11

Video System in Robotic Applications

Vincenzo Niola, Cesare Rossi, Sergio Savino and Salvatore Strano
University of Naples "Federico II"
Italy

1. Introduction

The "Artificial Vision" permits industrial automation and system vision able to act in the production activities without humane presence. So we suppose that the acquisition and interpretation of the imagines for automation purposes is an interesting topic.

Industrial applications are referred to technological fields (assembly or dismounting, cut or stock removal; electrochemical processes; abrasive trials; cold or warm moulding; design with CAD techniques; metrology), or about several processes (control of the row material; workmanship of the component; assemblage; packing or storages; controls of quality; maintenance).

The main advantages of these techniques are:

1. elimination of the human errors, particularly in the case of repetitive or monotonous operations;
2. possibility to vary the production acting on the power of the automatic system (the automatic machines can operate to high rhythms day and night every day of the year);
3. greater informative control through the acquisition of historical data; these data can be used for successive elaborations, for the analysis of the failures and to have statistics in real time;
4. quality control founded on objective parameters in order to avoid dispute, and loss of image.

The use of a vision system in a robot application, it concurs to increase the robots ability to interact with their work space, to make more efficient their management.

In this chapter some "Artificial Vision" applications to robotics are described:

- robot cinematic calibration;
- trajectories recording;
- path planning by means of vision system;
- solid reconstruction with a video system on a robot arm.

2. Vision usefullness

The man perceives the characteristics of the external world by means of sense organs. They allow that a sure flow of information, regarding for example the shape, the color, the temperature, the smell of an object, reaches the brain; in this way each man possesses a complete description of that is around him. More in a generalized manner, the man can itself be seen as a system that, for survival reasons, must interact with the external world, and, to be able to make it, he has need of a sensory apparatus able to supply him continuously information. This affirmation can be extended also to not biological systems, like, for example, the automatic machines or the robots. These equipment, carrying out a determined task, interact with the external world and they must be fortified with devices that are able to perceive the world characteristics , this devices are called sensors. A robot or one whichever automatic machine, that is equipped with sensors, is be able to perceive the "stimula " of the external world, in which it works.

What is the difference between sensors and organs of sense, at the operating level?

When we perceive a any sound, for example the voice of a known person, we are able to distinguish the stamp, to establish if it is acute or serious, to feel the intensity and volume. If instead, the same voice is acquired through a microphone, converting its signal in digital and processing it by means of an electronic calculator, the information that we can deduce, increase and they make more detailed: we will be able, for example, to determine the main frequency components, to measure the amplitude in decibel, to visualize the wave shape. In other words, sensors, beyond to having the representative function of the truth, concur also to extrapolate information at quantitative level and they allow us to lead a technical analysis on the acquired data.

Main aim of this chapter is to show how it is possible to equip robot of sight

The main problem of the reliable and precise robot realization, has been to implement the hardware e software structures, that constitute a sturdy and efficient control system.

How does motion control work in the man? The human body has a much elevating number of degrees of freedom and this renders very arduous the nervous system task. For this reason a highly centralized control structure is necessary. It is possible to describe the job of such structure through a simple example: let's image a man, that wants to take an object that is disposed on a table distant some meters. The man observes the table and the object position, while the brain elaborates the trajectory and the nervous impulses to transmit to muscles, characterizing reference points that are acquired from image observed by eyes. Subsequently the man begins to move and after some step, he reaches the table and takes the object. From this example, it can be asserted that, excluding the memory contribution and an elevated development of the other senses, it is not possible to carry out a task without to see.

Therefore the sense of the sight it has a twofold function in the human body motion:

1. to characterize the targets in the space.
2. to control the position and the guidelines of the several parts of the body that move.

In the robots, the motion control is, usually implemented only with joints position transducers. It is clear that, if joints translation and rotations are known, its spatial configuration is known completely; therefore, the second function that has been attributed to human sight is realized. A blind control is less suitable to catch up a target in the work space. In fact, to guide the robot end effector to a point, it is necessary to know, with precision, its cartesian coordinates and "translate" them in the joints space by means of inverse cinematic. For this reason, it is useful to increase robots sensory abilities, equipping them "off sight", by means of vision systems with opportune sensors. In this chapter it will be described in how it is possible to characterize the targets in the work space and to determine the values of the Denavit-Hartenberg cinematic parameters, by means of opportune techniques, and with two television cameras, so as to make simpler, more accurate and efficient both the robot management and the motion planning.

3. Vision process

About the term "vision" applications to industrial robots, the meaning of this word must be enriched and cleared with technical slight knowledge.
In literature, use of the vision like instrument for technical applications, is called "machine vision" or "computer vision".
It is important to explain, in the first place, which is the aim of the computer vision: to recognize the characteristics of objects that are present in the acquired images of work space and to associate them theirs real meant.
The vision process can be divided in following operations:

- Perception
- Pre-elaboration
- Segmentation
- Description
- Recognition
- Interpretation

Perception is the process that supplies the visual image. With this operation, we mean the mechanism of photogram formation by means of a vision system and a support, like a computer.
Pre-elaboration is the whole ~~of~~ noise reduction techniques and images improvement techniques.
Segmentation is the process by means of which the image is subdivided in characteristics of interest.
Description carries out the calculation of the characteristics that segmentation has evidenced, it represents the phase in which it is possible to quantify that only qualitatively has been characterized: lengths, areas ,volumes, ecc.
Recognition consists in assembling all the characteristics that belong to the object, in order to characterize the object. By means of the last phase, called interpretation, it is established the effective correspondence between a characterized shape and the object that is present in the real scene.
To say that a robot "sees", does not mean simply that it has a reality representation, but that it is able to recognize quantitatively the surrounding space, that is to recognize distances, angles, areas and volumes of the objects that are in the observed scene.

