3)$ the three rotation angles respectively around the x, y, z axis of the viewed coordinate system and $(\lambda_1, \lambda_2, \lambda_3)$ the amplitude of the translation respectively along the x, y, z axis of the viewer coordinate system.

- *the inner degrees of freedom.* These parameters define the state of the different inner articulations of the observed object. Each degree of freedom can be a rotation or a translation around or along a particular 3D axis. We will note $(\theta_4, ..., \theta_l, ..., \theta_p)$ the angles of the different inner rotations and $(\lambda_4, ..., \lambda_l, ..., \lambda_q)$ the amplitudes of the different inner translations.

Thus, the problem we intend to address can be described as follow :

Given : - the perspective projection (pinhole) as model of image formation,
          - the knowledge of the intrinsic parameters of the acquisition system,
          - the CAD model of the viewed object,
          - a sufficient set of matching between image segments and model ridges.

Find the extrinsic parameters and the inner degrees of freedom characterizing the pose of the viewed articulated object, in the acquisition reference system, compatible with the image content.

2.2 Description of the model

In our system, the articulated object is composed by several CAD models (one by part) connected together by articulations which describe the possible relative displacements between parts.

Each CAD model corresponds to the polyhedral approximation of the real part of the object. It is composed by a set of 3D vertices, a set of links between vertices to define the ridges and a set of links between ridges to build the different faces.

Each articulation is characterized by a set of degrees of freedom. Each degree of freedom is defined by its type (*rotation* or *translation*) and its 3D axis. At most, an articulation consists in three rotations and three translations. In this limit case, the relative movement of the two related parts will be arbitrary.

In the following, we suppose that the viewed object is composed by m parts and contents $p - 3$ inner rotations and $q - 3$ inner translations. To move a model point P_i , belonging to the tth part, to the point P_i' , we can apply to it the extrinsic rotations and translations and also the transformations associated to the degrees of freedom of the articulations situated between the reference part and the tth part. We choose to first apply the rotations. Let P_{iR} be the transformed point by rotation of the point P_i . The point P_{iR} is obtained by :

$$P_{iR} = [\bar{R}_{\theta_p}]^* ... [\bar{R}_{\theta_l}]^* ... [\bar{R}_{\theta_4}]^* [R_{\theta_3}][R_{\theta_2}][R_{\theta_1}]P_i$$

where $[R_{\theta_i}]_{i=1,2,3}$ are the homogeneous rotation matrix of the three extrinsic rotations and $[\bar{R}_{\theta_l}]^*_{l \in [4..p]}$ the transformation matrix associated with the lth inner object rotation. As this last rotation is applied around a 3D axis (not necessarily passing through the origin) the matrix $[\bar{R}_{\theta_l}]$ is in fact the product $[T_{\theta_l}]^{-1}[R_{\theta_l}][T_{\theta_l}]$ where $[T_{\theta_l}]$ is the translation matrix such that the rotation axis passes through the origin and $[R_{\theta_l}]$ is the rotation matrix of angle θ_l around this axis.

To obtain the final attitude P_i' of the point P_i , it is necessary to take into account the different translations.

$$P_i' = [T_{\lambda_q}]^* \ldots [T_{\lambda_l}]^* \ldots [T_{\lambda_4}]^* [T_{\lambda_3}][T_{\lambda_2}][T_{\lambda_1}]P_{iR}$$

where $[T_{\lambda_i}]_{i=1,2,3}$ are the homogeneous translation matrix of the three extrinsic translations

and $[T_{\lambda_l}]^*_{l \in [4..q]}$ the translation matrix associated with the lth inner object translation.

The symbol [ ]* means that the matrix $[\tilde{R}_{\theta_l}]$ or $[T_{\lambda_l}]$ is equal to the identity matrix if the lth inner degree of freedom doesn't take place between the reference part and the tth part (does not modify the attitude of P_i).

2.3 The mathematical equations

Let n image segments s_i be matched respectively with n model ridges L_i . Let us also suppose that all the vectors and points are expressed in the camera coordinate system.

