Matched Filters for Bin Picking

JEAN-DANIEL DESSIMOZ, MEMBER, IEEE, JOHN R. BIRK, MEMBER, IEEE, ROBERT B. KELLEY, MEMBER, IEEE, HENRIQUE A. S. MARTINS, AND CHI LIN I

Abstract—Currently, a major difficulty for the widespread use of robots in assembly and material handling comes from the necessity of feeding accurately positioned workpieces to robots. "Bin picking" techniques help reduce this constraint.

This paper presents the application of matched filters for enabling robots with vision to acquire workpieces randomly stored in bins. This approach complements heuristic methods already reported.

The concept of matched filter is an old one. Here, however, it is redefined to take into account robot end-effector features, in terms of geometry and mechanics. In particular, the proposed filters match local workpiece structures where the robot end-effector is likely to grasp successfully and hold workpieces. The local nature of the holdsites is very important as computation costs are shown to vary with the fifth power of structure size. In addition, the proposed filters tend to have a narrow angular bandwidth.

An example, which features a parallel-jaw hand is developed in detail, using both statistical and Fourier models. Both approaches concur in requiring a very small number of filters (typically four), even if a good orientation accuracy is expected (two degrees).

Success rates of about 90 percent in three or fewer attempts have been experimentally obtained on a system which includes a small minicomputer, a 128 × 128 pixel solid-state camera, a prototype Cartesian robot, and a "universal" parallel-jaw hand.

Index Terms—Artificial intelligence, bin picking, robot control, robot vision, scene analysis.

I. Introduction

TODAY, fully automated assembly lines are working in many industrial systems. In most cases, however, the assembly line requires components to be initially palletized or stacked in some kind of magazine. Off-line, magazines are usually loaded by workers. This paper addresses the problem of feeding a buffer, an assembly line, a machine, or, generally speaking, a goal site from bins that store randomly oriented workpieces.

The problem of feeding parts from bins is usually considered very difficult to solve because the assumption is made that a workpiece needs to be identified and located precisely in space before being grasped. This paper explores the inadequacies of that assumption, and, at the same time, discusses the bin picking problem in detail.

Several methods to acquire randomly stored workpieces are examined, and their respective advantages and limitations described in Section II. The solution introduced here, the matched filter one, is then thoroughly delineated. In Section III, matched filters are defined in general terms. Section IV considers various properties of the bin picking problem which allow a reduction of the computational load associated with matched filters. Section V introduces the design of matched filters for workpiece acquisition, and an example is extensively developed in Section VI. Finally, some extensions of the presented methods are suggested in the Conclusion, Section VII.

II. Current Methods for Bin Picking

Traditionally, workpieces have been fed to machines from bins by specialized equipment or human labor. There, however, is a large class of situations where a robot should bring cost-effective solutions. Our study addresses the latter technique.

In the simplest case, robots without vision can pick workpieces from bins. "Blindly," the end-effector mechanically scans a bin until an object is acquired. Magnetic [1], one-fingered [2], two-fingered [3], and vacuum cup [4] hands have been used for such a purpose. While simplicity is an obvious advantage of these techniques, they become increasingly inadequate as the probability of meeting potential holdsites along the (blind) end-effector path decreases. This problem is exacerbated by the realtively long time constants associated with mechanical motions.

Proximity sensors may increase the active range of the hand, thereby improving the probability of detecting holdsites, but they do not remove the fundamental speed limit imposed by arm motions. By contrast, (remote) vision can drastically reduce the impact of that limit, replacing most mechanical motions by electronic scanning.

The tremendous bandwidth of existing vision sensors (tube sensors may deliver information at a rate in excess of 10^7 bits/s) contributes to the possibility of recognizing and locating parts randomly piled in bins. Limitations, however, are still associated with vision data processing. While not intractable, the problem of recognizing and locating parts in bins requires a lot of computation, and existing computers do not provide a practical solution.

Machine vision allows workpiece pose estimation with algorithms of various complexity and success, according to application [5]. Isolated workpieces that rest on a planar background can generally be represented by binary images. In most cases,

Manuscript recieved July 25, 1983; revised March 26, 1984.

J.-D. Dessimoz was with the Department of Electrical Engineering, University of Rhode Island, Kingston, RI. He is now with the Ecole d'Ingenieurs de l'Etat de Vaud, Yverdon-les-Bains, Switzerland.

J. R. Birk was with the Department of Electrical Engineering, University of Rhode Island, Kingston, RI. He is now with Hewlett-Packard, Palo Alto, CA.

R. B. Kelley is with the Department of Electrical Engineering, University of Rhode Island, Kingston, RI.

H. A. S. Martins was with the Department of Electrical Engineering, University of Rhode Island, Kingston, RI. He is now with the Instituto Superior Tecnico, Lisbon, Portugal.

0162-8828/84/1100-0686$01.00

these workpieces have a small number of stable states. It is relatively easy to identify the silhouette of each state and to estimate accurately the corresponding pose in the background plane (two translations and one rotation).

When the number of states is infinite (four or five continuous degrees of freedom), the problem becomes much more difficult to handle. Moreover, if parts overlap, scene analysis becomes less reliable (hidden components cannot guide analysis). Under restricted conditions, as when workpieces are relatively flat, it is possible to attack the three degrees of freedom problem with partial occlusion [6], [7]. In general, however, overlapping leads to uncertainties in a six-dimensional space (three translations and three rotations). This is a typical situation when parts are stored in bins. As of today, no viable solutions for *pose estimation* in such a context have been discovered.

A system approach to the bin picking problem suggests the exploitation of the gripper-related knowledge in order to decrease the burden of scene analysis otherwise left to vision. In order to improve the performances of techniques mentioned above, the essential requirement is to replace the mechanical motions of the gripper searching for accessible parts by electronic scanning of the sensor. Our goal is then modest—to detect potential holdsites in images—and leaves part recognition and pose estimation for a latter stage, if required at all.

In a first stage, sensor data indicate a potential holdsite; that is, a place in the bin where the robot hand is likely to grasp *something* (yet unknown). Then the workpiece can usually be grasped (this should be acknowledged by simple sensors in the hand), isolated, brought against any suitable background, and even constrained in degrees of freedom (e.g., dropped on a planar surface).

Several ad hoc methods have proved effective for removing various classes of workpieces from bins with a robot (e.g., [8]). Their performances (processing time, accuracy, cost) are compatible with industrial constraints. They have set milestones on the way towards a general bin picking station. This paper will now explore a new solution, perhaps more versatile, and present an example in some detail. This solution relies heavily on matched filters.

III. Matched Filter Definition

Matched filters have been well known for many years in signal processing. However, template matching is more popular processing. Matched filtering differs basically from template matching (e.g., [9]) in taking noise components into account (e.g., [10]). Basically, image analysis is inhibited in frequency bands where noise power is large.

Several key ideas must be added to the basic matched filter concept in order to reach industrial applicability. First, only *local* properties which are noise-, workpiece-, and hand-dependent need to be matched. Second, a large number of filters corresponding to different workpiece pitch, yaw, and roll values can often be replaced by a much smaller set of orthogonal filters ("eigenfilters"). Third, provided that normalization is not required, the high frequency band which will eventually be rejected by matched filters can be cut out in a very early stage of the processing, allowing a low-rate resampling and thereby reducing the computational load. Actually, this can be achieved easily by adequately defocusing the sensor, and then sampling at coarser resolution.