4. The perspective transform

In this paragraph, an expression of perspective transformation is proposed, in order to introduce the perspective concepts for the application in robotic field.
The proposed algorithm uses the fourth row of the Denavit and Hartemberg transformation matrix that, for kinematics' purposes, usually contains three zeros and a scale factor, so it is useful to start from the perspective transform matrix.

4.1 The matrix for the perspective transformation [1,5]

It is useful to remember that by means of a perspective transform it is possible to associate a point in the geometric space to a point in a plane, that will be called "image plane"; this will be made by using a scale factor that depends on the distance between the point itself and the image plane.

Let's consider fig.1: the position of point P in the frame O,x,y,z is given by the vector w, while the same position in the frame Ω,ξ,η,ζ is given by vector w_r and the image plane is indicated with R; this last, for the sake of simplicity is supposed to be coincident with the plane ξ,η.

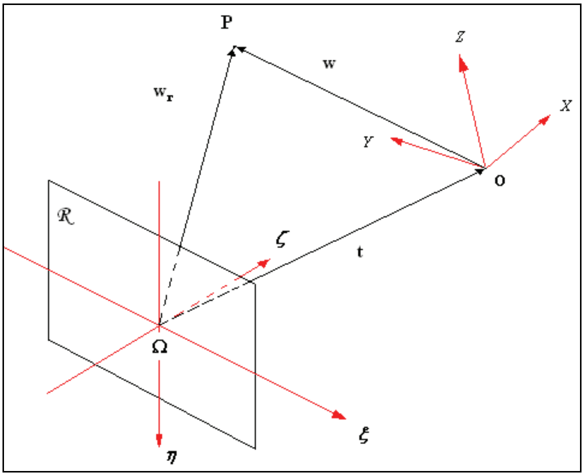


Figure 1. Frames for the perspective transformation

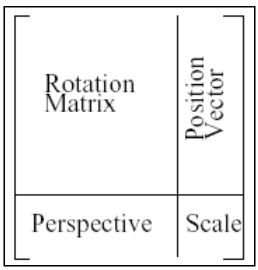
The vectors above are joined by the equation:

$$\begin{Bmatrix} w_{r,x} \\ w_{r,y} \\ w_{r,z} \\ sf \end{Bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & t_\xi \\ R_{21} & R_{22} & R_{23} & t_\eta \\ R_{31} & R_{32} & R_{33} & t_\zeta \\ 0 & 0 & 0 & sf \end{bmatrix} = \begin{Bmatrix} w_x \\ w_y \\ w_z \\ sf \end{Bmatrix} \tag{1}$$

where sf is the scale factor; more concisely equation (1) can be written as follows:

$$\tilde{w}_r = T \cdot \tilde{w} \tag{2}$$

where the tilde indicates that the vectors are expressed in homogeneous coordinates.
The matrix T is a generic transformation matrix that is structured according to the following template:



The scale factor will almost always be 1 and the perspective part will be all zeros except when modelling cameras.
The fourth row of matrix [T] contains three zeros; as for these last by means of the prospectic transform three values, generally different by zero, will be determined.
Lets consider, now, fig.2: the vector w*, that represents the projection of vector w_r on the plane ξ,η.

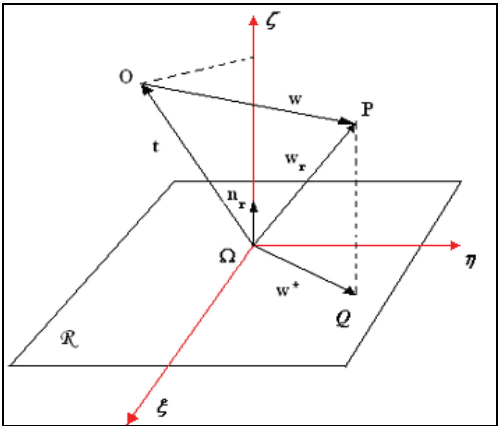


Figure 2. Vectors for the perspective transformation

The coordinates of point P in the image plane can be obtained from the vector w_r, in fact, these coordinates are the coordinates of w*, that can be obtained as follows:
Let's consider the matrix R:

$$R = \begin{bmatrix} \hat{\xi}^T \\ \hat{\eta}^T \\ \hat{\zeta}^T \end{bmatrix} \tag{3}$$

where $\hat{\xi}$ $\hat{\eta}$ $\hat{\zeta}$ are the versor of the frame $\{\Omega,\xi,\eta,\zeta\}$ axes in the frame $\{O,x,y,z\}$.
In fig.2 the vector t indicates the origin of frame O,x,y,z in the frame Ω,ξ,η,ζ and the projection of P on the plane ξ,η is represented by point Q, which position vector is w*. This last, in homogeneous coordinates is given by:

$$\tilde{w}^* = \begin{pmatrix} w_{r,\xi} \\ w_{r,\eta} \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \hat{\xi}^T w + t_\xi \\ \hat{\eta}^T w + t_\eta \\ 0 \\ 1 \end{pmatrix} \tag{4}$$

