

Focus based Feature Extraction for Pallets Recognition

Rita Cucchiara*, Massimo Piccardi⁺ and Andrea Prati*

*D.S.I., Univ. di Modena e Reggio Emilia, Via Campi, 213/b,
Modena, Italy ⁺D.I., Univ. di Ferrara, Via Saragat, 1, Ferrara, Italy
{rita, prati}@dsi.unimo.it,mpiccardi@ing.unife.it

Abstract

Visual recognition for object grasping is a well-known challenge for robot automation in industrial applications. A typical example is pallet recognition in industrial environment for pick-and-place automated process. The aim of vision and reasoning algorithms is to help robots in choosing the best pallets holes location. This work proposes an application-based approach, which fulfil all requirements, dealing with every kind of occlusions and light situations possible. Even some "meaning noise" (or "meaning misunderstanding") is considered. A pallet model, with limited degrees of freedom, is described and, starting from it, a complete approach to pallet recognition is outlined. In the model we define both virtual and real corners, that are geometrical object proprieties computed by different image analysis operators. Real corners are perceived by processing brightness information directly from the image, while virtual corners are inferred at a higher level of abstraction. A final reasoning stage selects the best solution fitting the model. Experimental results and performance are reported in order to demonstrate the suitability of the proposed approach.

1 Introduction

Object recognition is one of the most important topic of machine and robot vision; it is the basic task for handling robot navigation and automatic visual inspection [11]. In many cases the object to be recognized is well known and a model based on its structural and geometrical parts can be a-priori defined and used for the final reasoning step on the visual features extracted by image data. Between models and model matching approaches, the most commonly adopted method is tree search that allows for classifying the computed features in one or a set of acceptable models. Classification is provided by evaluating the presence and the attributes of some features and the relations between them. Examples of typical features are object corners or object profiles and main borders. Corners are normally adopted since they should be invariant with the object translation and scaling, and therefore allow for reliable recognition independently of the point of view. In this paper we address a recognition problem that is mainly bi-dimensional since we aim to detect objects that are tri-dimensional, but observed from points of view from where only the frontal surface can be seen. The difficulty of the recognition task arises from the complexity of the indoor scene, consisting of a real industrial environment where luminance conditions can be highly variable and many spurious objects must be disregarded.

The recognition task is a very common robotic application: the detection of pallets. Pallets can be completely present in the scene or only partially viewed depending on the camera position. This has been accomplished by a sequence of tasks able to extract visual primitives used to recognize industrial pallets, in order to allow the forking of them by a mobile robot. In particular, robot has to pick up the higher (or first) pallet in a stack. Nevertheless, our system has been developed to easily relax this constraint, so to deal with different specifications. The first hypothesis is that the sensor (hosted on the robot) is located so as to acquire possible pallet profiles as in the example images of Fig. 1. To achieve the goal, both whole pallet and partially visible ones have to be recognized. The goal of the system is to identify the central foot, in order to direct the robot towards it.



Figure 1: **Example images**

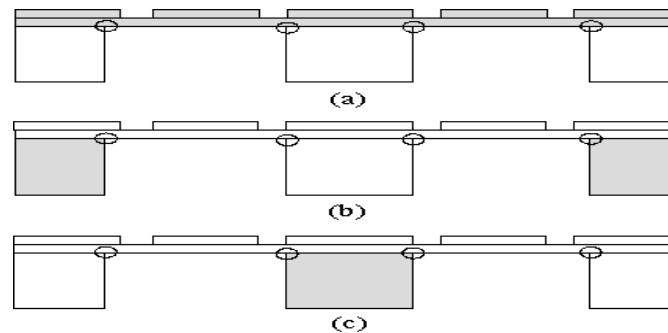


Figure 2: **Pallet model: (a) loading plane (b) side feet (c) central foot**

A detailed model of the pallet has been a-priori defined (Fig. 2), in order to extract from images the visual primitives needed. In particular, we define a hierarchy of visual features, consisting in:

- a set of salient features specifically oriented to the model used for defining a region of interest (ROI) where to better focus the search
- a set of features (eventually redundant) exploited for providing classification in order to identify the object corresponding with the model and to select possible locations.