Each segment s_i of the image is characterized by its unit director vector v_i and a point p_i . Then we can easily compute the unit vector normal N_i to the interpretation plane Π_i of s_i (plane through the optical center of the camera and the segment s_i).

Using the following notations : $v_i = (a_i, b_i, 0)$, $O_c p_i = (x_i, y_i, f)$, $N_i = (A_i, B_i, C_i)$

we obtain $\qquad N_i = \dfrac{v_i \times O_c p_i}{\| v_i \times O_c p_i \|} \qquad \Leftrightarrow \qquad N_i = \left(A_i = \dfrac{b_i f}{\sqrt{f^2 + d_i^2}}, B_i = \dfrac{-a_i f}{\sqrt{f^2 + d_i^2}}, C_i = \dfrac{d_i}{\sqrt{f^2 + d_i^2}} \right)$

where O_c is the optical center of the camera and $d_i = a_i y_i - b_i x_i$ is the distance between the supporting line of s_i and the image center (u_0, v_0) .
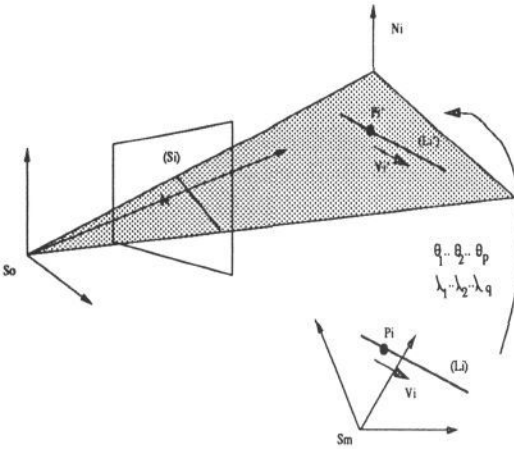

Figure 2 : General formulation.

The equation of the interpretation plane Π_i is : $A_i x + B_i y + C_i z = 0$

We are looking for the vectors $\Theta = (\theta_1, \ldots \theta_l, \ldots \theta_p)$ and $\Lambda = (\lambda_1, \ldots \lambda_l, \ldots \lambda_q)$ which define the object location minimizing the sum of the distances between the extremities of each model ridges and the interpretation plane of the corresponding image segments (see figure 2). This minimisation is made in two steps.

To compute the vector Θ we merely express that the transformed vector V_i' by the rotations of the director vector V_i of the model ridge L_i must lie in the interpretation plane of the corresponding image segment.

This can be written as a scalar product :

$$< N_i, V_i' > = < N_i, [R_{\theta_p}]^* \ldots [R_{\theta_l}]^* \ldots [R_{\theta_4}]^* [R_{\theta_3}][R_{\theta_2}][R_{\theta_1}]V_i > = 0$$

When the rotations are obtained, the translations are determined by expressing that a point P_i of the model ridge L_i rotated and translated also lies in the interpretation plane of the corresponding image segment. We obtain :

$$< N_i, P_i' > = < N_i, [T_{\lambda_p}]^* \ldots [T_{\lambda_l}]^* \ldots [T_{\lambda_4}]^* [T_{\lambda_3}][T_{\lambda_2}][T_{\lambda_1}]P_{iR} > = 0$$

with

$$P_{iR} = [\tilde{R}_{\theta_p}]^* \ldots [\tilde{R}_{\theta_l}]^* \ldots [\tilde{R}_{\theta_4}]^* [R_{\theta_3}][R_{\theta_2}][R_{\theta_1}]P_i$$

The main advantage in writing these equations is that the computation is performed successively for the rotations and the translations and that the complexity of the process decreases by solving two smaller systems.