In the same figure, n_r is the versor normal to the image plane R , and n will be the same versor in the frame $\{O,x,y,z\}$. The perspective image of vector w* can be obtained by assessing a suitable scale factor. This last depends on the distance d between point P and the image plane. The distance d is given from the following scalar product:

$$d = n_r^T w_r \tag{5}$$

Let's indicate with $w_{\{\Omega,\xi,\eta,\zeta\}}$ the vector w in the frame $\{\Omega,\xi,\eta,\zeta\}$:

$$\tilde{w}_{\{\Omega,\xi,\eta,\zeta\}} = \begin{pmatrix} w_\xi \\ w_\eta \\ w_\zeta \\ 1 \end{pmatrix} \tag{6}$$

Because $\hat{\xi}$ $\hat{\eta}$ $\hat{\zeta}$ are the versor of the frame $\{\Omega,\xi,\eta,\zeta\}$ axes in the frame $\{O,x,y,z\}$, it is possible to write the coordinates of the vector $w_{\{\Omega,\xi,\eta,\zeta\}}$ in the frame $\{\Omega,\xi,\eta,\zeta\}$:

$$\begin{aligned} w_\xi &= \hat{\xi}^T \cdot w = \xi_x w_x + \xi_y w_y + \xi_z w_z; \\ w_\eta &= \hat{\eta}^T \cdot w = \eta_x w_x + \eta_y w_y + \eta_z w_z; \\ w_\xi &= \hat{\zeta}^T \cdot w = \zeta_x w_x + \zeta_y w_y + \zeta_z w_z; \end{aligned} \tag{7}$$

In the frame $\{\Omega,\xi,\eta,\zeta\}$, it is possible to write w_r as sum of $w_{\{\Omega,\xi,\eta,\zeta\}}$ and t:

$$\tilde{w}_r = \tilde{w}_{\{\Omega,\xi,\eta,\zeta\}} + \tilde{t} = \begin{pmatrix} w_\xi + t_\xi \\ w_\eta + t_\eta \\ w_\zeta + t_\zeta \\ 1 \end{pmatrix} \tag{8}$$

Let's introduce the expressions:

$$\begin{aligned} D_x &= \frac{\left(\xi_x w_x + \xi_y w_y + \xi_z w_z + t_\xi\right) \cdot n_{r,\xi}}{w_x}; \\ D_y &= \frac{\left(\eta_x w_x + \eta_y w_y + \eta_z w_z + t_\eta\right) \cdot n_{r,\eta}}{w_y}; \\ D_z &= \frac{\left(\zeta_x w_x + \zeta_y w_y + \zeta_z w_z + t_\zeta\right) \cdot n_{r,\zeta}}{w_z}; \end{aligned} \tag{9}$$

it is possible to write:

$$d = n_r{}^T w_r = \begin{pmatrix} D_x \\ D_y \\ D_z \\ 0 \end{pmatrix}^T \cdot \begin{pmatrix} w_x \\ w_y \\ w_z \\ 1 \end{pmatrix} = D^T \cdot w \tag{10}$$

In the equation (10) the vector D is:

$$D = \begin{pmatrix} D_x \\ D_y \\ D_z \\ 0 \end{pmatrix} \tag{11}$$

As vector w* is given by:

$$\tilde{w}^*{}_p = \begin{pmatrix} \hat{\xi}^T w + t_\xi \\ \hat{\eta}^T w + t_\eta \\ 0 \\ n_r{}^T w_r \end{pmatrix} \tag{12}$$

The perspective matrix $[T_p]$ can be obtained:

$$\tilde{w}^*{}_p = T_p \cdot \tilde{w} \Rightarrow T_p = \begin{bmatrix} \xi_x & \xi_y & \xi_z & t_\xi \\ \eta_x & \eta_y & \eta_z & t_\eta \\ 0 & 0 & 0 & 0 \\ D_x & D_y & D_z & 0 \end{bmatrix} \tag{13}$$

The terms Dx, Dy, Dz assume infinity values if the vector w has one of his coordinates null, but this does not influence on generality of the relation $\tilde{w}^*{}_p = T_p \cdot \tilde{w}$, in fact in this case, the term that assume infinity value, is multiplied for zero.

