Vision Robotique Appliquée au Sertissage Automatique de Fils Electriques

Robot Vision in Automated Electrical Wire Crimping

David Vernon

Department of Computer Science, University of Dublin, Trinity College, Dublin, Ireland.

RESUME

Ce rapport décrit la vision robotique appliquée au sertissage automatique de fils électriques. Ce procédé de fabrication consiste à choisir un fil d'un plateau, de saisir un bout de fil avec un manipulateur robotique, et de mettre le bout dans une sertisseuse automatique. La sertisseuse attache la fixation puis le manipulateur remet le fil dans un plateau secondaire. On suppose qu'on ne met qu'un ou deux fils l'un sur l'autre à la fois dans le premier plateau. L'une des techniques discutées ici est la segmentation des image des niveaux de gris au moyen d'operations de seuillage et ce rapport décrit un algorithme automatique pour choisir un seuil. Cet algorithme est fondé sur la théorie (de Marr et Hildreth) de detection des contours. Ce rapport décrit un algorithme qui amincit les images binaires en suite, et qui produit un modèle squelettique, un pel de largeur, des fils. On décrit un autre algorithme heuristique qui analyse ces squelettes et qui extrait les coordonnées et l'orientation d'un point applicable où le manipulateur robotique peut saisir le fil. Cet algorithme-ci donne les coordonnées et l'orientation du bout de fil pour le mettre dans la sertisseuse. Toutes les coordonnées et les orientations sont rapportées dans un référentiel du monde réel. On se sert des polynômes de troisième ordre pour modeler la distorsion de caméra et aussi pour modeler la correspondence entre les coordonnées de l'image et les coordonnées du monde réel.

SUMMARY

The use of robot vision in automating an electrical wire crimping process is described. This process comprises selecting a wire from a tray, grasping this wire near one end with a robot manipulator, and placing the end in a crimping press. The press is activated, attaching the crimp, and the wire is then placed in an auxiliary output tray. It is assumed that the wires are layered no more than one or two deep. Techniques involved in this application include grey-scale image segmentation by thresholding. An automatic threshold selection algorithm based on the Marr-Hildreth theory of edge detection is detailed. An algorithm is described which thins subsequent binary images, producing a one-pixel wide skeleton model of the wires to be crimped. A skeleton-based heuristic analysis algorithm is also described which selects and identifies the coordinates and orientation of an appropriate point at which the robot manipulator may grasp the wire. The algorithm also derives the position and orientation of the nearby wire-end to facilitate insertion into the crimping press. All sensed positions and orientations supplied are given in a real-world reference frame; third-order polynomial spatial warping functions are used to model both the camera distortion and image to real-world coordinate reference frame transformation.

Robot Vision in Automated Electrical Wire Crimping

Vision Robotique Appliquée au Sertissage Automatique de Fils Electriques

David Vernon

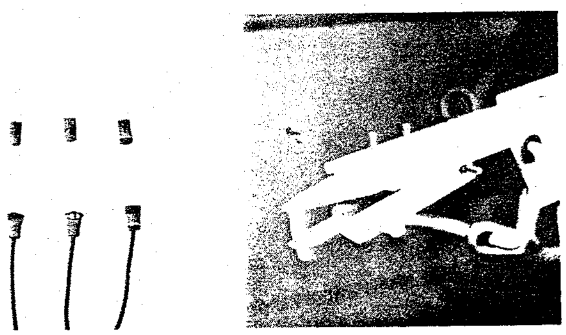
1.0 INTRODUCTION.

The manufacture of many electrical and electronic sub-assemblies involves the use of insulated electrical wiring. Such wiring varies considerably in length and must be properly prepared before being incorporated in the unit being manufactured. This preparation includes cutting the wire to length, stripping the insulation from both ends, and either tinning the ends with solder or attaching crimps (see diagram 1(a)). Presently, the first two of these processes may be effected automatically using high-volume automatic wire cutters and strippers, but the third process is normally accomplished by a human operator manually dipping the wire ends in flux and solder, in the first case, or crimping the wire using a crimping press, in the second.

This paper is concerned with the use of visual sensing to facilitate the automation of this crimping process by a robot manipulator.

Conventional first-generation robotic systems, which do not incorporate visual sensing, require that the work-pieces (in this case, the wire strips) are all uniformly oriented and uniquely presented to the robot arm [SIMO80, p.20]. Accomplishing this with wire strips is not a trivial task as the wire will, in general, adopt an arbitrary curvilinear profile. It would require specialised jigs to present the wire correctly to the end-effector, and such machinery would have to be able to adapt to wire strips of different length and thickness. The use of visual sensing to select and identify the wire strip offers a legitimate alternative solution.