The classical definition of matched filters can be explained as follows. Consider a digital image $f(i, j)$ and a pattern to be detected $p(i, j)$ defined on a domain D. Conceptually, a simple solution to detecting the pattern would call for a filter $h(i, j)$. When applied to f, $h(i, j)$ would give peaks in the output image $g(i, j)$ at locations where the pattern is present. The convolution product, typical of a filtering operation is the following:

$$g(i,j) = \sum_{m,n \text{ in } D} h(m,n) \cdot f(i-m, j-n). \tag{1}$$

Differences between a picture f and a pattern p can be locally modelled in various ways. The most common and theoretically appealing representation relies on Euclidian distance:

$$E = \sqrt{\sum_{m,n \text{ in } D} [(f(i+m, j+n) - p(m,n)]^2}\ . \tag{2}$$

Therefore,

$$E^2 = \sum_{m,n \text{ in } D} f^2(i+m, j+n) + p^2(m,n) - 2f(i+m, j+n) \cdot p(m,n). \tag{3}$$

The first term represents the energy of the input image in a domain D. It generally varies with the position of D in the image, but does not depend on the pattern. The second term represents the energy of the pattern and is constant. The third term is really interesting, because that is where a similarity between pattern and image may actually appear. It is defined as the cross correlation between p and f.

Equations (1) and (3) lead to the well-known result that the filter should basically perform the correlation between the pattern to be detected and the analyzed image:

$$h(i,j) = p(-i, -j). \tag{4}$$

A convolution in the space domain corresponds to a product in the Fourier domain. Changing the sign of the independent variable leads to the complex conjugate of the transform; therefore, we have the following equations:

$$G(u, v) = H(u, v) \cdot F(u, v) \tag{5}$$

$$G(u, v) = P^*(u, v) \cdot F(u, v) \tag{6}$$

where F, G, H, P, are the Fourier transform of g, f, h, p, respectively. $P^*(u, v)$ is the conjugate of $P(u, v)$ and corresponds to $p(-i, -j)$.

As introduced so far, the matched filter is rather simple. It is important, however, to see how it is defined when noise is present in the system. In this case, the following definition applies:

$$H(u, v) = k \frac{P^*(u, v)}{W(u, v)} \tag{7}$$

where W is the Fourier transform of the noise and k is a constant (to be discussed later). Note in (7) that the frequency bands where the noise is large tend to be attenuated. On the other hand, if the noise is white, the matched filter is essentially the same as if there were no noise at all (6). In industrial situa-

tions where bin picking must be performed, low (spatial) frequency variation of the illumination and workpiece reflectance variations are the main noise sources.

Template matching can be reduced to cross correlation in (2) if the energy of the image is relatively uniform everywhere. If not, the first term in (2) must be taken into account, and this leads, in particular, to the following normalized cross correlation:

$$g(i,j) = \frac{\sum_{m,n \text{ in } D} p(m,n) \cdot f(i+m, j+n)}{\sum_{m,n \text{ in } D} f^2(i+m, j+n)}. \quad (8)$$

Operating on images, the normalized cross correlation detects patterns which have the same shape as the template, even though they possibly have a different amplitude. The constant k in (7) can serve the same purpose. Although it appears as a constant in (7), k usually varies with the space coordinates of the picture when normalization is performed. To make this explicit, we can write

$$k(i,j) = \frac{1}{\sum_{m,n \text{ in } D} f^2(i-m, j-n)}. \quad (9)$$

Normalized matching is sometimes necessary, although in practice, it is often ignored in order to save computation time.

IV. Computation Reduction

Applied naively in order to detect each workpiece in the picture of a bin, the matched filter is very time consuming. Consider a workpiece represented by $M \times M$ picture elements (pixels). For each value of pitch, yaw, and roll, a different filter is defined. The angular resolution may be considered to increase linearly with M (for example, the resolution for pitch is best along the window perimeter). This leads to about M^3 filters, each requiring approximately M^2N^2 operations (multiplication and addition) per picture of size $N \times N$.

The computation load can be expressed as follows:

$$L = cM^5N^2 \quad (10)$$

where c is constant.

The computation load can be reduced when workpieces are matched locally, which affects M and its exponent, and when the bandwidth of the filter output is considered, which affects both M and N.

A. Matching Local Structures

Consider our specific problem, bin picking. Generally, the pattern to be matched may be drastically reduced, if the goal of detecting the workpiece type and pose is postponed, and the more practical subgoal of picking "something" out of the bin is selected. Vision algorithms must then detect potential holdsites, i.e., places where a robot end-effector is likely to grasp a workpiece. Potential holdsites appear in images as local structures that reflect both workpiece and end-effector characteristics. Indeed, even for the same object, holdsites are usually located differently when different types of end effectors are used (e.g., vacuum cup versus parallel-jaw hand).

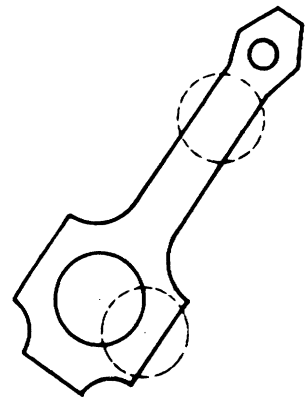


Fig. 1. Dashed circles enclose two potential holdsites. In each case, the segment of interest is much smaller and presents more symmetries than the workpiece as a whole.

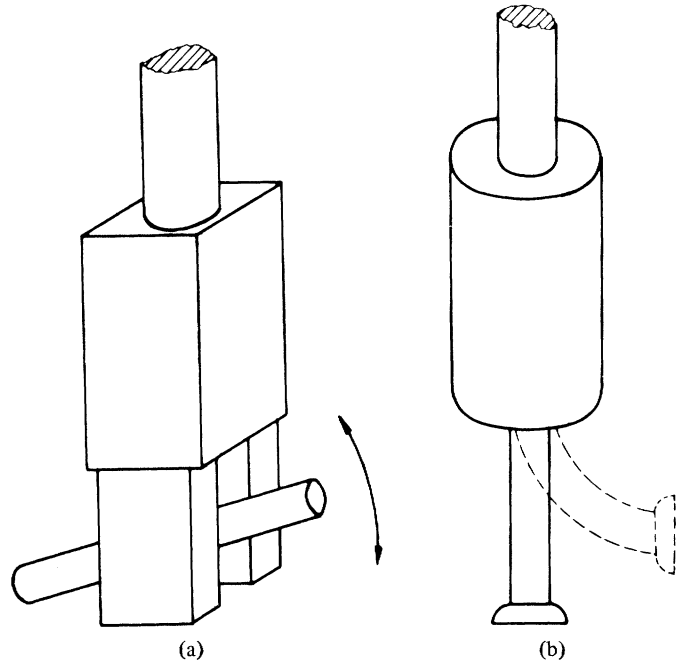


Fig. 2. Workpiece pose need not be completely known for a good grasping. The gripper in (a) allows a random orientation in one plane, while the vacuum cap in (b) can accommodate rotation about two axes.

For rigid and semirigid workpieces, holdsites can be small in respect to overall object size. Therefore, M decreases, and the pattern to detect leads to much less computation (Fig. 1). Notice in (10) that the computation load increases with the fifth power of the pattern size.

Additional benefits appear when the robot hand is taken into account. First, the end-effector may be applied without requiring knowledge of all workpiece translations and rotations. For example, assume that a parallel-jaw hand is used in order to pick an object [Fig. 2(a)]. A rotation of the object in the symmetry plane between the jaws usually does not affect grasping. Similarly, if a compliant vacuum cup is used, small rotations about two axes can be ignored [Fig. 2(b)].