2.4 Computing the rotations

The signed distance of the vector V_i' to the interpretation plane Π_i is given by :

$$D(\Theta,N_i,V_i) = <N_i, V_i'> = A_i u_i' + B_i v_i' + C_i w_i' \text{ where } V_i' = \left[R_{\theta_p}\right]^* ... \left[R_{\theta_l}\right]^* ... \left[R_{\theta_1}\right] V_i = (u_i', v_i', w_i')$$

We are looking for the vector $\Theta = (\theta_1, ...\theta_l, ...\theta_p)$ such that $\sum_{i=1}^{n} D(\Theta,N_i,V_i) = 0$. As the system is non linear, we use an iterative Newton-Raphson approach to solve it. At each step k of the iterative process, up to the first order we have :

$$D(\Theta,N_i,V_i) \approx D(\Theta_k,N_i,V_i) + \frac{\partial D(\Theta_k,N_i,V_i)}{\partial \Theta}(\Theta - \Theta_k)$$

where $\Theta_k = (\theta_{k1}, ...\theta_{kl}, ...\theta_{kp})$ is the rotation vector obtained at step k .

If Θ is the solution, we have $D(\Theta,N_i,V_i) = 0$ and we can use the following approximation :

$$D(\Theta_k,N_i,V_i) = \frac{\partial D(\Theta_k,N_i,V_i)}{\partial \theta_1}(\theta_{k1} - \theta_1) + ... \frac{\partial D(\Theta_k,N_i,V_i)}{\partial \theta_l}(\theta_{kl} - \theta_l) + ... \frac{\partial D(\Theta_k,N_i,V_i)}{\partial \theta_p}(\theta_{kp} - \theta_p)$$

The equation having p unknowns, we need at least p matchings. But in this limit case, if we want to estimate the orientation of each part of the viewed object, the matchings must be distributed on all parts such that each of them have a sufficient number of matchings to estimate its orientation.

Let $\Delta\Theta_k = [\theta_{kl} - \theta_l]$ for $l = 1..p$ be the angular correction vector, $D_k = [D(\Theta_k,N_i,V_i)]$ for $i = 1..n$ be the error vector and J_k be the Jacobian matrix of the function $D(\Theta_k,N_i,V_i)$ relatively to the variables θ_l . Using these notations, the system to solve is : $D_k = J_k\Delta\Theta_k$. The obvious way to solve the system is to write down the normal equations and solve the linear system :

$$^tJ_kD_k = {}^tJ_kJ_k\Delta\Theta_k$$

or to use a singular value decomposition of matrix J_k . At each step, we obtain a correction vector $\Delta\Theta_k$ to apply to the rotation vector Θ_k . The iterative process is repeated until a stable attitude is reached. Of course in any case the J_k matrix must be computed and this calculation will be detailed in the next section.

2.5 Computing the partial derivatives of the rotations

We want, for each value of $i = 1..n$, to compute the p derivatives ($l = 1..p$) :

$$\frac{\partial D(\Theta_k,N_i,V_i)}{\partial \theta_l} = <N_i, \frac{\partial [R_{\theta_p}]^* ... [R_{\theta_l}]^* ... [R_{\theta_1}]V_i}{\partial \theta_l}|_{(\theta_1 = \theta_{k1}, ...\theta_l = \theta_{kl}, ...\theta_p = \theta_{kp})}>$$

The partial derivatives of the rotation, at step k, can be written :

$$\frac{\partial [R_{\theta_p}]^* ... [R_{\theta_l}]^* ... [R_{\theta_1}]}{\partial \theta_l}|_{(\theta_1 = \theta_{k1}, ...\theta_l = \theta_{kl}, ...\theta_p = \theta_{kp})} = [R_{\theta_p}]^* ... \frac{\partial [R_{\theta_l}]^*}{\partial \theta_l}|_{\theta_l = \theta_{kl}} ... [R_{\theta_1}]$$