The approach taken in this paper endeavours to constrain the work environment so that the scene presented to the camera is not truly random. This is achieved by stipulating that the wires are arranged no more than one or two layers deep in the tray and the tray constitutes a clearly visible, and contrasting, visual background. In addition, it is assumed that the wires are 'almost flat', that is, their spatial variation in the third dimension is minimal.



(a) (b)

Diagram 1. (a) Various Crimps, both attached and unattached to wires. (b) Robot end-effector.

A special-purpose end-effector has been constructed to simultaneously grip the wire and push it down onto the tray (see diagram 1(b)). This mechanism for grasping the wire, together with the above constraints, allows the assumption that the wire actually lies in the plane of the tray. This obviates the need to explicitly extract the z component of the wire position; only the x- and y-coordinates of the grasp and end points are required. Tsuji et al. have developed a robot vision system which identifies the three-dimensional geometry of electrical wires but their technique assumes the presence of clearly-defined shadows and published results have demonstrated their technique for scenes containing only two wires [TSUJ82].

2.0 METHOD.

The solution to this sensing problem reduces to acquiring a grey-scale image, segmenting the image into two regions (the wires and the background) by thresholding, and explicitly modelling these wires by skeletonising the image. The wire skeletons are subsequently analysed by searching for some wire which might be easily grasped by the servo system and then extracting all salient information which the robot requires to accomplish the grasping and crimping action. A similar approach has been taken in problems concerned with the analysis of paper pulp fibres [KASV78] and of asbestos fibres [DIXO78].

2.1 Segmenting The Image.

Two distinct approaches, boundary detection and region growing, may be taken to the segmentation problem [BALL82, p.116]. A commonly-used and simple region-based technique is grey-level thresholding. In cases where an object is represented by uniform grey-level and rests against a background of different grey-level, thresholding at an appropriate level will assign a value of 0 to all pixels with a grey-level less than the threshold and a value of 1 to all pixels with a grey-level greater than the threshold. This segments the image into two disjoint regions, one corresponding to the background, and the other to the object.

In more detail, a threshold operation may be viewed as a test involving a function T of the form

$$T(i,j,N(i,j),g(i,j))$$

where $g(i,j)$ is the grey-level of the point (i,j) and $N(i,j)$ denotes some local property of the point (i,j), e.g. the average grey-level over some neighbourhood. If $g(i,j) > T(i,j,N(i,j),g(i,j))$ then (i,j) is labelled an object point, otherwise it is labelled as a background point, or conversely [WESZ78, p.259]. This is the most general form of the function T and three classes of thresholding may be distinguished, based on restrictions on T. These are global, local, global thresholding, where $T = T(g(i,j))$, that is of specific interest here.

The problem for reliable segmentation is to select the appropriate threshold. Weszka [WESZ78] provides a good survey of threshold selection techniques. One approach to threshold selection [KATZ65] evaluates the gradient of the image to be segmented and uses the average grey-level of those pixels having high gradient magnitudes as an estimate of the threshold value. Points in an image with high gradient values normally correspond to edges and such edges frequently lie on the boundary between object and background. Typically, the grey-level of this boundary pixel will lie between that of the object and the background and so thresholding at this level is likely to segment the image into object and background. The problem lies in deciding what constitutes a 'high' gradient magnitude and one is again presented with a (new) threshold selection task.

The technique described here is similar to this approach but, instead of using pixels having large gradient values, only those pixels explicitly on the edge/boundary are used, these being found with a reliable edge detector which is not dependent on thresholds. Such a detector is derived from the Marr-Hildreth theory of edge detection [MARR80].

Briefly, the Marr-Hildreth theory of edge detection is based on the Laplacian of an image that has been convolved with a 2-D Gaussian function. The Laplacian is the sum of the second (unmixed) partial derivatives; the second derivative of a point of high spatial frequency (i.e. a point at which the image grey-level changes very sharply) generates high positive and negative values on either side of the intensity change. Isolating these positive-to-negative-, or zero-, crossings effectively identifies points of high intensity change, i.e. edges. The Gaussian is used to smooth the image and Gaussian functions of different standard deviation yield edges derived at different scales within the image. By correlating edges detected at different scales, true or significant edges may be generated. A property of convolution allows the convolution with the Gaussian and the evaluation of the Laplacian to be combined as the convolution with the Laplacian of the Gaussian. This simplification affords significant computational savings.