Another benefit derives from the property that, considered locally, workpiece components are much more likely to present symmetries than when viewed as a whole (see Fig. 1). For instance, it would probably be inacceptable to consider an entire

connecting rod as cylindrical. However, the approximation is much more valid for the central shaft and any segment of it than for the whole object. The larger ring of a connecting rod has a similar circular symmetry about another axis. Both properties affect the dimensionality of the computational load, reducing it to a dependence in M^4, M^3, or even less for some workpieces. Notice also that images are essentially orthographic in our applications. Distance between camera and bin is large with respect to bin depth. Thus, distorsions due to perspective viewing can be neglected.

B. Slow-Rate Resampling or "Eigenfilters"

Intuitively, it seems that the matched filter technique requires a large number of filters in order to detect the actual orientation of a holdsite. However, both the traditional signal processing theory and vector space considerations point here to a reasonably small number of filters.

The point spread functions of matched filters for bin picking tend to be defined on domains much larger than one pixel (e.g., [11]). When viewed in the Fourier domain, matched filters often appear as relatively narrow passbands.

By examining this matched filter bandwidth, the minimum sampling rate for images can be estimated and the corresponding de-aliasing can be simply achieved by defocusing the camera lens. Essentially, this means that images can often be coarsely sampled before matched filtering. If a higher resolution were provided, the additional information would be rejected by the matched filter, leading to the same result at a higher computational cost.

In the previous section, matched filtering has been defined for one and two dimensions. This is not adequate when signals are of dimension three or more. Translations in the image plane can be expressed in a two-dimensional space; the most common third dimension corresponds to the rotation in the image plane of local structures to detect. The same analysis as above should be conducted in a polar coordinate system in order to identify the minimum angular sampling rate.

A practical way of assessing the angular bandwidth of a matched filter consists in monitoring its output when applied to a white (pseudo) noise image. This directly defines the minimum number of samples required, i.e., the minimum number of filters (each at a different orientation) that should be applied to images in order to detect local structures of a given type. The example of Section VI typically calls for four filters, if a signal to aliasing noise ratio better than 10 is chosen.

The model used above is the classic Fourier, which basically expresses a signal as a sum of sinewaves. An alternative is possible where a space is generated by vectors corresponding to a set of filters which differ by a certain orientation angle. For example, a filter corresponding to a parallel-jaw gripper is copied 360 times with one degree of rotation in order that the resulting set of filters covers all the hand rotation possibilities. These filters, and thus their corresponding vectors, are usually *not* independent, and when the space is orthogonalized, the actual space dimension, and thereby the minimum number of (eigen) filters, may be found.

In the example discussed in Section VI, four (eigen) filters prove sufficient for a two degree angular resolution, and two of them already yield a practically useful similarity measure for holdsite detection. In the former case, the structure orientation angle is related to the phase of the processed image, and in the latter case, the similarity is estimated by its magnitude.

V. Design of Matched Filters for Workpiece Acquisition

Workpieces and robot end-effectors must be known in geometric terms so that a potential holdsite can be detected. For instance, a parallel-jaw gripper would securely hold opposing parallel edges or opposed concave corners ("⟩⟨" pattern), (e.g., [12]).

During part of the training phase, the geometric representation of potential holdsites must be transformed into light intensity maps similar to the one perceived by vision sensors. This mapping operation is often done in graphics (e.g. [13]), but is fully deterministic in that context. In particular, on every visible workpiece surface element, there are fixed amounts of incident light and effective shadowing. In our context, a slightly different formulation is necessary, essentially in order to take into account the stochastic nature of shadowing.

Image sensors perceive the light intensity reflected by scenes. The light intensity is related to geometric properties by two laws:

$$\text{Lambert's Law} \qquad I_r = I_i \cdot c \cdot \cos^2 a \tag{11}$$

$$\bar{I}_r = I_{io} \cdot (1 - e^{-3d/c_w}). \tag{12}$$

The first law is deterministic and states that the reflected light intensity I_r from a workpiece surface element varies with the incident light I_i on that element, with the angle a of the local surface normal in respect to the sensor axis, and with the surface reflectance constant c. This law is valid for diffuse reflection. When reflection is specular (mirror-like), the amount of reflected light decreases much faster than $\cos^2$.

The second law addresses a stochastic phenomenon and states that the average amount of light incident on a surface element I_i decreases as a function of the depth d of a surface element below the top layer of a bin of parts. We propose it as a result of empirical considerations.

Assuming diffuse illumination above the bin, the amount of light falling on a surface element decreases with depth, because of the partial shadowing of higher workpiece layers. Indirect illumination is exponentially attenuated by losses in consecutive reflections. In the model of (12), the constant c_w corresponds to a shadow and reflectance factor and is workpiece dependent. It can be physically interpreted as the depth of the deepest part below top layer, which can be observed in a typical bin of parts. (I_i on that deepest visible part amounts to less than 5 percent of I_{io}.)

In principle, the signal to detect P is now identified. Properties of the noise W should be evaluated. W is defined on the same domain as P. Most noticeably, noise includes slow (spatial) variations in illumination both on top of the bin, where it is not perfectly uniform, and as a function of depth, when various layers are processed. Therefore, the very low (spatial) frequency component should be rejected.

There is another noise component which is scene dependent.

This noise component appears as potential holdsites (in the sense of a good similarity with P), which correspond to sites where the robot hand *cannot* acquire the object. (An attempt can be physically made, or training can be assisted by an operator.) Perhaps the most adequate method to extract it relies on an iterative technique, applied during a training phase. [See (14)-(17).]

Mathematically the technique can be described as follows, for one matched filter:

$$H^{(n)} = (1 - k)\, H^{(n-1)} + k\, \frac{P}{1 + \delta(o,o) + W^{(n)}} \tag{13}$$

$$W^{(n)} = \begin{cases} W^{(n-1)} & \text{if no potential holdsite is detected or if a potential holdsite is detected and proven successful} \\ (1 - k)\, W^{(n-1)} + k\, B^{(n)} & \text{if a potential holdsite is detected and proven impractical} \end{cases} \tag{14}$$

$$W^{(o)} = 0; H^{(o)} = \frac{P}{1 + \delta(o,o)} \tag{15}$$

$$B = F[b^{(i,j)}] \tag{16}$$

and a potential holdsite is detected if

$$y^{(n-1)} \geq E. \tag{17}$$

The subscript (n) or $(n - 1)$ indicates the iteration number. $\delta(o,o)$ is a Dirac function, corresponding to the average illumination; y is the output of the filter when an image is processed. Typically, the image represents a bin full of workpieces, and E is a threshold above which the image is found locally very similar to the pattern to detect. In (16), $b^{(i,j)}$ is defined as the image in the domain D centered on a potential holdsite as defined by (17). $F[x]$ denotes the Fourier transform of x.

VI. Parallel-Jaw Filter: An Experiment

The matched filter presented in this section can be said to be applicable to those parts that can be picked by a parallel-jaw gripper (see Fig. 3). Experiments have been performed which illustrate the use of the matched filter technique in a practical context.

A. Filter Definition

The pattern to be matched, in order to allow a parallel-jaw gripper to pick up a part, can be roughly described as two dark regions surrounding a bright region [see (11) and (12)]. This pattern can be seen in Fig. 4. The bright region represents the part; the dark regions represent places where the gripper can be placed, i.e., where the part does not touch adjacent pieces.

The dimensions of the window are both gripper and piece dependent; the size of the dark regions has to be large enough to "hold" the jaws of the gripper; the size of the bright region should match a typical structure of the part. The filter was computed as the differences in the average pixel intensity in the light region and in the dark region, each pixel with a weighting factor of one. The average weight of the whole filter is zero, thereby rejecting illumination variation.