As the rotations corresponding to the inner degrees of freedom of the articulated object, are applied around a particular 3D axis, we use the quaternion representation of a rotation. By this way, the rotation matrix of angle θ_{kl} around a 3D axis with unit director vector $\Delta_l = (u_l, v_l, w_l)$ is written :

$$[R_{\theta_{kl}}] = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1 q_2 - q_0 q_3) & 2(q_1 q_3 + q_0 q_2) & 0 \\ 2(q_1 q_2 + q_0 q_3) & q_0^2 + q_2^2 - q_3^2 - q_1^2 & 2(q_2 q_3 - q_0 q_1) & 0 \\ 2(q_1 q_3 - q_0 q_2) & 2(q_2 q_3 + q_0 q_1) & q_0^2 + q_3^2 - q_1^2 - q_2^2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where $[q_0 = \cos(\theta_{kl}/2), q_1 = u_l \sin(\theta_{kl}/2), q_2 = v_l \sin(\theta_{kl}/2), q_3 = w_l \sin(\theta_{kl}/2)]$

To simplify the computation of the derivatives, at each step of the iterative process we modify the reference attitude of the model in the viewer coordinate system. In fact, the model is located according with the parameters found at the kth step, consequently and relatively to this new model definition, its current attitude is characterized, in the viewer coordinate system, by $\theta_{kl} = 0$ for $l = 1..p$. The computation of the partial derivatives of the rotation is now straightforward .

$$[q_0 = 1, q_1 = 0, q_2 = 0, q_3 = 0] \qquad \left[\frac{\partial q_0}{\partial \theta_{kl}} = 0, \frac{\partial q_1}{\partial \theta_{kl}} = u_l, \frac{\partial q_2}{\partial \theta_{kl}} = v_l, \frac{\partial q_3}{\partial \theta_{kl}} = w_l \right]$$

$$[R_{\theta_{kl}}] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \frac{\partial [R_{\theta_{kl}}]}{\partial \theta_l} = \begin{bmatrix} 0 & -w_l & v_l & 0 \\ w_l & 0 & -u_l & 0 \\ -v_l & u_l & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus, the partial derivative of the distance between the vector V_i' and the plane Π_i is simply the triple product of the vectors N_i, Δ_l and V_i as now $V_i' = V_i$:

$$\frac{\partial D(\Theta_k, N_i, V_i)}{\partial \theta_l} = [N_i, \Delta_l, V_i]$$

Finally, obtaining the matrices of the linear system, to solve at each step of the iterative process, needs very few computation :

$$J_k = \begin{pmatrix} [N_1, \Delta_1, V_1]^* & . & [N_1, \Delta_1, V_1]^* & . & [N_1, \Delta_p, V_1]^* \\ [N_i, \Delta_1, V_i]^* & . & [N_i, \Delta_1, V_i]^* & . & [N_i, \Delta_p, V_i]^* \\ [N_n, \Delta_1, V_n]^* & . & [N_n, \Delta_1, V_n]^* & . & [N_n, \Delta_p, V_n]^* \end{pmatrix} \qquad D_k = \begin{pmatrix} <N_1, V_1> \\ <N_i, V_i> \\ <N_n, V_n> \end{pmatrix}$$

where the symbol $[\ ]^*$ indicates that the triple product is equal to zero if the corresponding rotation is an inner degree of freedom without effect on the location of the considered matched ridge.

2.6 Computing the translations

Once the rotations are computed, it is easy to estimate the values of the translations amplitude. We are looking for the vector $\Lambda = (\lambda_1, \ldots \lambda_l, \ldots \lambda_q)$ such that the distance $\sum_{i=1}^{n} D(\Lambda, N_i, P_i') = 0$. Let us recall that P_i' is the transformed point by the translations of the point P_{iR}, itself the transformed point of P_i by the previously estimated rotations.

$$P_i' = [T_{\lambda_q}]^* \ldots [T_{\lambda_l}]^* \ldots [T_{\lambda_1}] P_{iR} \quad \text{where} \quad P_{iR} = [\bar{R}_{\theta_p}]^* \ldots [\bar{R}_{\theta_l}]^* \ldots [\bar{R}_{\theta_4}]^* [R_{\theta_3}][R_{\theta_2}][R_{\theta_1}] P_i$$