4.2 The perspective concept

From equation (13) some useful properties can be obtained in order to define how a geometric locus changes its representation when a perspective transform occurs.
As for an example of what above said, let us consider the representation of the displacement of a point in the space: suppose that the displacement occurs, initially, in the positive

direction of x axis. Say this displacement Δw, the point moves from the position P to the position P′, that are given by the vectors:

$$w = \begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix} \quad \text{and} \quad w' = \begin{pmatrix} w'_x \\ w_y \\ w_z \end{pmatrix} \tag{14}$$

If the perspective transforms are applied we have :

$$p = Tp \cdot w \quad \text{and} \quad p' = Tp \cdot w' \tag{15}$$

the displacement in the image plane is given by:

$$\Delta p = p' - p \tag{16}$$

that is to say:

$$\Delta p = \begin{Bmatrix} \dfrac{\xi_x \cdot [w_x'(D^T w) - w_x(D'^T w')]}{(D^T w)(D'^T w')} \\ \dfrac{\eta_x \cdot [w_x'(D^T w) - w_x(D'^T w')]}{(D^T w)(D'^T w')} \\ 0 \end{Bmatrix} \tag{17}$$

In this way, a displacement Δw along the x axis corresponds to a displacement Δp in the image plane along a straight line which pitch is . So the x axis equation in the image plane is:

$$\eta = (\eta_x / \xi_x) \cdot \xi + \frac{\xi_x t_\eta - \eta_x t_\xi}{\xi_x} \tag{18}$$

The interception was calculated by imposing that the point which coordinates are belongs to the x axis. In the same way it is possible to obtain the y axis and the z axis equations:

$$\text{y axis :} \quad \eta = (\eta_y / \xi_y) \cdot \xi + \frac{\xi_y t_\eta - \eta_y t_\xi}{\xi_y} \tag{19}$$

$$\text{z axis :} \quad \eta = (\eta_z / \xi_z) \cdot \xi + \frac{\xi_z t_\eta - \eta_z t_\xi}{\xi_z} \tag{20}$$

By means of equations (18), (19) and (20) it is possible to obtain a perspective representation of a frame belonging to the Cartesian space in the image plane; that is to say: for a given body it is possible to define it's orientation (e.g. roll, pitch and yaw) in the image plane.

4.3 Perspective transformation in D-H robotic matrix [5]

For kinematics purposes in robotic applications, it is possible to use the Denavit and Hartemberg transformation matrix in homogeneous coordinates in order to characterize the end-effector position in the robot base frame by means of joints variable, this matrix usually contains three zeros and a scale factor in the fourth row. The general expression of the homogenous transformation matrix that allows to transform the coordinates from the frame i to frame i-1, is:

$$A_1^{i-1} = \begin{bmatrix} C\vartheta_i & -C\alpha_i \cdot S\vartheta_i & S\alpha_i \cdot S\vartheta_i & a_i \cdot C\vartheta_i \\ S\vartheta_i & C\alpha_i \cdot C\vartheta_i & -S\alpha_i \cdot C\vartheta_i & a_i \cdot S\vartheta_i \\ 0 & S\alpha_i & C\alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{21}$$

For a generic robot with n d.o.f., the transformation matrix from end-effector frame to base frame, has the following expression:

$$T_n^0 = A_1^0 \cdot A_2^1 \cdot A_3^2 \cdot ...... \cdot A_n^{n-1} \tag{22}$$

With this matrix it is possible to solve the expression:

$$\{P\}_0 = T_n^0 \cdot \{P\}_n \tag{23}$$

where and are the vectors that represent a generic point P in frame 0 and frame n.
It is useful to include the perspective concepts in this transformation matrix; in this way it is possible to obtain a perspective representation of the robot base frame, belonging to the Cartesian space, in an image plane, like following expression shows:

$$\{P\}_p = T_p \cdot \{P\}_0 = T_p \cdot T_n^0 \cdot \{P\}_n = \left[T_p\right]_n^0 \cdot \{P\}_n \tag{24}$$

where is the perspective image of generic point P and is the perspective transformation matrix from end-effector frame to an image plane.
With this representation the fourth row of the Denavit and Hartemberg matrix will contain non-zero elements. A vision system demands an application like this.

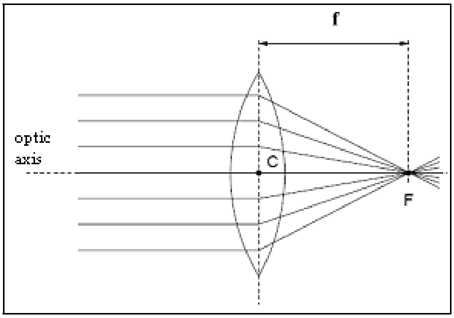
5. The camera model

When vision systems are used for robotic applications, it is important to have a suitable model of the cameras.
A vision system essentially associates a point in the Cartesian space with a point on the image plane. A very common vision system is the television camera that is essentially composed by an optic system (one or more lenses), an image processing and managing system and an image plane; this last is composed by vision sensors. The light from a point in the space is conveyed by the lenses on the image plane and recorded by the vision sensor.
Let us confine ourselves to consider a simple vision system made up by a thin lens and an image plane composed by CCD (Charged Coupled Device) sensors. This kind of sensor is a

device that is able to record the electric charge that is generated by a photoelectric effect when a photon impacts on the sensor's surface.
It is useful to remember some aspects of the optics in a vision system.

5.1 The thin lenses model [2,4,13,14]

A lens is made up by two parts of a spherical surfaces (dioptric surfaces) joined on a same plane. The axis, normal to this plane, is the optical axis. As shown in fig.3, a convergent lens conveys the parallel light rays in a focus F at distance f (focal distance) from the lens plane.