Presented like this, the filter exhibits dimensionality reduction by discarding two of the three angles that specify the pose

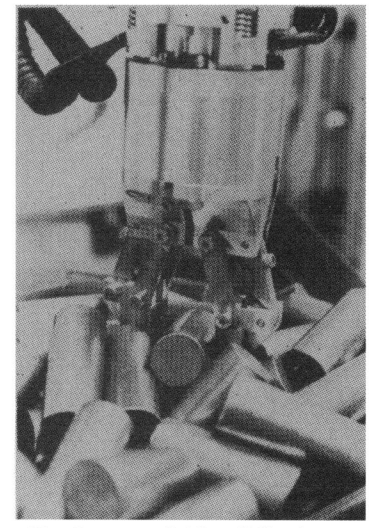


Fig. 3. A parallel-jaw (PJ) gripper.

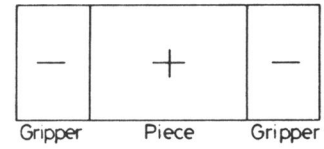


Fig. 4. Parallel-jaw (PJ) filter. Correspondence between geometrical and reflectance properties is illustrated here. The "+" area corresponds to a bright zone where a workpiece is visible, while the "–" area corresponds to deep locations in the bin, where the gripper jaws should not encounter obstructions.

of the part, specifically those that correspond to rotations within the jaws, and by using a much smaller pattern than the one that would be necessary to match a full piece. One problem remains, though, which is the rotation in the image plane, and which, if not taken into account, will make the filter unpractical for most purposes.

To overcome this problem the first thing to be noticed is the fact that the jaws are assumed symmetric, thus the angular domain to be sampled is reduced to a range of 180 degrees.

Second, the gripper could accommodate a certain margin of error for direction estimation within the plane, say 9°. This would require an 18° sampling interval, leading to a total of 10 filters, which is still quite large.

1) Rotational Sampling: Experiments have been performed where the parallel-jaw (PJ) filter described above is applied to a (pseudo) white-noise image. The bandwidth of the output was observed. The results are discussed here which support the considerations of Section IV-B.

As the image consists of white noise, the output of the filter is stochastic. The experiment was repeated many times in order to estimate the average power specturm.

The test image was divided into 36 nonoverlapping regions. In each region the PJ filter was applied 180 times, at a 1° rotation interval. The Fourier transform of the output was evaluated, leading to a power spectrum.

Referring to Fig. 5, using an ideal white noise image as the input, the output will be exactly the transfer function of the matched filter. Hence, by analyzing the Fourier transform of the output, we can get a very good idea about the cutoff frequency of a particular filter. Then, according to the Shannon

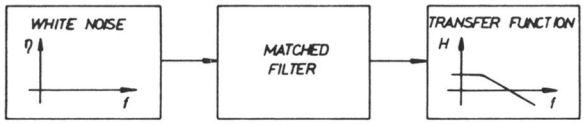


Fig. 5. Estimation of matched filter transfer function through the use of white noise.

TABLE I
COLLISION FRONT SUCCESS RATE. TWO HUNDRED CASTINGS WERE REMOVED. BINS WERE LOADED TEN TIMES WITH 20 CASTINGS.

Filter Dimension (W, Lm, Lp)	BW90 (cutoff freq.)	Power ratio (%)	DC component (magnitude)
(9.4.9)	1	92.9	868.67
(7.4.9)	2	96.8	1304.50
(5.4.9)	2	98.5	1615.56
(3.4.9)	2	95.9	1864.83
(9.3.7)	1	90.1	1362.03
(7.3.7)	1	94.1	1932.03
(5.3.7)	1	91.9	2281.69
(3.3.7)	2	96.7	2551.58
(9.2.5)	2	97.5	1382.42
(7.2.5)	2	98.4	1685.67
(5.2.5)	2	98.1	1553.97
(3.2.5)	2	93.6	1510.25
(9.5.5)	1	95.0	751.42
(7.5.9)	2	98.0	1185.47
(5.5.9)	2	98.3	1484.50
(3.5.9)	2	93.6	1730.42
(9.4.7)	1	90.2	1299.50
(7.4.7)	1	95.3	1778.64
(5.4.7)	1	92.1	2034.17
(3.4.7)	2	96.8	2258.31
(9.3.5)	2	98.1	1538.92
(7.3.5)	1	97.0	1995.28
(5.3.5)	1	97.8	2025.13
(3.5.5)	2	98.2	2056.11

theorem, the maximimum number of samples (i.e., filters at different orientation) needed can be determined.

Practically, the procedures are the following: create a pseudo-white noise image. Apply the particular matched filter of interest to the image with 512 rotations in the range $(0, \pi)$; the rotations are made with regular coordinate transformation and bilinear approximation. The range $(\pi, 2\pi)$ is omitted because of symmetry. Then take the fast Fourier transform (power spectrum) of the output. To avoid the bias due to nonideal whiteness of the image, we repeat the above steps at 36 independent (nonoverlapping area) points, and take their average as a good estimate of the ideal power spectrum. Then set 90 percent of the total power as the criterion for bandwidth (BW). We obtain the results shown in Table I.

Some typical results in linear-log scale are shown in Fig. 6(a)–(d). They correspond to filters having dimensions (7,4,9), (7, 5, 9), (9, 3, 7), and (9, 4, 7), where the first value is the width W, the second is the length of the jaw region L_m, and the last, L_p, is the length of the center area in Fig. 4. As expected, those filters are low pass and very narrowly band-limited. Another interesting point is that, in general, the filters with square central region appear to have the most narrow passbands among the PJ filters.

Notice that for certain PJ filter sites the bandwidth varies, but is limited for practical values to the range 1–2 cycles per 360°. Intuitively speaking, the filter output swings twice between low and high values as the filter makes a 360° rotation

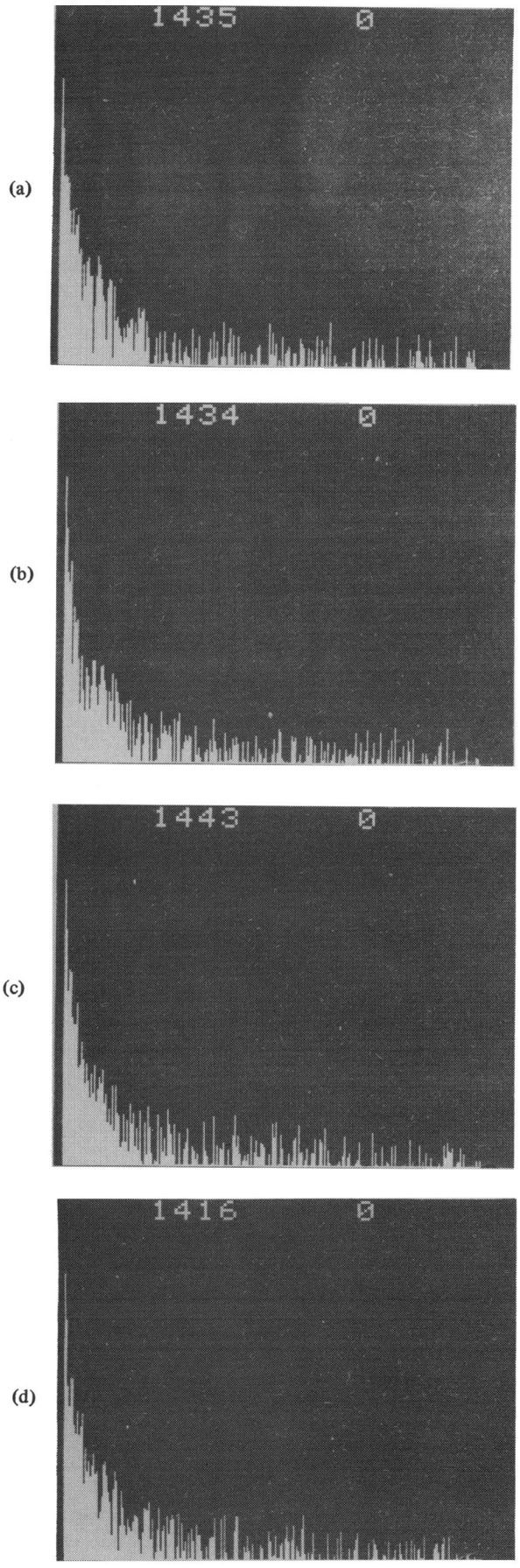


Fig. 6. Power spectrum at the output of various (width, L_m,L_p) filters: (7, 4, 9) filter in (a), (7, 5, 9) filter in (b), (9, 3, 7) filter in (c), and (9, 4, 7) filter in (d). All spectra have been estimated in 36 different nonoverlapping areas of a white noise image and averaged.