Each element of the translation matrix $[T_{\lambda_l}]$ can be written as follow (where (u_l, v_l, w_l) are the components of the unit director vector Δ_l of the translation). The distance $D(\Lambda, N_i, P_i)$ corresponds to a scalar product :

$$[T_{\lambda_l}] = \begin{bmatrix} 1 & 0 & 0 & \lambda_l.u_l \\ 0 & 1 & 0 & \lambda_l.v_l \\ 0 & 0 & 1 & \lambda_l.w_l \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$<N_i,[T_{\lambda_q}]...[T_{\lambda_l}]...[T_{\lambda_1}]P_{iR}>=\begin{bmatrix}A_i\\B_i\\C_i\end{bmatrix}\begin{bmatrix}X_{iR}+\lambda_1u_1+...\lambda_lu_l+...\lambda_qu_q\\Y_{iR}+\lambda_1v_1+...\lambda_lv_l+...\lambda_qv_q\\Z_{iR}+\lambda_1w_1+...\lambda_lw_l+...\lambda_qw_q\end{bmatrix}$$

If we consider the n equations corresponding to the n matchings between image segments and model ridges, we must solve the following linear system $D = J\Lambda$ with

$$J=\begin{pmatrix}<N_1,\Delta_1>^* & . & <N_1,\Delta_l>^* & . & <N_1,\Delta_q>^*\\<N_i,\Delta_1>^* & . & <N_i,\Delta_l>^* & . & <N_i,\Delta_q>^*\\<N_n,\Delta_1>^* & . & <N_n,\Delta_l>^* & . & <N_n,\Delta_q>^*\end{pmatrix}\qquad D=\begin{pmatrix}<N_1,P_{1R}>\\<N_i,P_{iR}>\\<N_n,P_{nR}>\end{pmatrix}$$

and where the symbol $<\ >^*$ indicates that the scalar product is equal to zero if the corresponding translation is an inner degree of freedom without effect on the location of the matched ridge.

3 Experiments

3.1 Localization of an articultated object from one monocular image

In this paragraph, we present experiments relative to the pose estimation of two robot arms. The models of the robots used, refer to existing industrial machines. The first one (robot # 1) is a robot arm composed of five parts animated by four rotations and one translation. The second robot (robot # 2) has height parts animated by five inner rotations and one inner translation. The computation of the robot pose using these models needs respectively the estimation of eleven and twelve attitude parameters.

Working from synthetical images permits to evaluate (under good conditions) the accuracy of the method since we know the attitude of the model used to generate the images. Experimental results from a such image are given in table 1.

Figure 3 presents the target grey-level image and the manually selected segments obtained by Deriche contour extractor [DER-87] and a polygonal approximation technique. The robot arm model is shown in the attitude corresponding to the initial pose used by the iterative process (figure 4). Note the important difference between the initial attitude parameters of the model and those corresponding to the final image target. In this experiment, the reference parts correspond to the part containing the ridges labelled (1-2-3-4).
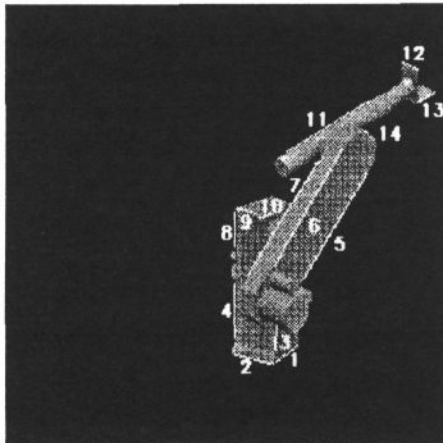

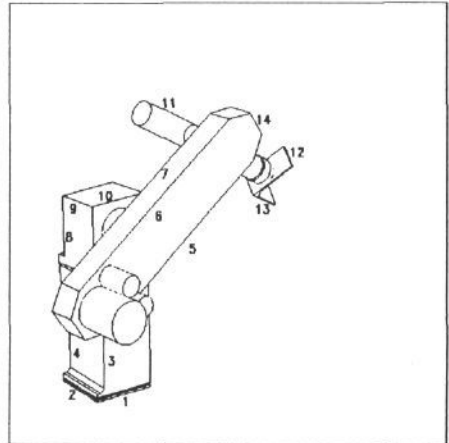Figure 3 : Synthetical images and selected segments.

