A system based on stereo vision for unpacking pieces of wood with an industrial robot

B. Allotta*, C. Colombo†, F. Bertocci*, L. Scarselli*, M. Sgarbi°, G. Bioli°, V. Colla°

° PERCRO Lab, Scuola Superiore Sant'Anna
Polo Sant'Anna Valdera, Viale R. Piaggio 34, 56025 Pontedera (PI)
E-mail: sgarbi {gbioli,colla}@sssup.it
* Università di Firenze, Dipartimento di Energetica "S. Stecco"
E-mail: allotta@unifi.it fbertocci@medea-italia.it
† Università di Firenze, Dipartimento di Sistemi e Informatica
E-mail: colombo@unifi.it

Abstract

The paper describes a system based on stereo vision which allows the automatisation of the unpacking of a stack of partially manufactured pieces of wood by means of an industrial robot. The proposed solution is not expensive and does not require heavy constraints on the working environment; its general features make it easy to apply with small modifications also in different contexts.

1 Introduction

The technology and expertise related to vision system has nowadays become so consolidated that such systems can play an important role in the production processes of many industrial sectors to the aim of increasing production, reducing the costs and improving the working conditions and job quality of the personnell. Typical task of vision systems concerns, among the others, quality controls [1], objects positioning and teleguidance [2], objects selection and classification. The control unit of industrial robots are often integrated with a vision system that provides useful information for the robot control and motion planning.

While in some high-tech industries (such as those operating in the fields of automotive, electronic components, domestic appliances, etc...) the importance of vision systems is widely recognised and their presence is commonly accepted and continuosly increased, in other industrial sectors, that are very important especially in the Italian and Tuscan scenario, such as wood, paper and leather manufacturers, vision systems are still considered with strong suspicion, although the technical level of the machinery commonly found in these industries is very high. There are indeed non negligible obstacles to the application of vision systems in the abovementioned fields, due to the difficult conditions of the working environment: poor lighting, powder, chemical substances, humidity, high temperatures, frequent collisions are common factors that, summed with the low instruction level of the technical personnel normally employed in the production stages, make the use of vision systems unappealing, as they are frequently considered not robust, unreliable and expensive; thus, many dangerous and/or repetitive operations are still performed by human operators. The only way to encourage the interest of the manufacturer toward the vision systems is to make them envisage the possibility of eliminating some working shifts, especially during the night, and/or of increasing the production by overcoming some "bottle necks" in the production cycles which, at the moment, are due to the needing of connecting two automatic phases of the production itself (normally performd by machineries working at very high speeds) by means of the intervention of human operators.

The paper describes a system based on stereo vision to be applied in the wood industry, but whose characteristics are quite general and make it re-usable with small modifications also in other industrial fields. Moreover, the system could be useful also in non-industrial applications, where repeatitive posistioning of similar objects must be performed at quite high velocities by means of a robot. The emphasis of the paper is on the vision system, i.e. on the solutions that has been devised for conditioning the original working environment so that the system can work and on the tuning of the system parameters, rather than on the robot control, that indeed could be optimised in order to reduce the times required to perform a single operation.

The paper is organised as follows: Sec. 2 is devoted to the description of the problem for whose solution the system has been designed, while Sec. 3 describes the foundamental idea of the proposed solution. Sec. 4 treats the method adopted for determining the system parameters; Sec. 5 describes the standard procedure for obtaining the coordinates of a point in the space from its two images captured by two cameras and Sec. 6 is focused on the evaluation of the point in which each pieces is gripped by the robot end–effector. Sec. 7 describes the experimental setup that has been builded for the final experiments and, finally, some concluding remarks are reported in Sec. 8.

2 Description of the problem

In the industrial production of furnitures and other widely used objects, the wood pieces are manufactured by numerically controlled machining centers that work at very high velocities and produce a considerable number of pieces per hour; such pieces are manually packaged in frameworks usually named *pallet*. Once the pallet is complete, it must be moved toward the machining center performing the subcessive manufacture; afterwards the pieces are provided, singularly or in small groups, to the machine itself by a human operator. The operations of pallet packing and unpacking are very ripetitive and tedious for the operator and can be also dangerous, if, for instance, the machining centers produce noise and/or unhealthy powder. Also from the production point of view, the needing for the human intervention in these operations heavily slows down the production cycle, makes at least partially useless the effort for the achievement of high speeds performed by the wood machining centers designers. On the other hand, the automatisation of the pallet packing and unpacking operations would reduce the costs related to the human work and would allow to continue the production even when no human supervision is available, for instance during the night.