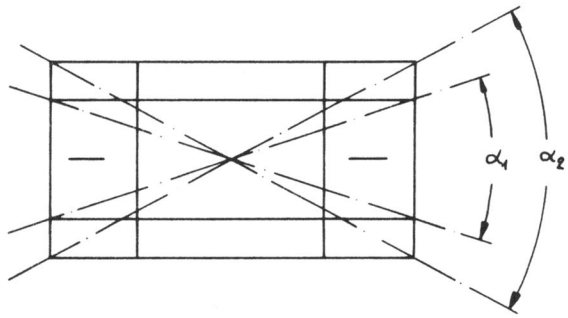


Fig. 7. Angular ranges corresponding to various widths.

on an image region. The size parameters of the PJ filter for our application where connecting rods and yokes are acquired lead to a bandwidth of two cycles per revolution. The Nyquist theorem therefore calls for no more than four angular samples, i.e., four filters at regular angle interval.

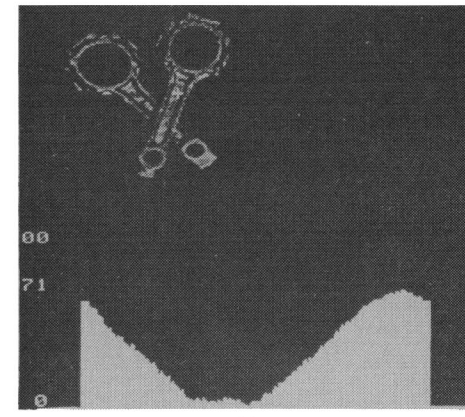
Once the filter is defined, according to the gripper and the workpiece being considered, its angular spectrum can be analyzed and the number of necessary filters is easily obtained. The approach described above is fully automatic while the ad hoc solution given in the sequel requires human operator ingenuity.

2) Using "Eigenfilters": The previous approach defines the minimum number of filters by considering aliasing and signal to noise ratio. It would be preferable to have it directly related to holdsite detection or orientation estimation error. The latter strategy is adopted here, as the number of basic filters is increased until experimental data match the desired performances.

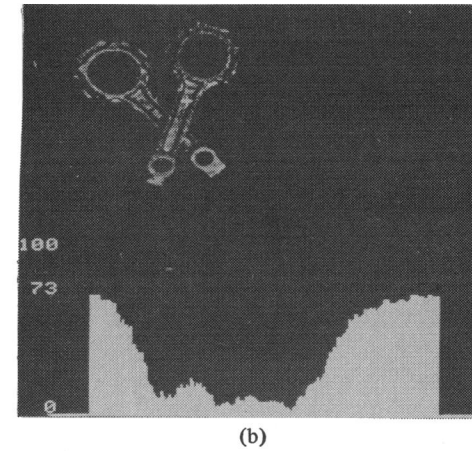
The ideal situation would be the achievement of an "eigenfilter" situation where the use of a very small number (e.g., two) of orthogonal (or at least independent) filters, would yield acceptable performances. Depending on the spanning angle of the filter, i.e., the angular sector that it occupies, more or less overlapping will be obtained between successive rotated filters (see Fig. 7). Even filters that are 45° out of phase would contain some common information about the pattern to be matched. The spanning angle corresponds here to an angular averaging, which in effect is a low-pass (angular) filtering.

A test was designed where filters of varying widths were passed over an operator-selected holdsite on a bin of connecting rods, for half a rotation. Two examples are shown in Fig. 8(a) and (b). The response curve suggested that the larger the width, the lower the resolution of the filter, as demonstrated by a flatter peak response of smaller amplitude.

The test also suggested that for a sufficiently wide filter, a cosine square dependency was being obtained with angle. (Applying the Nyquist theorem, we would then need only *two* filters, 90° apart, the output to other filters being obtained by interpolation.) With one such filter, another test was performed, where a synthetic image was created consisting of a perfect holdsite, corresponding to the filter in dimensions. The filter was applied twice, with a 90° interval. This is equivalent to applying two filters which are 90° out of phase. We shall refer to them as the vertical or the horizontal filter. Assuming a sine square,



(a)



(b)

Fig. 8. PJ filter response versus angle for various widths. The filter is applied at the point addressed by a cursor. In (a), width = 9, length minus = 4, length plus = 11. In (b), width = 3, length minus = 4, length plus = 11. Refer to Fig. 4 for the definition of variables.

cosine square variation for the filters' responses,

$$H = M \cdot \sin^2 a$$
$$V = M \cdot \cos^2 a. \tag{18}$$

The peak magnitude and angle were estimated as

$$M = H + V$$
$$a = +/- \tan^{-1} (H/V)^{1/2}. \tag{19}$$

The sign ambiguity in the angle determination was eliminated with the use of another pair of filters 45° out of phase, i.e., at 45 and 135°.

The errors on magnitude and angle determination can be seen in Fig. 9. Estimation can be considered good over the entire range. For the angle determination they are even lower if we combine the most accurate range of each filter pair. The resulting maximum error in orientation estimation is about 2° for the entire range. In the magnitude estimation there is also some gain, but not as large. The new error plot can be seen in Fig. 10.

This result suggested the following procedure to apply the filter. First use only two filters, the horizontal and vertical for simplicity of computation; with the output images estimate the magnitude of the response; find the maxima in the response

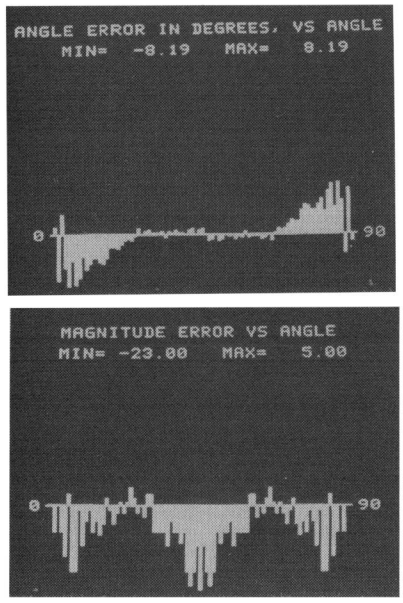


Fig. 9. Error plots for test 1 (2 filter estimation).