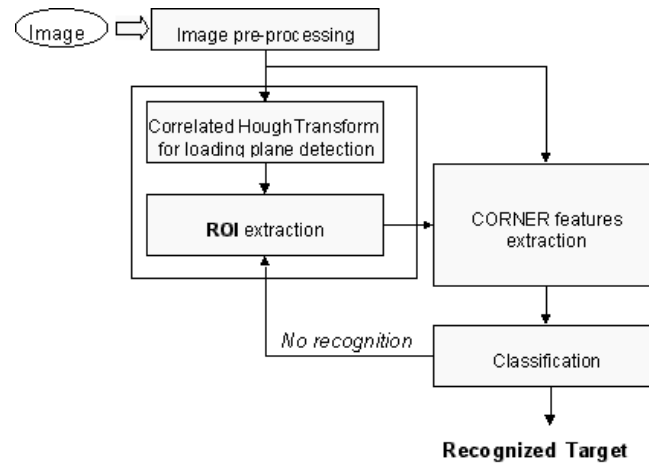
In the paper we first describe how the most salient feature is computed in the image. In section 2 we present a suitable modification of the well-known Hough transform that is robust enough for selecting the loading planes with high reliability. After the loading plane has been located, a ROI (Region Of Interest) is extracted from the image. According to this model, we can define the visual features needed for the recognition, called *virtual corners* and *real corners*. The former ones are inferred by a higher level of abstraction able to detect them by intersection of straight lines. The last ones are detected using one of the corner detection algorithms present in the literature [15]. Both approaches have been considered to better explore their potentiality in section 3. A final reasoning phase (section 4) based on decision trees, is able to classify the object as a complete and correctly recognised pallet or as a partial pallet or eventually a non-pallet object. A set of real images has been collected, in order to deal with all possible situations. Some of them are used as a training set for tuning both manually and automatically (with tools such as C4.5 [13]) the classification tree parameters. Other images are used as a final test set. Section 5 shows some experimental results, both by superimposing the detected pallets to the original images and by performance expressed as miss detections and false detections. The conclusion presents final considerations on the profitability of the presented approach.

2 Architecture of the recognition system and ROI Extraction

In complex applications, restricting the focus of attention could be very useful both for improving performance and for avoiding errors in the recognition task. Since the goal of the system is (as stated in previous section) the recognition of the higher loading plane of the pallet, we can focus our attention to a limited ROI centred on the most salient feature characterizing the loading plane found. In this ROI other eventually redundant object features can be computed and used in a classification system for matching the feature set on the model feature set. According to this, the system architecture has been structured as shown in Fig. 3.

First, the original image is pre-processed and enhanced by contrast stretching [12] to deal with different luminance conditions. Due to the rectilinear nature of the loading plane, the most immediate approach is to detect straight lines, i.e. using the Hough transform (HT) [8]. Of the several variants to the HT, we have adopted two: the CHT (Correlated HT) and the EHT (Edge-based HT), both based on straight lines. The Correlated Hough Transform (CHT) [2] is an enhancement of the GBHT (Gradient based HT) [5]: GBHT works in the space between 0 and 2π and weights the HT votes with gradient of the image points; moreover it votes in the gradient direction only. The CHT exploits all the GBHT information and moreover correlates the HT space in order to detect long and thin objects that can be described by two main roughly straight edges at a short distance from each other [3].

The features selected by the CHT well fit with the three characteristics of the upper and lower straight lines delimiting the loading plane (described above). Couples of straight lines with high CHT votes are inserted into a list in decreasing order from top to bottom in image space: in the ROI extraction step this list is scanned beginning from its top, to achieve, if possible, the recognition of the current candidate of pallet in the focused region. Once the loading plane has been identified by its salient feature, a ROI containing it is

Figure 3: **The overall architecture**Figure 4: **ROI extraction example**

extracted from the image. To accomplish this, first the length of the ROI is computed and then the height is obtained by it with simple considerations on the geometrical structure of the candidate pallets. Resulting ROI is shown in Fig. 4. In this ROI the candidate pallet is searched by extracting geometric features, such as corners, and using a decision tree for classifying spatial relation between them. If a pallet is not recognized another CHT peak is selected as starting loading plane and another ROI is computed. The process is iterated from the highest part on the image up to the lowest part or until CHT peaks are detected.