The paper describes a system based on stereo vision which allows the automatisation of the unpacking of a pallet composed by partially manufactured pieces of wood like the one depicted in Fig. 1 by means of an industrial robot.

The proposed system can find applications both in similar contexts and in non–industrial applications, as no assuptions have been made neither on the type of robot adopted nor on the surrounding environment; at the moment foundamental conditions for the system usage are normal lighting and pieces to be moved that are quite matt and mainly planar. The attempt to overcome at least these latter two constraints is the object of the present work, as well as the further reduction of the needing for environmental conditioning. Nevertheless the system is already appliable to a quite large number of situations.

Figure 1: An industrial pallet containing partially manufactured pieces of wood (in particular parts of chairs).

3 The proposed solution

The industrial derivation of the problem that have been faced had the positive effect to force the system designers to take into account two foundamental objectives:

- to reduce at the most the costs related to the vision system, that is designed as a tool of an industrial robotic system, which is by itself very expensive;
- to make the system installation as easy as possible, so as not to need fondamental changes in the working environment.

As it clearly appears in Fig. 1, in an industrial pallet the pieces are very compactly arranged. Such arrangement does not guarantee enough space between two neighboring pieces for allowing the grip with a robotic tool equipped with two jaws; thus the pieces was ordered so as to make the grip easier for an industrial manipulator. Moreover the considered pieces have very different shapes and high tolerances are admitted with respect to their nominal dimensions. Stereo vision fits well to the proposed situation, as it allows to recognise the three dimensions of the pieces, despite of their dimensional variability, to infer their spatial position and orientation and, therefore, to work on slightly structured pallets, without very significative modifications of their original form.

The ordered arrangement of the pieces on each level of the pallet has been obtained with a couple of wooden sticks, each one carrying a series of colored markers. The markers are small cilinders with a diameter of 8 mm; their upper base is colored with a different color for each stick (see Fig. 2). The sticks allow to separate the different levels of pieces, while the markers allow a sufficient separation among the pieces so as to allow the robot grip. The color of the markers provide a useful reference in the tridimensional reconstruction of the pieces, to the aim of ensure an easy and safe grip.

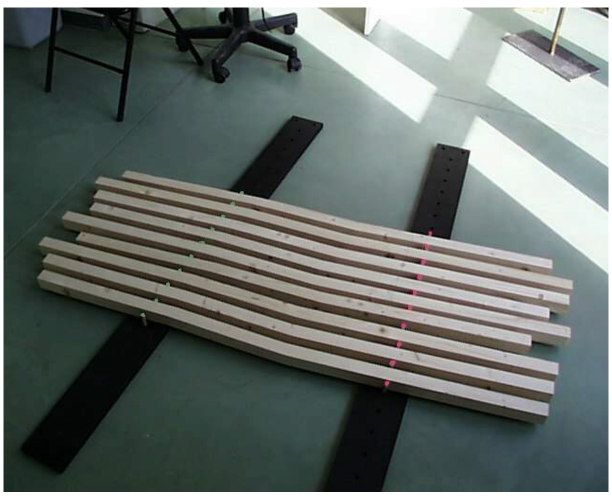

Figure 2: The system for pallet structuring that exploits colored markers.

The vision system is composed of a couple of television cameras linked to a PC by a Frame Grabber (MATROX Meteor II). The PC that has been used is a Pentium II 300 MHz (with a 128 Mb RAM) where the MS Windows NT4 operating system is installed. The experimental setup that has been used for the first trials is depicted in Fig. 3, but, once the whole software had been developed, more significative experiments was performed with a true industrial robot.

4 The system calibration

The system need to be calibrated when the cameras are moved, in order to determine the cameras positions and orientations with respect to a fixed reference frame.

Figure 3: The system for pallet structuring that exploits colored markers.

The image coordinates (x_{im}, y_{im}) (expressed in pixels) of a point P are linked to its spatial coordinates $[X^w, Y^w, Z^w]^T$ (expressed in the desired unit and referred to the chosen spatial reference frame) by the following equations [3] [4]:

$$\begin{aligned} x_{im} - o_x &= -f_x \cdot \frac{r_{11}X^w + r_{12}Y^w + r_{13}Z^w + T_x}{r_{31}X^w + r_{32}Y^w + r_{33}Z^w + T_z} \\ x_{im} - o_y &= -f_y \cdot \frac{r_{21}X^w + r_{22}Y^w + r_{23}Z^w + T_y}{r_{31}X^w + r_{32}Y^w + r_{33}Z^w + T_z} \end{aligned} \quad (1)$$