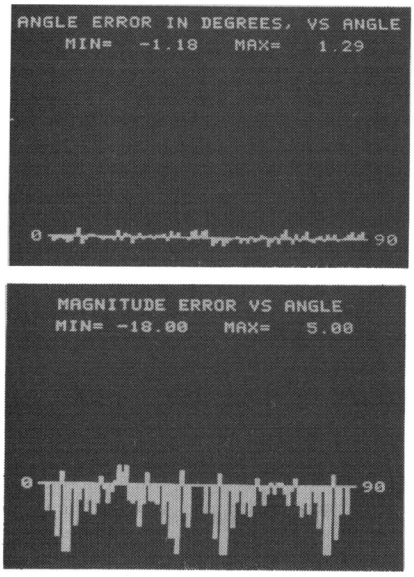


Fig. 10. Error plots for test 2 (4 filter estimation).

and select a fixed number of holdsites; for those, and only those, apply the filter at 45 and 135°; use one pair to estimate the direction, the other to resolve symmetry ambiguity. The pair to be used for direction estimation is the one that has more balanced response, i.e., the one for which the ratio of the larger response to the smaller is closer to unity. The rationale for this choice can be understood by analyzing the error plots just presented, as within the best estimating range of a filter pair, both filters yield closer responses than outside it.

B. Detection of Potential Holdsites

Matched filtering is only one component of the overall algorithm which allows a robot with vision to acquire workpieces from bins.

The breakdown of the different phases of the parallel-jaw filter algorithm can be seen in Fig. 11. A brief description of

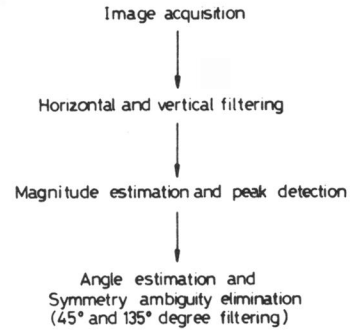


Fig. 11. Main steps for potential holdsite detection.

each of the phases follows, with explanation of the several parameters involved in each. This algorithm has been programmed in a mixture of assembler and Fortran to run on a minicomputer. The procedures that form the core of the algorithm have then been translated in Pascal for documentation and transportability purposes.

1) Image Acquisition: A gray scale image of the bin was acquired and stored in memory as a two dimensional array G,

$$0 \leqslant G(i,j) \leqslant \text{MAX}; \quad 1 \leqslant i,\ j \leqslant N \tag{20}$$

where MAX is the maximum intensity level and N the size of the image. For notation convenience assume that this is already the reduced resolution image desired.

The two filtered images are obtained using a recursive computation. Only the vertical filtering computation is described as the procedure was similar for the horizontal filter. The filtering is only computed where the window fits completely within the image.

Assume a window of width W, and length of the dark or minus regions, equal to L_d and of the bright or plus region, L_p. Assume also that both W and L_p are odd integers.

A buffer B is created containing the sum of W pixels along the image columns. The recursive computation of B is

$$\begin{aligned} B_i(j) &= B_{i-1}(j) - G(i-w,j) + G(i-1+w,j) \\ & \qquad j = 1, \cdots, N \\ & \qquad i = w+1, \cdots, N+1-w \\ & \qquad w = (W-1)/2 \end{aligned} \tag{21}$$

with starting conditions

$$B_w(j) = \sum_{i=1}^{W} G_0(i,j); \quad j = 1, \cdots, N. \tag{22}$$

For each occurrence of B, two buffers, M and P, are recursively computed, holding the sums of the pixels in the minus and plus zones of the filter as

$$\begin{aligned} M_i(L+j) = M_i(L+j-1) &- B_i(j) \\ &- B_i(j+L_p+L_m) \\ &+ B_i(j+L_m) \\ &+ B_i(j+L_p+2L_m) \end{aligned} \tag{23}$$

$$P_i(L+j) = P_i(L+j-1) - B_i(j+L_m) + B_i(j+L_p+L_m)$$
$$j = 1, \cdots, N+1-2L$$
$$i = w, \cdots, N+1-w$$
$$L = (L_p+1)/2 + L_m \tag{24}$$

with starting conditions

$$M_i(L) = \sum_{k=1}^{L_m} B_i(k) + B_i(k+L_p+L_m) \tag{25}$$

$$P_i(L) = \sum_{k=1}^{L_p} B_i(k+L_m). \tag{26}$$

The vertical image V is obtained from the buffers P and M as

$$V(i,j) = (1/W)\,[(P_i(j)/L_p) - (M_i(j)/(2L_m))] \quad \text{if} > 0 \tag{27}$$
$$= 0, \quad \text{otherwise.} \tag{27}$$

The two filtered images are then summed to estimate the magnitude response. The S largest local maxima are chosen as the selected holdsites if their magnitudes are above a minimum threshold T. In case all peaks are below that threshold the bin is considered empty.

The local maxima are obtained in the following way. An ordered list of peaks is created with length S. When a new magnitude is computed it is compared against the list of entries. If its magnitude is smaller than the smallest entry it is rejected. If the entry is found somewhere within the list the distance to all entries above it is computed. If it differs by less than a minimum distance D from one of those entries it is rejected, otherwise it is inserted in the list; all entries below it that are closer to it by less than D are then removed from the list.

Two filtering windows identical to the ones used for vertical and horizontal filtering, rotated 45° from those ones, are passed over the image in a brute force way, i.e., as a double summation of pixel values, their coordinates being computed using standard plane rotation.

Direction of the holdsite is then computed following the criterion explained above.

Sometimes the algorithms detect false holdsites, i.e., regions in images which do not correspond to sites where workpieces can be grasped. An example of such a situation might be a spurious reflectance pattern in an empty bin. The false holdsites lead to failures because the robot gripper makes a motion but cannot acquire any workpiece. In order to avoid deadlocks, some bookkeeping must be made of past failures. Essentially, subsequent holdsite detection should be inhibited where failures have occurred.

C. Experimental Results

Fig. 12 (a)-(c) shows the result of applying the parallel-jaw filter algorithm to different pieces, the connecting rod and yoke castings, and titanium cylinders. Shown are the input gray scale image, the horizontal and vertical filters output images, and

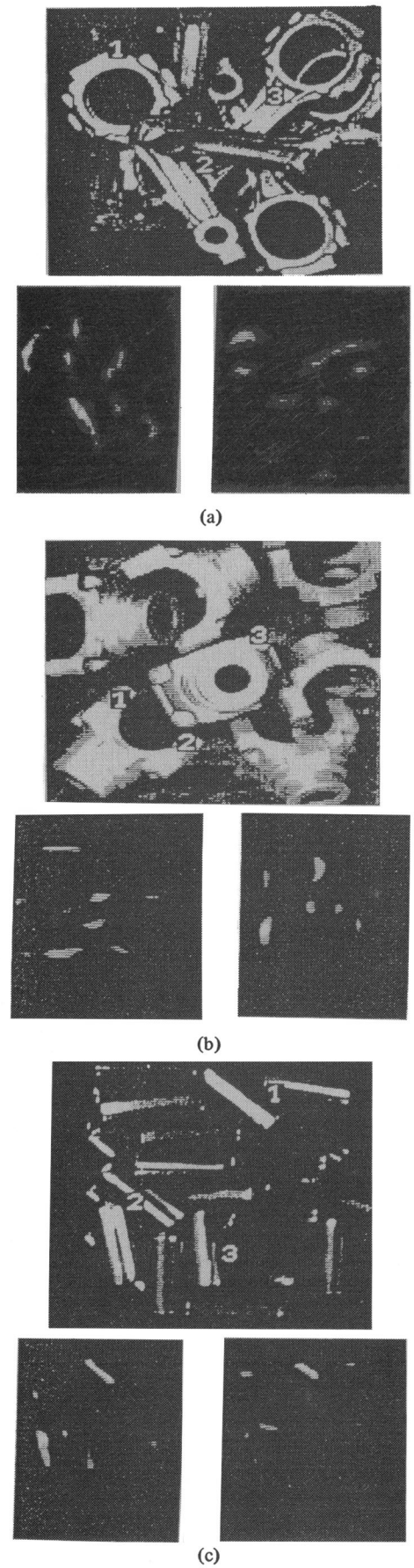


Fig. 12. PJ filter applied to various workpieces: in (a), connecting rod castings; in (b), yoke castings; in (c), large titanium cylinders.