3 Corner feature extraction

The mobile robot needs not only the information about the region where pallet could be located, but also the exact location of the pallet's central foot in order to fork it. Therefore information about pallet structure as upper forking holes corners or feet position have to be extracted from the image. In the present work we have adopted two parallel and independent approaches for supporting the decisional phase: virtual corners extraction and real corners extraction. In other words, two types of visual features associated with cor-

ners have been extracted in order to recognize a pallet in a ROI by means of a redundancy of information. To this aim we define:

- ◊ *Virtual corners*: the ideal corners that we could imagine at the intersection on the main horizontal and vertical lines composing the pallet. Therefore we call them "virtual" since they are not perceived in the imaged but are mathematically computed after straight line detection
- ◊ *Real corners*: the object corners that are perceivable in the image due to the contrast between object and background grey levels; real corners are just an approximation of the actual object corners since their computation with image processing is affected by errors due to shadows and luminance variations.

Each method needs edge filtering to extract vertical straight lines belonging to pallet feet or borderlines surrounding pallet regions. Thus, the corner detection module interacts with ROI extraction module, to focus the analysis of point features in the current region of interest, avoiding false visual primitives to be selected and obtaining an improvement in computation complexity.

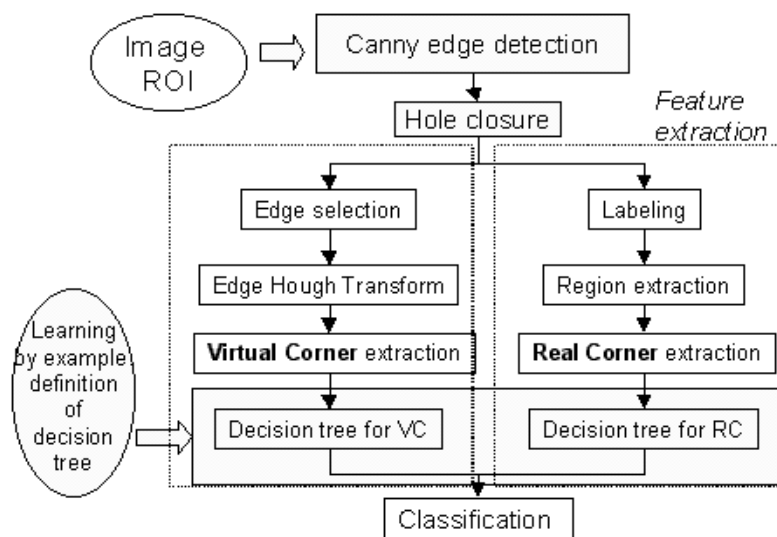


Figure 5: **Corner feature extraction**

Corner features are extracted with a sequence of image processing and analysis tasks. These steps are summarized in Fig. 5. In parallel with the ROI computation, a first phase of edge detection on the image enhanced by contrast stretching is provided. The operator used in this work is the *Canny edge detector*, given that is one of the most effective algorithms traditionally adopted in grey level images. It allows for computing possibly closed edges [1]. Some simple but indispensable image processing operators are used for improving corner detection. In particular morphological *hole closure* is exploited for improving the computed edges.

Virtual corners are extracted as virtual intersection between the pallet main lines. It is performed with the classic Edge Hough Transform (EHT) computed on the Canny edge

image after an edge selection step. The edge selection is a semantic noise reduction: it eliminates all edge segments that appear to be too short w.r.t the object geometry. The selection of the threshold is depending on the ROI length, which is itself dependent on the distance of the potential target to the camera. The Hough votes are accumulated only for that angles, representing vertical direction (within a given tolerance), so as to search vertical lines only (representing possible vertical borders).