where

$$\mathbf{M} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & T_x \\ r_{21} & r_{22} & r_{23} & T_y \\ r_{31} & r_{32} & r_{33} & T_z \end{bmatrix}$$

is the matrix of the *extrinsecal parameters*, namely the roto-translation matrix that allows to overlap the image reference system to the spatial reference system, that is defined by the user, and

$$\mathbf{A} = \begin{bmatrix} f_x & 0 & o_x \\ 0 & f_y & o_y \\ 0 & 0 & 1 \end{bmatrix}$$

is the matrix of the *intrinsecal parameters*, such as the focal length, expressed in the horizontal and vertical units of the pixels (f_x, f_y), and the coordinates of the image center (o_x, o_y).

To calibrate the system means to determine all the above mentioned parameters. To this aim, the *calibration pattern* depicted in Fig. 4 has been realised, that is an object with a very simple shape and easy to be recognised within an image. Such pattern is a cube in plexiglass with a 50 cm long edge; 24 red sqares with a 35mm long side are depicted on two of its sides.

Figure 4: The calibration pattern.

The calibration procedure consists of the acquisition of two different simultaneous images of the calibration pattern, from which the coordinates of the centroids of the red squares are evaluated; as the spatial position of the squares centers is *a priori* known, eq. 1 allows to determine the parameters of the two cameras, that characterise the geometry of the stereo system. The use of centroids is simpler and more reliable with respect to the use of corners, but is subjected to perspective errors.

The calibration procedure must be performed once, unless the position and/or orientation of the two cameras is changed. An accidental change of the cameras positions implies the repetition of the calibration procedure; for this reason the cameras that focuse the calibration pattern must be located so as to view the scene containing the pieces to manipulate.

The above described calibration procedure has the advantage of the simplicity, but indeed its precision is limited to the portion of the robot workspace in which the calibration pattern is located. Another more precise calibration procedure uses the robot itself, as its end-effector could be exploited to carry to predefined positions an object with simple shape, possibly pointwise and colored, that could be easily pointed out in the images provided by the two cameras.

5 The triangulation procedure

The triangulation procedure allows to determine the point in wich each piece must be gripped with a precision sufficient for the robotic grip. Also in this case eq. 1 is applied, but now the cameras parameters are known and the spatial coordinates of the point of interest are evaluated. The problem is thus characterised by 2 equations and 3 unknowkns, thus two couples of image coordinates of the same point in the space are needed in order to evaluate its three-dimensional position vector; so doing the system is overdimensioned, as 4 equations and 3 unknowkns are present, therefore an approximation will inevitably occurr in the evaluation of the position vector itself [5] [6].

6 Evaluation of the gripping point

Figure 5 shows an example of the kind of coupled images that have been exploited for determining the gripping point.

The partially manufactured pieces of wood that have been used in the experiments are composed of one, two or three linear segments, as depicted in Fig. 6, that shows also the tolerance that have been admitted on the length of some of these segments.

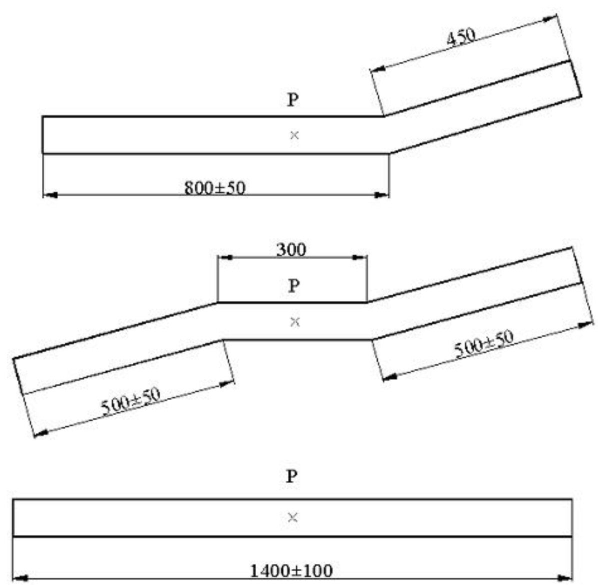

Figure 6: The three possible shapes of partially manufactured pieces of wood that have been used in the experiments; the length of the segments are expressed in mm.