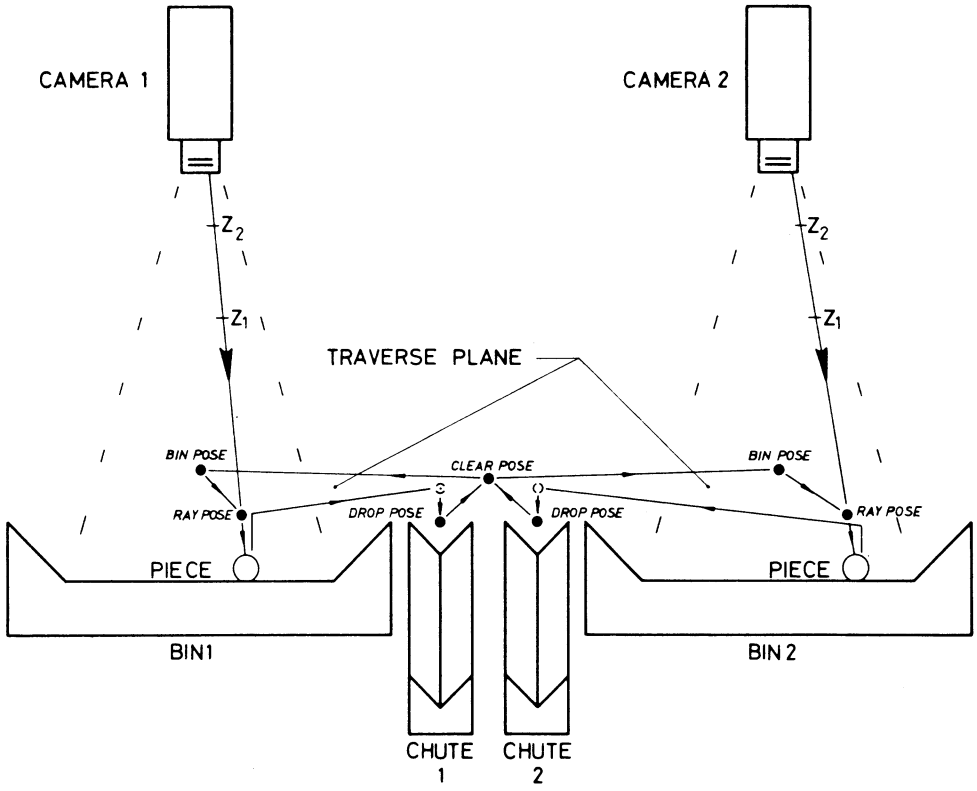


Fig. 13. Schematic of the dual bin execution cycle, with key poses labeled.

the holdsites selected. Notice that the filter output has only half of the resolution of the input images, as the coarser sampling is made possible by the passband effect of the matched filters.

Following potential holdsite detection, robots with vision have to physically acquire workpieces. For our experiments, the system architecture included a robot MARK IV, a Cartesian robot developed at URI, a pair of bins, two overhead cameras (GE TN 2000) looking down in two bins, and two receiving sites for the acquired parts. The supply bins had a flat bottom area the same size of the field of view. To provide collision-free access to all the pieces, the sides of the bin were slanted. The alignment of the two cameras in relation to the robot's coordinate frame was done in a semiautomatic way, generating a camera model which is described in [13]. This model provided relatively free position of the cameras which were placed in an almost vertical orientation above the bins.

The software ran on a minicomputer (Control Automation LSI-2). To speed up execution, acquisition was performed in parallel from two bins. Two tasks (see Fig. 13) were run simultaneously by the real-time executive. One of the tasks performed the image analysis on one of the bins while the other performed the acquisition on the other. The acquisition task had priority over the image analysis task, which was only allowed to take control of the CPU during arm motions.

Cycle Times: If the image analysis task was run alone, an execution time for the parallel-jaw filter algorithm of about 5 s was achieved. Image complexity slightly affects the computation time.

With the system architecture used, paralleling the image analysis with an acquisition attempt, the overall cycle took about 9 s/piece. This time was generally dominated by arm motions in the sense that, even if the image analysis was performed in no time at all, the reduction in cycle time would be negligible.

TABLE II

Removed on first attempt (p1)	57 %
Removed on second attempt (p2)	24 %
Removed on third attempt (p3)	6 %
Not removed in three attempts (new picture taken)	13 %

It is difficult to assess the success rate of vision algorithms for bin picking. Ultimately, it is only by trying a detected holdsite with an actual end-effector that the success or failure can be observed. Thus, the algorithms require a complete system in order to be tested.

System performances will be effected by many factors which are totally unrelated to vision, for example, the size of a vacuum cup. Even when the chance is taken to assess the success rate for a certain algorithm and system, the results are of limited value. What would be far more interesting, if possible, would be to assess the performances for a large class of objects or even, for all existing objects. The practical limitations for such an assessment are obvious and the best achievable rating would be restricted to a "standard workpiece set," yet to be defined.

Some informal and formal statistics were collected observing several runs of the system. They are all based on allowing three holdsite selections per bin at most. The second selection would be considered only in case the first attempt was a failure, and the third in case the second was also a failure. In the case all sites resulted in failures, a new image was taken and analyzed.

A test run was made for the "collision fronts" algorithm and the connecting rods castings. Results are show in Table II

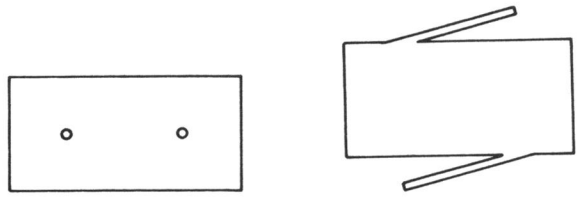


Fig. 14. Nonlinear filters seem necessary, when very small structures make grasping difficult. In (a), small holes must be detected, if a vacuum cup hand is used, to avoid leaks. In (b), narrow fins might break if a parallel-jaw hand is not carefully positioned.

and indicate a probability p, larger than 50 percent for coming out of the bin with one part at first trial and of almost 90 percent not to come out empty. Considering that this piece was once considered the "difficult" problem for bin picking, the results are very good. Notice in Table II that experimental results agree rather well with the statistical predictions that define the probability p_2 of being successful in exactly two trials as $p_1 - p_1^2$, and similarly, p_3 as $p_1 - 2p_1^2 + p_1^3$.

For the parallel-jaw filter algorithm no formal results were obtained as at that point in time the Robot Laboratory was in the process of changing from the MARK IV robot to an industrial robot, a Unimation PUMA 600, from the original PJ gripper to a new model, PJ II, to a more powerful computer, a TI 990/12, and also making an effort to write all programs in Pascal to facilitate technology transfer. For the period of time where the old system was still running, the parallel-jaw filter algorithm seemed to outperform the former collison fronts algorithm for the two workpiece types of our experiments, connecting rod and yoke casting. The numbers presented in Table II can then be thought of as a lower limit on performance of the parallel-jaw filter algorithm.