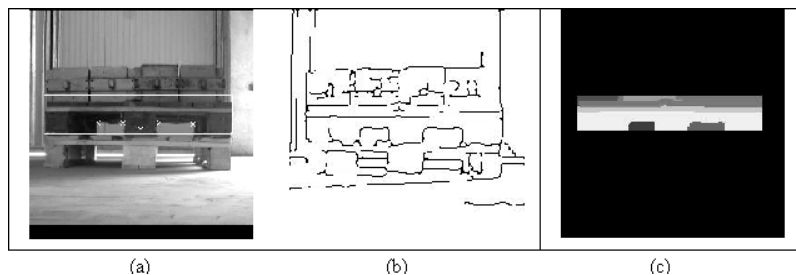
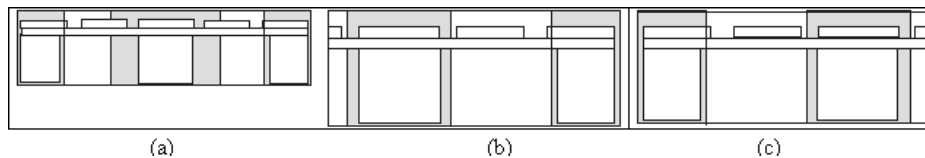
Virtual corners identification is achieved by intersecting only internal vertical lines with lower straight line of loading plane, since upper corners of the forking holes are defined as intersection between them. Given the line equations of the loading plane line and the four possible vertical lines, $\rho_i = x \cos \theta_i + y \sin \theta_i$ and $\rho_l = x \cos \theta_l + y \sin \theta_l$ with $l=1..4$, the corner coordinates can be immediately extracted as:

$$x = \frac{(\rho_i - \rho_l \frac{\cos \theta_i}{\cos \theta_l}) \sin \theta_l}{\cos \theta_l - (\sin \theta_i - \tan \theta_l \cos \theta_i)} \quad y = \frac{(\rho_i - \rho_l \frac{\cos \theta_i}{\cos \theta_l})}{(\sin \theta_i - \tan \theta_l \cos \theta_i)}$$

It's important to note that virtual corners don't match to real corners in original image exactly, given that they are a result of mathematical techniques carrying unavoidable approximations (due in particular to space discretizations); in spite of this, they are important in pallet detection, given that they preserve the information of proportionality and collinearity between corners. Nevertheless, when Canny edge detectors neglects some edges, some corners could be missed.

For this reason, another method has been tested for extracting object corners. In this case we search for **real corners**, following the previously described meaning, i.e. grey level points of the image that could belong to an object corner. In this case we process the grey level image directly. Nonetheless, the grey level image, even in the ROI only, present many particulars, shadows and objects that should be eliminated in order not to have an explosion of the number of possible real corner. For this reason we apply the grey level corner computation only in a region obtained after a region growing and labelling process, defined by some steps:

- after Canny edge extraction and edge closure, the next step regards the image regions extraction, using a hybrid linkage region growing [7]. Grey values based region growing algorithms [16, 10] are not applied due to the possibility of having pallet with different colours in his components, then without a complete homogeneous pallet region to extract. An example of edge images and on the labelled regions in the ROI are shown in Fig. 6
- To obtain only the pallet region from the labelled image an heuristic algorithm is applied. His working principle is simply the selection of the region having the greatest number of pixels in some relevant ROI sectors depending on the ROI position in the image. If ROI is totally internal to the image or is long as the entire image itself, then the central foot is hypothesised to be placed in the centre of ROI and the lateral feet are hypothesised to be in lateral sectors of ROI. Thus the pixels that participate to the vote process are the ones in grey sectors in Fig. 7a. Instead, if ROI is placed in the lateral sides of image, the pallet is hypothesised to be partially present in the scene, so only pixels belonging to lateral or central foot are counted in their heuristic sectors in ROI (as in models of Fig. 7b and 7c)

Figure 6: **Region growing task**Figure 7: **Pallet region extraction**

Real corners are extracted from the pallet region obtained by the previous method. These point features are not interpreted as intersections of image lines but as pattern intensities [15]. Such features prove stable across sequences of images, and are therefore interesting to track objects across sequences [4, 14, 9]. For further details on corner extraction refer to [15]. To select only forking holes corners, a predefined model can be made. In particular a discriminating feature is the gradient's angle of the evaluated corner. Assuming that the pallet region is white over a black background, and the coordinate system is as Fig. 8(a), there are two main types of corners: 315° and 225° . Only corners with a gradient's angle in $315^\circ \pm T$ or $225^\circ \pm T$ (where T is a predefined tolerance, assumed to be 15°) are selected into a list, while the others are rejected. Finally a template matching is applied to the remaining corners in the list, selecting only those with a quadratic error lower than a threshold. The templates used, representing the ideal corner wanted, are shown in Fig. 8.