The point in the center of the middle line of the central segment has been chosen as the optimal gripping point. In order to grip the piece, the robot control unit need to know the spatial coordinates (X^g, Y^g, Z^g) of

a) b)

Figure 5: Two simulatneous images of experimental pallet provided by the two cameras.

the gripping point and the orientation of the piece. As the gripping point is located on the upper side of each piece, that can be considered planar, Z^g coincides with the height with respect to a fixed reference frame of the upper base of the pallet. The first step of the procedure determines such plain by exploiting the presence of the colored markers. The two acquired images of the pallet are filtered in order to enhance the colors; the image coordinates of each colored marker, that is considered pointwise, are computed and suitably ordered and, afterwards, their spatial coordinates are evaluated through triangulation. Z^g is thus computed by means of the least squares procedure. In order to recover X^g and Y^g, virtual images of the scene as it was seen from a direction perpendicular to the upper plane of the pallet itself are obtained from the two available images. An example of such virtual images is reported in Fig. 7, where it is evident that all the pixels not belonging to planes parallel to the upper plane of the pallet are subjected to heavy distorsions. This is not important, because they do not affect the procedure.

The two virtual images that are obtained by the real images provided by the two cameras and contain a top view of the pallet are obviously very similar, thus one only of them is exploited in the remaining part of the procedure.

From the image providing a top view of the piece of interest, that is furtherly filtred for contrast enhancing, the piece contours are determined and its medium line is pointed out: such line is approximated by means of one segment, in the case of straight pieces, or eventually two or three jointed segments, as de-

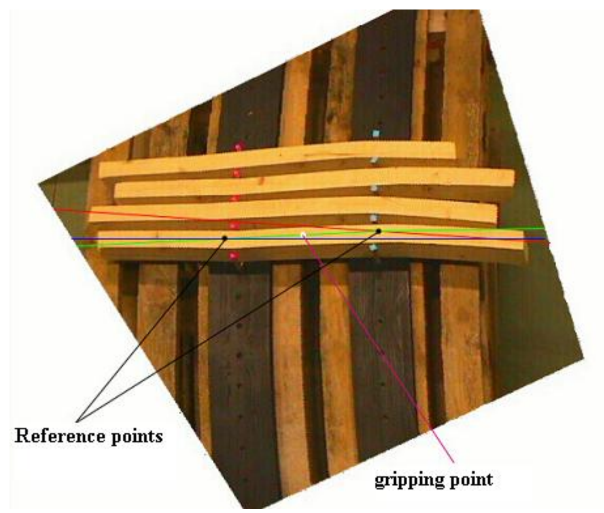

Figure 7: An example of virtual image of the pallet. The straight lines containing the three jointed segments that approximate the medium line are reported superimposed to the image of a piece

picted in Fig. 7. Afterwards, on the basis of the pieces geometry and by exploiting the intersections between the straight lines containing the above mentioned segments (the so–called "reference points" that are put into evidence in Fig. 7), the first two coordinates of the gripping point, i.e. X^g and Y^g, are evaluated. Fig. 7 shows the results of such operation in the case of the reported image. Moreover, thanks to the approximation of the medium line of the pieces with straight lines, the orientation of the robot end–effector that is used for gripping is directly evaluated from the spatial orientation of the line on which the gripping point is located.

7 The experimental setup and results

The system has been interfaced with an industrial manipulator that grasps the piece by means of a simple device equipped with three suckers, that is depicted in Fig. 8.

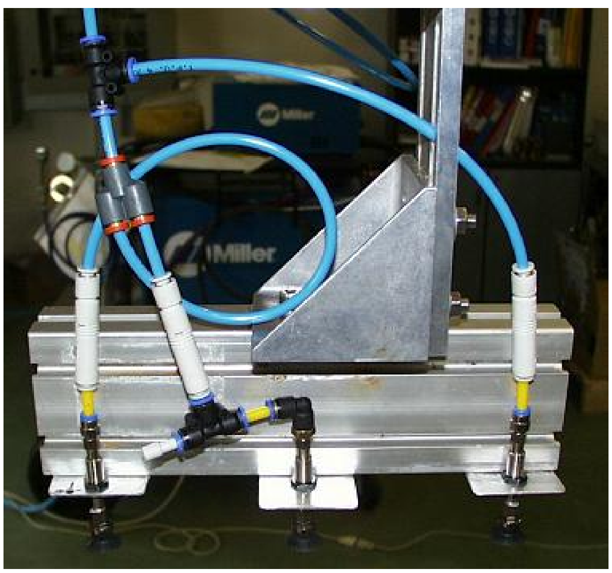

Figure 8: The gripping device.

Fig. 9.a illustrates the setup that has been build in order to perform the experiments in a realistic scenario: an industrial robot with 6 joints was used for moving the pieces; its control has not been optimised, namely no study has been performed on the possibility of optimising the trajectories and thus reducing the time required for each piece. Fig. 9.b shows the robot gripping the piece.