While Marrs theory requires the correlation of edge segments derived using Gaussians of different standard deviation, empirical studies indicate that the edges detected by one operator alone are reliable. The operator implemented here uses a Gaussian with a standard deviation of two.

The threshold selection procedure first uses a Marr-Hildreth detector to locate edges in the image and the mean grey-level of the image pixels at these edge points is computed. This mean represents the global threshold value.

2.2 Generating Wire Skeletons.

The skeleton of an object may be thought of as a generalized axis of symmetry of that object [LEVI70, p.62]. Serra [SERR82, p.375] attributes the first formalisations of the concept of a skeleton to Motzkin [MOTZ35] and Blum [BLUM62]; indeed, the medial axis transform (MAT) proposed by Blum [BLUM67] is one of the earliest and most widely studied techniques for generating the skeleton.

The skeleton is frequently used as a descriptor of shape [PAVL78, p.246] and one of the important features of the MAT is that the shape boundary may be easily reconstructed from the medial axis. The concept of thinning the binary image of an object is related to such transformations in that it generates a representation of an approximate axis of symmetry of the shape by the successive deletion of pixels from the boundary of the objects. In general, such a thinned representation is not formally related to the original object shape and it is not possible to reconstruct the original boundary from the thinned object. Numerous thinning algorithms exist and a survey is given in [TAMU78].

If one treats a thinning algorithm as one that removes object pixels from an image according to some constraints then it remains to consider what these constraints must be.

First, the pixel must be a border pixel. This implies that it has at least one neighbouring pixel which is a background pixel (an 8-adjacency convention is assumed throughout). Given that the image is being scanned in a raster format, seeking potential pixels for removal, then if those border pixels are removed just as they are found, it is unlikely that the resulting thinned object will represent a skeleton of the object, i.e. that the remaining pixels are equidistant from the original boundary. A remedy for this is to make successive passes of the image, removing pixels with different border orientations on successive passes. For example, such a sequence might be North, South, East, West, North, South, ... etc. [ROSE82, p.232].

The second restriction is that deletion of a pixel should not destroy the objects connectivity. The number of skeletal lines after thinning should be the same as the number of objects in the image before thinning. This problem is related to the manner in which each pixel in the object is connected to every other pixel. Pixels are said to be connected to, and a component of, an object if it has a grey-scale of 0 and at least one adjacent object pixel. Consider now the following five-pixel object:

A		E
B	C	D

The pixel C 'connects' the two object segments AB and ED, that is, if C were removed then it would break the object in two; this pixel is 'critically-connected'.

Obviously, this property may occur in many more cases than this and critical-connectivity may be characterised as follows:

Given a pixel, labelled 9, and its eight adjacent neighbours (labelled 1-8)

1	2	3
8	9	4
7	6	5

and assume that writing the pixel number (e.g. 8) indicates presence, i.e. it is an object pixel, whereas writing with an over-bar (e.g. $\bar{8}$) indicates absence, i.e. it is a background pixel. Assume, also, normal Boolean sign conventions (+ indicates logical OR, and . indicates logical AND). The pixel 9 is critically-connected if

$$
\begin{aligned}
9.\{ & \lfloor(1+2+3).(5+6+7).\bar{4}.\bar{8}\rfloor \\
+ & \lfloor(1+8+7).(3+4+5).\bar{2}.\bar{6}\rfloor \\
+ & \lfloor 1.(3+4+5+6+7).\bar{2}.\bar{8}\rfloor \\
+ & \lfloor 3.(5+6+7+8+1).\bar{2}.\bar{4}\rfloor \\
+ & \lfloor 5.(7+8+1+2+3).\bar{4}.\bar{6}\rfloor \\
+ & \lfloor 7.(1+2+3+4+5).\bar{6}.\bar{8}\rfloor\}
\end{aligned}
$$

is true. Hence, the second restriction implies that if a pixel satisfies this condition then it should not be deleted.

A thinning algorithm should preserve an objects length (approximately, at least). To facilitate this, a third restriction must be imposed such that the arc ends, i.e. object pixels which are adjacent to just one other object pixel, are not deleted.

The final thinning algorithm, then, is to scan the image in a raster fashion, removing all object pixels according to these three restrictions, varying border orientation from pass to pass. The image is thinned until four successive passes (corresponding to the four border orientations), producing no changes to the image, are made. At this stage, thinning ceases.