4 Recognizing object with decision tree

A decision tree based on the extracted features has been defined in order to achieve final pallet recognition. The final goal is to compute the central foot position, taken as reference to guide the robot towards the pallet. With this approach, target recognition is viewed as a tree search aiming at associating extracted features with model features [6]. Search is performed over a binary tree search, where nodes represent pattern conditions and leaves the corresponding object classifications. We defined different possible object classifications, corresponding to the same object from different points of view, in order not only to understand if the object in the ROI is the true target but also if it is completely perceivable (and thus forkable) or if it is viewed only in part. In the classifier, each object is described by an *attribute set*: the number of detected corners, the pallet position (completely inside the ROI, right aligned, left aligned, "close"), the distances between

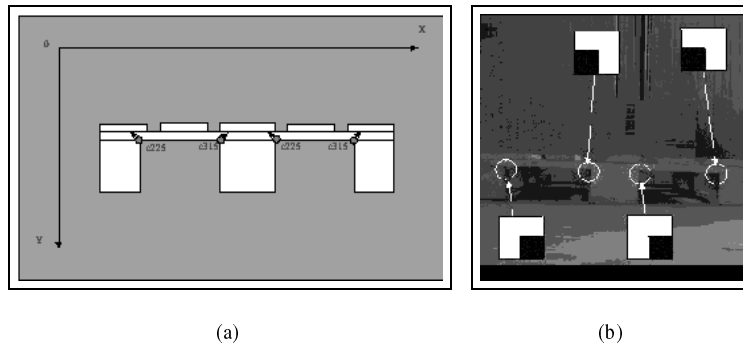


Figure 8: **Corner angles example and template matching**

the corners viewable from the detected position, the distance mutual ratios. To cover all situations of interest, we defined seven classes for classifications, defined with numbers from 0 to 6, where class 0 defines recognition failure while the remaining ones identify all the possible positions of holes to be forked.

We defined two different decision trees for virtual and real corners. Decision trees differ since in the two approaches we can have different types of errors: possibly, more miss-detections are provided by the virtual corner method, while real corner detection extracts generally more corner than the actual ones. Decision trees have been generated by an automatic symbolic decision tree generator, namely the C4.5 program [13]. This automatic classification system, starting from a training set made of several pre-classified images is able of defining an optimised decision tree.

5 Experimental results and performance analysis

The two methods are used in parallel to provide a useful comparison between feature extraction effectiveness and to arrange their results in order to improve the recognition reliability. In this section, experimental results from both methods are presented. To this aim, we have suitably defined the rationale of recognition "success" and "failure": *success* is defined as a correct pallet central foot detection, with a right class attribution, by mean constraints imposition over features extracted. *Failure* is defined as a non-detection of pallet (class 0) or an error in the pallet structure recognition, obtaining a false central foot position and a false classification. So failure is divided into two types: miss detection (no recognition) and false detection (false recognition). We used a test set of images coming from different industrial environments with various luminance conditions and with target located at different distances from the camera. We performed two types of tests: the first one using the same parameters in the decision trees for the overall test set. Obviously, due large differences in test images, specially in pallet dimension (dependently to the pallet-camera distance) specific parameter values could be instead adopted. For this reason we performed another test using different set of parameters for different test images set. The test image database consists of 56 test images which are divided in three classes: *PC*, *GC*, *SQ*. *PC* images represent pallets in a context with a poor contrast, *GC* images represent

pallets with a good contrast, both in an illuminated and dark environments, while *SQ* images represent an approaching sequence, with small and distant pallets.