The calibration procedure performed with the calibration pattern allows to achieve a maximum error in spatial reconstruction of about 3 mm, that is highly sufficient for the purpose of the considered application.

The accurancy obtained in the determination of the gripping point is not high; Fig. 10 shows the desired and computed gripping points for some pieces with three joint segments. The average absolute distance among the desired and the evaluated gripping point is about $10mm$: the estimated gripping point lies with good approximation on the real median axis of the pieces, namely the maximum distance of such point from the axis is of about 3 mm, while a greater error arises along the direction of the axis itself. Anyway, such accuracy is high enough for the present application, due to the light weight of the pieces and the particular shape of the gripping device, whose three suckers ensure a safe grip. It must be underlined that in the pieces used for the experiments, the angles between the two or three segments that eventually compose the pieces themselves are almost equal to π, namely the pieces are quite similar to a unique straight segment. This choice have been made according to the "strategy of the worst case," i.e. because the reproduced cathegory of the pieces is the more critical case and indeed can arise when manufacturing, for instance, parts of chairs or tables. Clearly if the two or three straight lines to be found are more perpendicular with respect to the presented case, the reference points can more easily and reliably be found and the average error in the evaluation of the gripping point dramatically decreases.

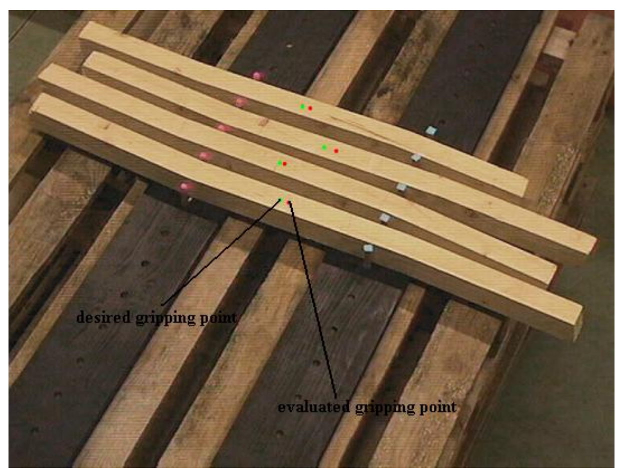

Figure 10: Some examples of location of desired and evaluated gripping points on the pieces.

8 Conclusion

A system based on stereo vision to be applied in the wood industry has been presented; the main ideas underlying this system are the separation of the partially manufactured pieces to be gripped and their

a) b)

Figure 9: a) The final experimental setup. b) The industrial robot grips the piece.

schematisation through their medium line, which is approximated via a stepwise linear function and this allows to determine the gripping point with a sufficient accuracy. The idea of exploiting a schematic although incomplete description of the pieces in order to find the relevant information (i.e. the position of the gripping point) is widely general and can be extended to different objects. The almost planar geometry of the pieces is a quite strict constraint, but is mainly due to the adopted gripping device, thus is does not depend on the strategy adopted for processing the images provided by the two cameras. On the other hand, an actual constraint for the system is that the system should not be made in metal or other materials that can reflect the light, as at the moment this kind of phenomena are neither considered nor somehow compensated. Another limitation for the system is indeed the needing of the colored markers for the separation of the pieces to be gripped; their presence halves the pallet density and this is a real obstacle for the application of the system without the presence of a human operator (for instance during overnight shifts); in the application for which the system as been studied, namely to feed a wood machining center that manufactures a great number of pieces per hour, the colored markers should be avoided but in different applications, where the number of pieces to be manufactured is not high or where the colored markers can be mounted on some pre-existent structures, the markers can be mantained. Current work is focused on the elimination of the markers constraint in order to make the system more appealing from the industrial point of view.

References

[1] Fravolini, Ficola, Gabrielli, La Cava, Ghirelli “Controllo qualit di prodotti industriali. Un approccio sistematico,” *Automazione e strumentazione*, n.3, March 1999.

[2] Meeden, “An incremental approach to developing intelligent neural network controllers for robots,” *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 26, n. 3, pp. 474-485, June 1996.

[3] P. Ioannis “Digital image processing algorithms,” *Prentice-Hall*, 1993.

[4] T. Pavlidis “Algorithms for Graphics and Image Processing,” *Computer Science Press, Inc*, 1982.

[5] D.H. Ballard, C.M. Brown “Computer Vision,” *Prentice-Hall*, 1982.

[6] E. Trucco, A. Verri “Introductory Techniques for 3-D Computer Vision”, *Prentice-Hall*, 1998.