VII. Conclusion

The paper has introduced a new technique for the acquisition of randomly piled workpieces by a robot with vision. It complements past algorithms especially for situations where scene representation by binary images is not adequate. Although it appears to allow the automatic acquisition of a significant class of workpieces, it is not universal. Topics related to the described technique and where research is presently needed include the following ones:

1) Definition of nonlinear "matched" filters for applications where a small structure would prevent the successful acquisition of workpieces. For example, even a relatively small hole in a workpiece would prevent a vacuum cup from acquiring it, or thin fins along a handle would prevent a parallel-jaw from acquiring an object (see Fig. 14). In both cases, however, a linear filter would probably not perceive these small contributions or they would be dominated by other sources, e.g., workpiece illumination variations.

2) Automatic definition of matched filters, from solid modeling representations of workpiece and robot end-effectors. Consider the selection of potential holdsites of a workpiece by a given robot end-effector during training phase for a new workpiece type. At present, an operator estimates and possibly manually tries locations where the workpiece can be securely held. In the future, geometric modeling should allow the simulation of the same operation, thereby leading to the automatic definition of holdsites.

3) Hierarchical approach where matched filters would not directly detect holdsites but rather extract holdsite components. Filter output could be interpreted at a higher level—or undergo additional iteration.

4) Acquisition of randomly piled workpieces requires complex systems. In particular, components nonrelated to vision have an influence on performances. Therefore, additional research is especially needed in end-effector design, for aspects related to kinematics (e.g., how many fingers), geometry (e.g., what finger shape), and sensory capabilities (e.g., artificial skin).

References

[1] A. Ferloni, I. Franchetti, P. Vincentini, and P. Fici, "Ordinatore: A dedicated robot that orientates objects in a predetermined direction," in *Proc. 10th Int. Symp. Industrial Robots*, Milano, Italy, 1980, pp. 655–658.

[2] A. Romiti, "Picking from a bin through tactile sensing," in *Proc. 11th Int. Symp. Industrial Robots*, Tokyo, Japan, Oct. 7–9, 1981, pp. 273–280.

[3] T. Sakata, "An experimental bin-picking robot system," in *Proc. 3rd Int. Conf. Assembly Automat.*, Stuttgart, Germany, May 25–27, 1982, pp. 615–626.

[4] R. Tella, J. R. Birk, and R. B. Kelley, "General purpose hands for bin picking robots," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-12, Nov.–Dec. 1982.

[5] J.-D. Dessimoz, J. Birk, R. Kelley, and A. Beckwith, "Between bins and goalsites: Workpiece pose estimation," in *Proc. 3rd. Int. Conf. Automat. Assembly*, Stuttgart, Germany, May 25–27, 1982, pp. 603–614.

[6] J.-D. Dessimoz, M. Kunt, G. H. Granlund, and J. M. Zurcher, "Recognition and handling of overlapping parts," in *Proc. 9th Int. Symp. Industrial Robots*. Washington, DC, Mar. 1979, pp. 357–366.

[7] W. Perkins, "A model-based vision system for industrial parts," *IEEE Trans. Comput.*, vol. C-27, pp. 126–143, Feb. 1978.

[8] R. Kelley, J. Birk, J.-D. Dessimoz, H. Martins, and R. Tella, "Acquiring connecting rod castings using a robot with vision and sensors," in *Proc. Int. Conf. Robot Vision Sensory Contr.*, Stratford-upon-Avon, U.K., April 1–3, 1981, pp. 169–178.

[9] R. Duda and P. Hart, *Pattern Classification and Scene Analysis.* New York: Wiley, 1973.

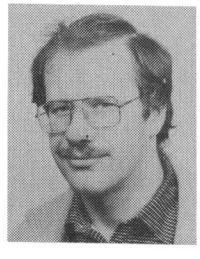
[10] W. K. Pratt, *Digital Image Processing.* New York: Wiley, 1978.

[11] J. Birk, J.-D. Dessimoz, and R. Kelley, "General methods to enable robots with vision to acquire, orient, and transport workpieces," presented at 9th NSF Grantees' Conf. Product. Res. Technol., Ann Arbor, MI, Nov. 2–4, 1981.

[12] S. D. Roth. "Ray casting for modeling solids." *Comput. Graph. Image Processing*, pp. 109–144, 1982.

[13] H. Martins, J. Birk, and R. Kelley, "Camera models based on data from two calibration planes," *Comput. Graph. Image Processing*, vol. 17, pp. 173–180, Oct. 1981.

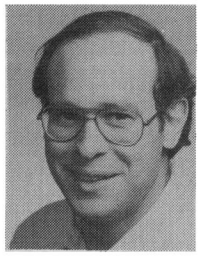
[14] R. Kelley, H. Martins, J. Birk, and J.-D. Dessimoz "Three vision algorithms for acquiring workpieces from bins," *Proc. IEEE*, Special Issue on Robotics, pp. 803–820, July 1983.



Jean-Daniel Dessimoz (S'79–M'80) was born in Conthey, Switzerland, on March 17, 1953. He received the B.A. degree in humanities from the College de Sion, Switzerland, in 1971, and the M.S. and Ph.D. degrees in electrical engineering from the Swiss Federal Institute of Technology, Lausanne, in 1976 and 1980, respectively.

He was with the Department of Electrical Engineering, University of Rhode Island, Kingston, from 1980 to 1982 where he was an Assis-

tant Professor. Since 1982, he has been sharing his time between the Ecole d'Ingenieurs de l'Etat de Vaud, Yverdon-les-Bains, Switzerland, where he is currently an Associate Professor, and consulting. His research interests include machine vision, robotics, digital signal processing, knowledge engineering, and computer architecture.



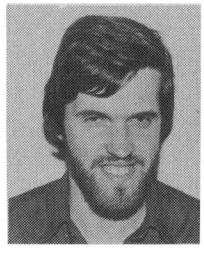
John R. Birk (S'67–M'70) was born in New York, NY, on June 10, 1944. He received the B.E. degree in mechanical engineering from Cooper Union, New York, NY, in 1966, and the M.S. and Ph.D. degrees in biological engineering from the University of Connecticut, Storrs, in 1968 and 1970, respectively.

He was with the Department of Electrical Engineering, University of Rhode Island, Kingston, from 1970 to 1982, where he was a Professor and was the first Director of the Robotics Research Center. Since 1982 he has been with Hewlett-Packard, Palo Alto, CA. His research interests include manipulators, computer control, scene analysis, and industrial automation.

Chi-Lin I, photograph and biography unavailable at the time of publication.

Robert B. Kelley (S'56–M'67) was born in Newark, NJ, on April 24, 1935. He received the B.S.E.E. degree from the New Jersey Institute of Technology, Newark, in 1956, the M.S.E.E. degree from the University of Southern California, Los Angeles, in 1958, and the Ph.D. degree from the University of California at Los Angeles in 1967.

He has been at the University of Rhode Island, Kingston, with the Department of Electrical Engineering since 1966, where he is currently a Professor and Technical Director of the Robotics Research Center. His research interests include image analysis, pattern recognition, artificial intelligence, robotics, and advanced automation.



Henrique A. S. Martins was born in Lisbon, Portugal, on June 22, 1953. He received the electrical engineering degree from the Instituto Superior Tecnico, Lisbon, Portugal, in 1976, and the M.S. and Ph.D. degrees in electrical engineering from the University of Rhode Island, Kingston, in 1980 and 1982, respectively.

He was a Teaching Assistant from 1974 to 1976 and an Instructor from 1976 to 1977 at the Department of Electrical Engineering, Instituto Superior Technico, Lisbon, Portugal. In 1977 he took a leave when he was awarded a Fulbright-Hays travel grant to pursue the M.S. and Ph.D. degrees in the U.S. He was a Research Assistant at the Department of Electrical Engineering, University of Rhode Island, Kingston, from 1977 to 1982. He is currently at the Instituto Superior Tecnico, Lisbon, Portugal.