Methods and used param.	Image type	<i>PC(18)</i>		<i>GC(31)</i>		<i>SQ(7)</i>	
	Error type	MD	FD	MD	FD	MD	FD
<i>Real corners</i> ¹		3	2	2	0	4	0
<i>Virtual corners</i> ¹		3	2	2	4	5	2
<i>Real and Virtual combined</i> ¹		2	3	0	1	2	2
<i>Real corners</i> ²		3	2	2	0	3	0
<i>Virtual corners</i> ²		3	2	2	4	1	0
<i>Real and Virtual combined</i> ²		2	3	0	1	1	0

Table 1: Experimental results

The final decision about recognition combines virtual corner and real corner detection methods: if only one approach recognizes a pallet, then we state that the detection process has success, and the centre of mass of central foot is obtained using the features extracted by this method. If both methods succeed, then the features used to compute the centre of pallet are those extracted by the real corner detection method (since it normally extract more precise corners). As we can see in Table 1, using a unique parameter set for all images, there is a relatively high number of miss-detections (MD) in images with far away pallets (*SQ* images), where many features are distorted or cluttered. For this reason, using specific values in critical conditions can solve the problem improving greatly recognition ability. Table 1 shows, for every method (and the combination of both), the number of miss detections (MD) and false detections (FD) using either a single set or specific values. Combination of methods improves the system recognition ability, especially on miss detection even using a single parameter set. Finally, the best results obtained with optimised parameters is achieved thank to the possibility of selecting a variable neighbourhood size in the computation of EHT local maxima.

6 Conclusion

The aim of the paper is to describe a complete recognition process in a robotic industrial application. A first relevant issue in the approach is the adoption of a hierarchy between visual features in order to extract only the more salient ones, capable of defining a region of interest where other features are computed. This step is necessary in complex environment and in images containing complex objects (that is typical of unconstrained indoor scenes); otherwise the extraction of more refined features (such as corners) can result in a too high number of possible features that must be evaluated and put in relationships each other with a combinatorial explosion of the tree search complexity. Moreover, the more interesting point of the work has been the adoption and comparison of two different approaches for corner detection that can be concurrently used in order to improve the recognition and the classification. Finally, the analysis shows that geometric relations between object parts are only theoretically invariant with object scaling: instead due to

¹Single parameter set

²Optimized values

discretisation errors and unavoidable approximations the object scaling can results in a miss-detection or wrong detection of important features that can affect the classification.

References

- [1] J. Canny. A computational approach to edge detection. *IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI)*, PAMI-8, 1986.
- [2] R. Cucchiara and F. Filicori. A highly selective ht based algorithm for detecting extended, almost rectilinear shapes. In *Proc. of CAIP 95*, September 1995.
- [3] R. Cucchiara and M. Piccardi. Eliciting visual primitives for detecting elongated shapes. *Image and Vision Computing Journal*, 17(5-6):347–355, 1998.
- [4] A. Fusiello, T. Tommasini, E. Trucco, and V. Roberto. Making good features track better. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1998.
- [5] F.O. Gorman and M.B. Clowes. Finding pictures edges through collinearity of feature points. *IEEE Transactions on Computers*, 25(4), April 1976.
- [6] W. Grimson. *Object Recognition by Computer: the Role of Geometric Constraints*. MIT Press, 1990.
- [7] R.M. Haralick and L.G. Shapiro. *Computer and Robot Vision*, volume 1. Addison Wesley, 1992.
- [8] P.V. Hough. *Methods and means to recognize complex patterns*. U.S. Patent 3,069,654, 1962.
- [9] T. Kanade and C. Tomasi. Detection and tracking of point features. In *Technical Report CMU-CS-91-132*, Carnegie Mellon University, April 1991.
- [10] L. Levisaldi and V. Cantoni. *La visione delle macchine*. Tecniche Nuove, 1989.
- [11] T.S. Newman and A.K. Jain. A survey of automated visual inspection. *Comp. Vision and Image Understanding*, 61(2):231–262, 1995.
- [12] W.K. Pratt. *Digital Image Processing*. Wiley-Interscience Publication, John Wiley & Sons Inc., 2nd edition, 1991.
- [13] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.
- [14] C. Tomasi and J. Shi. Good features to track. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 1994.
- [15] E. Trucco and A. Verri. *Introductory techniques for 3D computer vision*. Prentice Hall, 1998.
- [16] S.W. Zucker. Region growing: Childhood and adolescence. *Comput.Graphics Image Process*, 5, 1976.