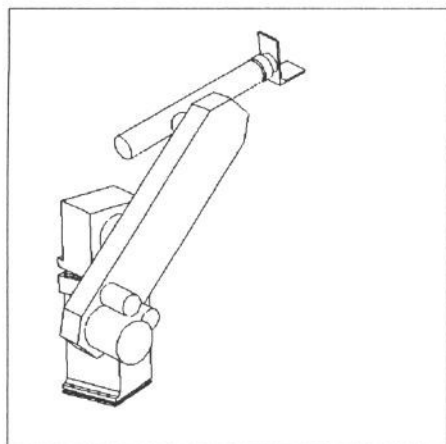Figure 4 : Initial pose of the robot #1 arm and selected ridges.

Figure 5 : Arm pose after the fourth and final step of the rotation estimation.
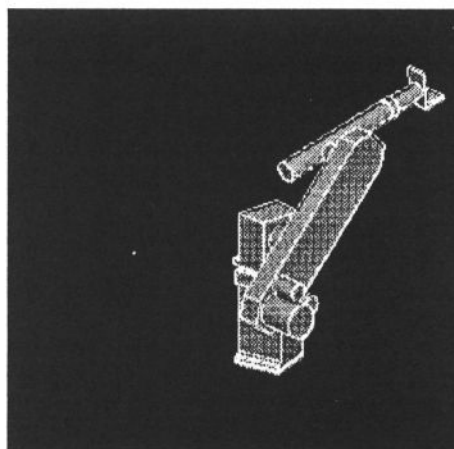
Figure 6 : Final arm pose after translation estimation.

In figure 5, we have drawn the attitude of the rotation found in the last (fourth) step of the iterative rotation estimation process. Fast convergence of the process is obtained since the more important corrections of the rotation are estimated at the first step. The process stop automatically when the correction to bring to the rotation parameters becomes equal to zero (in practice less than 0.01°). Using the resolution of a linear system, we can obtain the values of the remaining attitude parameters translation. In figure 6, we have superimposed to the target grey-level image the final pose of the model. Visually, the results seems quite good. More precise answer can be found in the table 1.

We can notice that the robot arm used in our experiment have cylindrical parts (example parts with ridges 11 in image 4). These parts are modelled as polyhedral parts with many generating lines. During the iterative process the matches are dynamically managed, the generating lines closest the real limbs of the cylindrical part (for the current attitude) are selected to be projected on the corresponding image segments.

Table 1 : Attitude parameters of the robot#1 arm corresponding to previous figures.

Attitude parameter	Initial value	Value used to create the synthetical image	Calculated value
Extrinsic parameters			
θ_1	336.0°	322.9°	323.0°
θ_2	325.6°	306.7°	306.8°
θ_3	14.1°	31.3°	31.3°
λ_1	-0.291 m	0.109 m	0.107 m
λ_2	-0.001 m	-0.001 m	0.000 m
λ_3	4.378 m	5.778 m	5.763 m
Inner degrees of freedom			
θ_4	-6.5°	18.1°	18.0°
θ_5	32.9°	46.3°	46.1°
θ_6	-63.4°	-31.9°	-31.4°
θ_7	-40.1°	-9.0°	-8.9°
λ_4	0.000 m	0.100 m	0.090 m

Table 1 shows the pose estimation results obtained for robot #1 and image presented in figure 3. The second column corresponds to the values characterizing the model attitude used to create the target image. The parameters of the initial pose of the model, used in the iterative process, are written in the first column and those of the computed attitude are given in the last column.