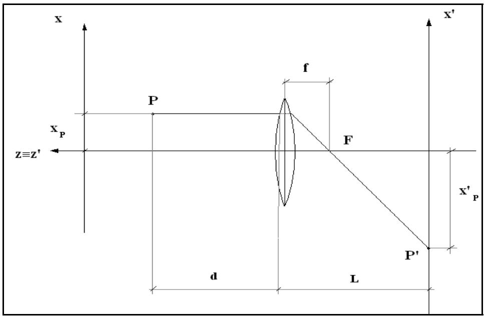


Figure 3. Convergent lens (left), Thin lens (rigth)

The focal distance f, in air, is given by:

$$f = (n - 1) \cdot \left(\frac{1}{R_1} - \frac{1}{R_2} \right) \tag{25}$$

where n is the refractive index of the lens and R_1 ed R_2 are the bending radius of the dioptric surfaces.
Now consider a thin lens, a point P and a plane on which the light-rays refracted from the lens are projected as shown in fig.3 the equation for the thin lenses gives:

$$\frac{1}{d} + \frac{1}{L} = \frac{1}{f} \tag{26}$$

It is possible to determinate the connection between the position of point P in the space and it's correspondent P' in the projection's plane (fig.3).
If two frames (xyx for the Cartesian space and x'y'z' for the image plane), having their axes parallel, are assigned and if the thickness of the lens is neglected, from the similitude of the triangles in fig.5 it comes:

$$\frac{x_P}{f} = -\frac{x'_P}{L - f} \tag{27}$$

with the equation of the thin lenses we can write:

$$x_P^{'} = -\frac{f}{d-f} \cdot x_P \tag{28}$$

If we consider that generally the distance of a point from the camera's objective is one meter or more while the focal distance is about some millimetres (d>>f), the following approximation can be accepted:

$$x_P^{'} \cong -\frac{f}{d} \cdot x_P \tag{29}$$

So the coordinates of the point in the image plane can be obtained by scaling the coordinates in the Cartesian space by a factor . The minus sign is due to the upsetting of the image.

5.2 The model of the camera [4]

As already observed a camera can be modelled as a thin lens and an image plane with CCD sensors. The objects located in the Cartesian space emit rays of light that are refracted from the lens on the image plane. Each CCD sensor emit an electric signal that is proportional to the intensity of the ray of light on it; the image is made up by a number of pixels, each one of them records the information coming from the sensor that corresponds to that pixel.
In order to indicate the position of a point on an image it is possible to define a frame u,v (fig.4) which axes are contained in the image plane. To a given point in the space (which position is given by its Cartesian coordinates) it is possible to associate a point in the image plane (two coordinates) by means of the camera. So, the expression "model of the camera" means the transform that associates a point in the Cartesian space to a point in the image space.
It has to be said that in the Cartesian space a point position is given by three coordinates expressed in length unit while in the image plane the two coordinates are expressed in pixel; this last is the smaller length unit that ca be revealed by the camera and isn't a normalized length unit. The model of the camera must take onto account this aspect also.
In order to obtain the model of the camera the scheme reported in fig.4 can be considered.

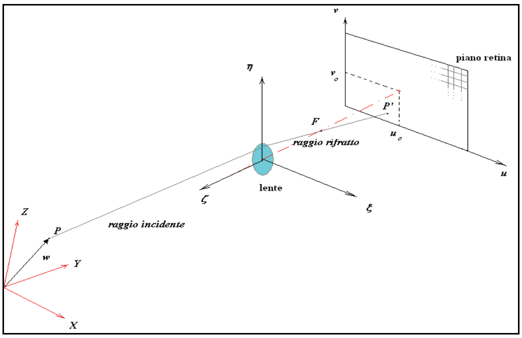


Figure 4. Camera model

Consider a frame xyz in the Cartesian space, the position of a generic point P in the space is given by the vector w. Then consider a frame ξ,η,ζ having the origin in the lens centre and the plane ξ,η coincident with the plane of the lens; hence, the plane ξ,η is parallel to the

image plane and ζ axis is coincident with the optical axis. Finally consider a frame u,v on the image plane so that u_0 and v_0 are the coordinates of the origin of frame ξ,η,ζ expressed in pixel.
As it was already told, the lens makes a perspective transform in which the constant of proportionality is –f . If this transform is applied to vector w, a w_l vector is obtained:

$$\tilde{w}_l = T_l \cdot \tilde{w} \tag{30}$$

Were the matrix T_l is obtained dividing by –f the last row of the perspective transformation matrix T_p.

$$T_l = \begin{bmatrix} \xi_x & \xi_y & \xi_z & t_\xi \\ \eta_x & \eta_y & \eta_z & t_\eta \\ 0 & 0 & 0 & 0 \\ -\frac{D_x}{f} & -\frac{D_y}{f} & -\frac{D_z}{f} & 0 \end{bmatrix} \tag{31}$$

Substantially, the above essentially consists in a changing of the reference frames and a scaling based on the rules of geometric optics previously reported.
Assumed x_l e y_l as the first two components of the vector w_l, the coordinates u and v (expressed in pixel) of P' (image of P) are :

$$\begin{cases} u = \dfrac{x_l}{\delta_u} + u_o \\ v = \dfrac{x_l}{\delta_v} + v_o \end{cases} \tag{32}$$