We can notice that, with synthetical images, the accuracy obtained is quite good. The mean error for rotation parameters is lower than 0.5° and the error for translation parameters is lower than 2 cm. Of course the maximum error is obtained for parameter λ_3 which corresponds to the translation along the view axis.
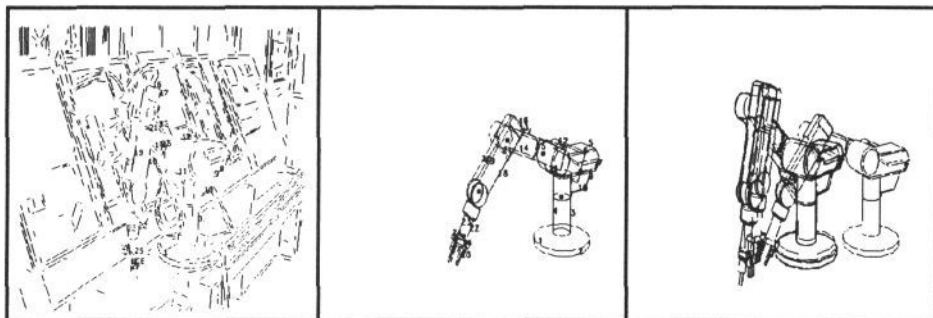

Figure 7 : Polygonal approximation and selected lines.

Figure 8 : Initial pose and selected ridges.

Figure 9 : Convergence of the iterative process.

The next presented experiment shows the result obtained from a real image taken in an industrial environment (figure 1-a). The straight line segments obtained after edges detection and polygonal approximation are shown in figure 7. The manually selected lines are also presented. The initial pose of the robot #2 arm and corresponding ridges are drawn in figure 8. The different steps of the iterative process are visualized in figure 9. Note that the major corrections are made during the first step. The final pose are superimposed to the grey-level image and presented in last figure.

This experiment could be seen as the protocol followed by an operator during a remote control process to estimate the pose of a viewed articulated object in dangerous environment.
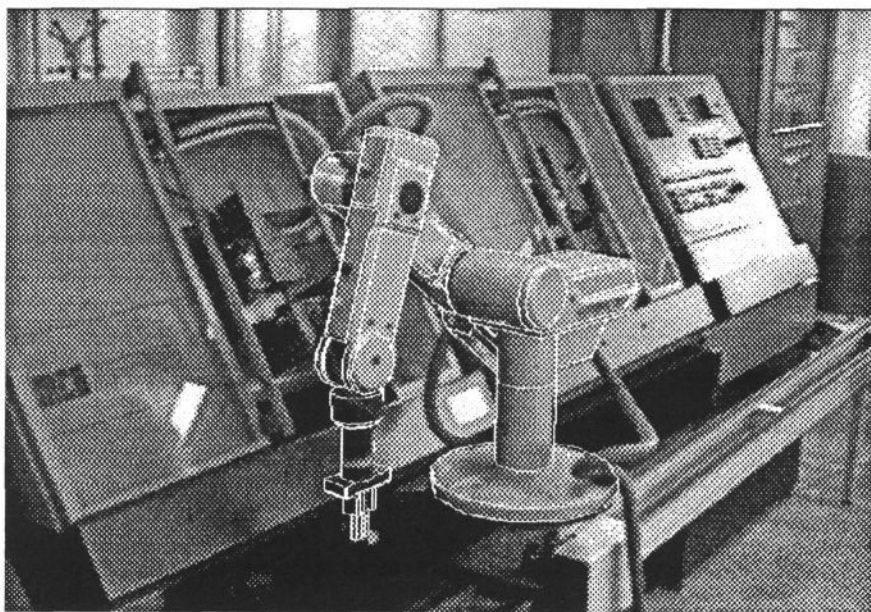

Figure 10 : Final pose estimation.