Where δ_u e δ_v are respectively the horizontal and vertical dimensions of the pixel.
So, by substituting equation (30) in equation (32) it comes:

$$\begin{cases} u = -\dfrac{f}{D^T w}\left[\left(\dfrac{1}{\delta_u}\cdot\hat{\xi} - \dfrac{u_o}{f}\cdot D\right)^T w + \dfrac{1}{\delta_u}\cdot t_\xi\right] \\ v = -\dfrac{f}{D^T w}\left[\left(\dfrac{1}{\delta_v}\cdot\hat{\eta} - \dfrac{v_o}{f}\cdot D\right)^T w + \dfrac{1}{\delta_v}\cdot t_\eta\right] \end{cases} \tag{33}$$

Finally, if we define the vector $m = [u \quad v]^T$, the representation in homogeneous coordinates $\tilde{m} = \left[m_1 \quad m_2 \quad -D^T w/f\right]^T = \left[u \quad v \quad -D^T w/f\right]^T$ of the previous vector can be written :

$$\tilde{m} = M \cdot \tilde{w} \tag{34}$$

Where M is the matrix :

$$M = \begin{bmatrix} \left(\frac{\xi_x}{\delta_u} - \frac{u_o D_x}{f}\right) & \left(\frac{\xi_y}{\delta_u} - \frac{u_o D_y}{f}\right) & \left(\frac{\xi_z}{\delta_u} - \frac{u_o D_z}{f}\right) & t_\xi / \delta_u \\ \left(\frac{\eta_x}{\delta_v} - \frac{v_o D_x}{f}\right) & \left(\frac{\eta_y}{\delta_v} - \frac{v_o D_y}{f}\right) & \left(\frac{\eta_z}{\delta_v} - \frac{v_o D_z}{f}\right) & t_\eta / \delta_v \\ -D_x / f & -D_y / f & -D_z / f & 0 \end{bmatrix} \tag{35}$$

that represents the requested model of the camera.

6. The stereoscopic vision

What above reported concurs to determine the coordinates in image plane (u,v) of a generic point of tridimensional space $w=[w_x\ w_y\ w_z\ 1]^T$, but the situation is more complex if it is necessary to recognise the position (w) of a point starting to its camera image (u, v). In this case the equations (33) becomes a system of 2 equation with 3 unknowns, so it has no solutions.
This obstacle can be overcome by means of a vision system with at least two cameras.
In this way, what above reported can be applied to the recording of a robot trajectory in the three dimensional space by using two cameras. This will emulate the human vision.
Let us consider two cameras and say M and M′ their transform matrixes. We want to recognise the position of a point P, that in the Cartesian space is given by a vector w in a generic frame xyz. From equation (34) we have:

$$\begin{cases} \tilde{m} = M \cdot w \\ \tilde{m}' = M' \cdot w \end{cases} \tag{36}$$

The first equation of the system (36), in Cartesian coordinates (non-homogenous), can be written as:

$$\begin{cases} (u \cdot D + f \cdot \mu_1)^T w = \mu_{14} \\ (v \cdot D + f \cdot \mu_2)^T w = \mu_{24} \end{cases} \tag{37}$$

Where:

$$\begin{aligned} \mu_1 &= \left\{\left(\frac{\xi_x}{\delta_u} - \frac{u_o D_x}{f}\right)\left(\frac{\xi_y}{\delta_u} - \frac{u_o D_y}{f}\right)\left(\frac{\xi_z}{\delta_u} - \frac{u_o D_z}{f}\right)\right\}; \\ \mu_2 &= \left\{\left(\frac{\eta_x}{\delta_v} - \frac{v_o D_x}{f}\right)\left(\frac{\eta_y}{\delta_v} - \frac{v_o D_y}{f}\right)\left(\frac{\eta_z}{\delta_v} - \frac{v_o D_z}{f}\right)\right\}; \\ \mu_{14} &= t_\xi / \delta_u \\ \mu_{24} &= t_\eta / \delta_v \end{aligned} \tag{38}$$

In the same way for the camera, whose transform matrix is M′, it can be written:

$$\begin{cases}(u'\cdot D'+f'\cdot\mu'_1)^T\, w = \mu'_{14} \\ (v'\cdot D'+f'\cdot\mu'_2)^T\, w = \mu'_{24}\end{cases} \tag{39}$$

By arranging eq.(26) and eq.(27) we obtain:

$$\begin{bmatrix}(u\cdot D+f\cdot\mu_1)^T \\ (v\cdot D+f\cdot\mu_2)^T \\ (u'\cdot D'+f'\cdot\mu'_1)^T \\ (v'\cdot D'+f'\cdot\mu'_2)^T\end{bmatrix}\cdot w = \begin{bmatrix}\mu_{14} \\ \mu_{24} \\ \mu'_{14} \\ \mu'_{24}\end{bmatrix} \tag{40}$$

This last equation represents the stereoscopic problem and consist in a system of 4 equation in 3 unknown (w_x,w_y,w_z). As the equations are more than the unknowns can be solved by a least square algorithm. In this way it is possible to invert the problem that is described by equations (33) and to recognise the position of a generic point starting to its camera image.