4 Conclusion

This paper presents a new method that permits to estimate, in the viewer coordinate system, the spatial attitude of an articulated object from a single perspective image. Its principle is based on the interpretation of some image lines as the perspective projection of linear ridges of the object model, and on an iterative search of the model attitude consistent with these projections.

To respect the inner mechanical articulations of the real object, the presented method computes the extrinsic attitude parameters (which characterize the pose of a reference part of the observed object in the viewer coordinate system) and simultaneously the values of the inner degrees of freedom. The parameters estimation is made in two steps. The first one is devoted to the calculus of the attitude parameters corresponding to a rotation and involves an iterative process. The second one permits by the resolution of a linear system to estimate the translation parameters. Each of these step involves the resolution of very simple linear system. Two examples of the pose estimation, from synthetical and real images are presented which show the accuracy and the reliability of this approach.

In present state, the presented method can be useful in some operations driven by remote control. Our next purpose will be to introduce Kalman formalism in order to estimate the covariance matrix relative to the pose parameters, and to use this approach along an image sequence to track the displacement of a robot arm.

5 Bibliography

[BAR-85] S.T. BARNARD, "Choosing a Basis for Perceptual Space", Proc. of Computer Vision Graphics and Image Processing, Vol 29, N°1, 1985, pp 87-99.

[DER-87] R. DERICHE, "Optimal Edge Detection Using Recursive Filtering", Proc. of first ICCV, London, June 1987, pp 501-505.

[DHO-89] M. DHOME and al, "Determination of the Attitude of 3D Objects from a Single Perspective Image", IEEE Trans. on PAMI, vol 11, N°12, December 1989, pp 1265-1278.

[DHO-90] M. DHOME and al, "Spatial Localization of Modelled Objects of Revolution in Monocular Perspective Vision ", Proc. of First European Conference on Computer Vision, Antibes, April 1990, pp 475-488.

[FOR-91] D. FORSYTH and al, "Invariant Descriptors for 3D Object Recognition and Pose", Proc. of DARPA/ESPRIT Workshop on Applications of Invariance in Computer Vision, Reykjavik, March 1991, pp 171-208.

[HOR-87] R. HORAUD, "New Methods for Matching 3-D Objects with Single Perspective View", IEEE Trans. on PAMI, vol PAMI-9, n°3, May 1987, pp 401-412.

[HOR-89] R. HORAUD and al, "An Analytic Solution for the Perspective 4 Points Problem", Proc. of Computer Vision Graphics and Image Processing, 1989.

[HUT-87] D.P. HUTTENLOCHER & S. ULLMAN, "Object Recognition using Alignment", Proc. of the First Int. Conf. on Computer Vision, London, June 1987, pp 102-111.

[KAN-81] T. KANADE, "Recovery of the Three Dimensional Shape of an Object from a Single View", Artificial Intelligence, Special Volume on Computer Vision, Vol 17, N°1-3, August 1981.

[LOW-85] D.G. LOWE, "Perceptual Organization and Visual Recognition", Kluwer, Boston, 1985, chapter 7.

[LOW-91] D.G. LOWE, "Fitting Parameterized Three-Dimensional Models to Images", Proc. of Trans. on PAMI, vol 13, N°5, May 91, pp 441-450.

[MUL-89] I.J. MULLIGAN and al, "A Model-Based Vision System for Manipulator Position Sensing", Proc. of Workshop on Interpretation of 3D Scenes, Austin, Texas, November 1989, pp 186-193.

[RIV-81] P. RIVES and al, "Recovering the Orientation and Position of a Rigid Body in Space from a Single View", Technical Report, INRS-Télécommunications, 3 place du commerce, Ile-des-soeurs, Verdun, H3E-1H6, Québec, Canada, 1981.

[SHA-86] T. SHAKUNAGA & H. KANEKO, "Perspective Angle Transform and Its Application to 3-D Configuration Recovery", Proc. of Int. Conf. on Computer Vision and Pattern Recognition, Miami Beach, Florida, June 1986, pp 594-601.