6.1 The stereoscopic problem [2]

Relation (40) represents the stereoscopic problem, it consists in a system of 4 equations in 3 unknown, in the form:

$$A\big(u, u, u', v', w\big)\cdot w = B \tag{41}$$

where A is a matrix that depends by two couple of camera coordinates (u,v) and (u′,v′), and by vector w, and B is a vector with parameters of cameras configuration.
It is possible to find an explicit form of this problem.
Starting to first equation of (33), it is possible to write:

$$u = -\frac{f}{D^T w}\left[\left(\frac{1}{\delta_u}\cdot\hat{\xi} - \frac{u_o}{f}\cdot D\right)^T w + \frac{1}{\delta_u}\cdot t_\xi\right] \Rightarrow$$

$$\frac{1}{\delta_u}\big(\xi_x\cdot w_x + \eta_x\cdot w_y + \zeta_x\cdot w_z\big)- \tag{42}$$

$$\frac{u_0}{f}\big(D_x\cdot w_x + D_x\cdot w_y + D_x\cdot w_z\big)+\frac{u}{f}\big(D_x\cdot w_x + D_x\cdot w_y + D_x\cdot w_z\big)=-\frac{t_\xi}{\delta_u}$$

By means of equation (9), it is possible to write:

$$\begin{gathered} D_x \cdot w_x + D_x \cdot w_y + D_x \cdot w_z = \\ w_x\left(\xi_x \cdot n_{r,\xi} + \xi_y \cdot n_{r,\eta} + \xi_z \cdot n_{r,\zeta}\right) + w_y\left(\eta_x \cdot n_{r,\xi} + \eta_y \cdot n_{r,\eta} + \eta_z \cdot n_{r,\zeta}\right) + \\ w_z\left(\zeta_x \cdot n_{r,\xi} + \zeta_y \cdot n_{r,\eta} + \zeta_z \cdot n_{r,\zeta}\right) + \\ \left(t_\xi \cdot n_{r,\xi} + t_\eta \cdot n_{r,\eta} + t_\zeta \cdot n_{r,\zeta}\right) \end{gathered} \tag{43}$$

If we define the elements:

$$\begin{gathered} N_\xi = \left(\xi_x \cdot n_{r,\xi} + \xi_y \cdot n_{r,\eta} + \xi_z \cdot n_{r,\zeta}\right); \\ N_\eta = \left(\eta_x \cdot n_{r,\xi} + \eta_y \cdot n_{r,\eta} + \eta_z \cdot n_{r,\zeta}\right); \\ N_\zeta = \left(\zeta_x \cdot n_{r,\xi} + \zeta_y \cdot n_{r,\eta} + \zeta_z \cdot n_{r,\zeta}\right); \\ k = \left(t_\xi \cdot n_{r,\xi} + t_\eta \cdot n_{r,\eta} + t_\zeta \cdot n_{r,\zeta}\right), \end{gathered} \tag{44}$$

equation (33) becomes:

$$\begin{gathered} \left(\frac{\xi_x}{\delta_u} - \frac{\left(u-u_0\right)\cdot N_\xi}{f}\right)\cdot w_x + \left(\frac{\eta_x}{\delta_u} - \frac{\left(u-u_0\right)\cdot N_\eta}{f}\right)\cdot w_y + \left(\frac{\zeta_x}{\delta_u} - \frac{\left(u-u_0\right)\cdot N_\xi}{f}\right)\cdot w_z + \\ \frac{u-u_0}{f}\cdot k = -\frac{t_\xi}{\delta_u} \end{gathered} \tag{45}$$

An analogous relation can be written for second equation of (33):

$$\begin{gathered} \left(\frac{\xi_y}{\delta_v} - \frac{\left(v-v_0\right)\cdot N_\xi}{f}\right)\cdot w_x + \left(\frac{\eta_y}{\delta_v} - \frac{\left(v-v_0\right)\cdot N_\eta}{f}\right)\cdot w_y + \left(\frac{\zeta_y}{\delta_v} - \frac{\left(v-v_0\right)\cdot N_\xi}{f}\right)\cdot w_z + \\ \frac{v-v_0}{f}\cdot k = -\frac{t_\eta}{\delta_v} \end{gathered} \tag{46}$$

By arranging equation (45) and (46), it is possible to redefine the stereoscopic problem, expressed by equation (40):

$$P\left(u, u, u', v'\right)\cdot w = S \tag{47}$$

In equation (47) P is a matrix 4x3, whose elements depend only by (u,v) and (u',v'), and B is a vector 4x1, whose elements contain parameters of cameras configuration.
The expression of matrix P is:

$$P=\begin{bmatrix}\frac{\xi_x}{\delta_u}-\frac{(u-u_0)\cdot N_\xi}{f} & \frac{\eta_x}{\delta_u}-\frac{(u-u_0)\cdot N_\eta}{f} & \frac{\zeta_x}{\delta_u}-\frac{(u-u_0)\cdot N_\xi}{f}\\ \frac{\xi_y}{\delta_v}-\frac{(v-v_0)\cdot N_\xi}{f} & \frac{\eta_y}{\delta_v}-\frac{(v-v_0)\cdot N_\eta}{f} & \frac{\zeta_y}{\delta_v}-\frac{(v-v_0)\cdot N_\xi}{f}\\ \frac{\xi'_x}{\delta_{u'}}-\frac{(u'-u'_0)\cdot N_{\xi'}}{f'} & \frac{\eta'_x}{\delta_{u'}}-\frac{(u'-u'_0)\cdot N_{\eta'}}{f'} & \frac{\zeta'_x}{\delta_{u'}}-\frac{(u'-u'_0)\cdot N_{\xi'}}{f'}\\ \frac{\xi'_y}{\delta_{v'}}-\frac{(v'-v'_0)\cdot N_{\xi'}}{f'} & \frac{\eta'_y}{\delta_{v'}}-\frac{(v'-v'_0)\cdot N_{\eta'}}{f'} & \frac{\zeta'_y}{\delta_{v'}}-\frac{(v'-v'_0)\cdot N_{\xi'}}{f'}\end{bmatrix} \tag{48}$$

The expression of vector S is:

$$S=\begin{Bmatrix}-\frac{t_\xi}{\delta_u}-\frac{u-u_0}{f}\cdot k\\ -\frac{t_\eta}{\delta_v}-\frac{v-v_0}{f}\cdot k\\ -\frac{t_{\xi'}}{\delta_{u'}}-\frac{u'-u'_0}{f'}\cdot k'\\ -\frac{t_{\eta'}}{\delta_{v'}}-\frac{v'-v'_0}{f'}\cdot k'\end{Bmatrix} \tag{49}$$

By equation (47) it is possible to invert the problem that is described by eqs. (33) and to recognise the position of a generic point starting to its camera image, by means of pseudoinverse matrix P+ of matrix P.

$$P\cdot w=S \Rightarrow P^T\cdot P\cdot w=P^T\cdot S\Rightarrow w=\left(P^T\cdot P\right)^{-1}\cdot P^T\cdot S\Rightarrow w=P^+\cdot S \tag{50}$$

By means of equation (50), it is possible to solve the stereoscopic problem in all configurations in which is verified the condition:

7. The camera calibration [2, 18]

In order to determine the coordinate transformation between the camera reference system and robot reference system, it is necessary to know the parameters that regulate such transformation. The direct measure of these parameters is a difficult operation; it is better to identify them through a procedure that utilize the camera itself.
Camera calibration in the context of three-dimensional machine vision is the process of determining the internal camera geometric and optical characteristics (intrinsic parameters) and/or the 3-D position and orientation of the camera frame relative to a certain world coordinate system (extrinsic parameters). In many cases, the overall performance of the machine vision system strongly depends on the accuracy of the camera calibration.

In order to calibrate the cameras a toolbox, developed by Christopher Mei, INRIA Sophia-Antipolis, was used. By means of this toolbox it is possible to find the intrinsic and extrinsic parameters of two cameras that are necessary to solve the stereoscopic problem. In order to carry out the calibration of a camera, it is necessary to acquire any number of images of observed space in which a checkerboard pattern is placed with different positions and orientations.
In each acquired image, after clicking on the four extreme corners of a checkerboard pattern rectangular area, a corner extraction engine includes an automatic mechanism for counting the number of squares in the grid. This points are used like calibration points, fig. 5.
The dimensions dX, dY of each of squares are always kept to their original values in millimeters, and represent the parameters that put in relation the pixel dimensions with observed space dimensions (mm).

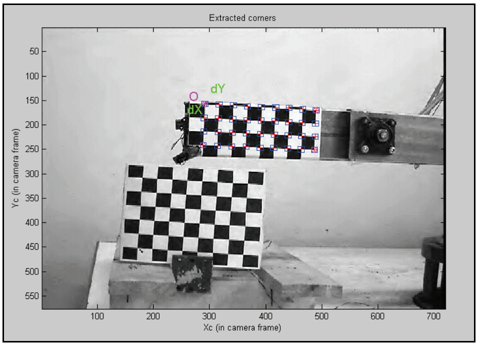


Figure 5. Calibration image

After corner extraction, calibration is done in two steps: first initialization, and then nonlinear optimization.
The initialization step computes a closed-form solution for the calibration parameters based not including any lens distortion.
The non-linear optimization step minimizes the total reprojection error (in the least squares sense) over all the calibration parameters (9 DOF for intrinsic: focal (2), principal point (2), distortion coefficients (5), and 6*n DOF extrinsic, with n = images number).
The calibration procedure allows to find the 3-D position of the grids with respect to the camera, like shown in fig. 6.

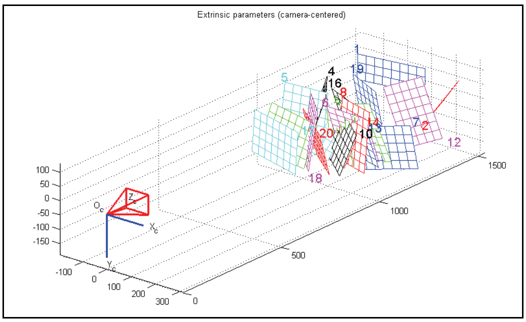


Figure 6. Position of the grids for the calibration procedure