2.3 Analysing The Image.

There are essentially two features that need to be extracted from the image:

- Identification of a good point at which the robot end-effector should grasp the wire and the orientation (tangent) of this point on the wire.
- The position and orientation of the wire end in relation to the point at which the wire is to be grasped.

The orientations are required because unless the wire is gripped at right angles to the tangent at the grasp point the wire will rotate in compliance with the finger grasping force. The orientation of the endpoint is important when inserting the wire in the crimping-press as the wire is introduced along a path coincident with the tangent to the wire at the end point.

A wire segment is defined as a subsection of a wire bounded at each end by either wire-crossing or an arc-end (wire segment end). Thus, a wire segment with two valid end points, at least one of which is an arc-end, and with a length greater than some predefined system tolerance, contains a feasible grasp point. This is a point some suitable fixed distance (15 mm) from the wire end. Additional constraints have been incorporated to avoid selecting wires in situations where the wires occlude each other near their ends.

Once the positions (i and j image coordinates) of both the grasp point and the end point are known, the orientation or the tangential angles of these two points must be estimated. The tangent to the wire at the grasp point is assumed to be parallel to the line joining two skeletal points displaced by two pixels on either side of the grasp point. The tangent to the wire end is assumed to be parallel to a line joining the end point and a skeletal point three pixels from the end. Both these tangential angles are estimated using the world coordinates corresponding to these pixel positions.

The complete analysis algorithm comprises two subsections: a raster scan mechanism and a wire tracking mechanism. The image is scanned in a raster fashion until an object pixel is encountered. If this pixel has not been previously visited its coordinates are passed to the tracking algorithm. If it has been encountered then it must lie on a path already rejected for grasping purposes and the scan simply proceeds. Scanning continues until all image pixels have been visited or until a valid grasping point is found by the tracking mechanism. A spacing of twenty between successive scan lines has been chosen since this affords a more efficient way of searching the image. This necessitates that twenty passes of the image may be required by the scanning algorithm, each pass starting at an origin displaced from the previous one by one pixel spacing.

The wire tracking mechanism involves following the wire path in all directions from the point passed to it by the scanning algorithm. It tracks until either a wire crossing or a wire end is encountered. If a wire segment with two valid endpoints (arc-ends or valid crossings), at least one of which is an arc-end, and with a length greater than a pre-defined system tolerance is encountered, then the point at a fixed short distance from the arc-end of this segment qualifies as a valid grasp point and is identified as such. The coordinates of the endpoint are then extracted by following the wire to the end; tangential angles are determined in accordance with the techniques discussed above.

Robot Vision in Automated Electrical Wire Crimping

Vision Robotique Appliquée au Sertissage Automatique de Fils Electriques

David Vernon

2.4 The Camera Model: Interfacing Between Images And The Real-world.

When using robot vision to identify the position (and orientation) of objects to be manipulated by a robot, the relationship between the image coordinate reference frame and the real-world reference frame must be established.

Simple approaches to this problem, e.g. [BALL82, pp.481-484], assume that the image/robot reference frame relationship is linear. However, there is a significant non-linear component in the present TCD camera/robot configuration. This is due to geometric distortion introduced by the imaging system and inaccuracies in the actual robot calibration. Consequently, this approach to generating a camera model was unsuccessful. So, in addition to defining the relationship between image and real-world, a useful solution must also model the spatial or geometric distortion in the camera and the non-linearity of a supposedly cartesian robot reference frame. To facilitate such a solution the requisite transformation is confined to a plane-to-plane non-linear mapping, normally referred to as a spatial warping function. This implies that for any given plane in the real-world one may generate a transformation between image coordinates (i,j) and robot coordinates (x,y), assuming z is constant and known. This relationship may be expressed by the following equation:

$$(x,y) = (W_x(i,j),\ W_Y(i,j))$$

Thus, given any image point (i,j) the corresponding robot x and y coordinates may be generated using the warping functions W_x and W_Y, respectively. Since analytic expressions for W_x and W_Y will rarely be known, a common approach is to model each spatial warping function by an nth order polynomial [PRAT78, p.430-432]. Third-order polynomials (in the two variables i and j) have been used for this application and the problem is now to derive the sixteen coefficients of each polynomial. The solution is facilitated by solving two sets of sixteen simultaneous equations, of the form of the polynomials in W_x and W_Y respectively, derived by associating sixteen control points in the real-world with their corresponding sixteen image points. The values of the coordinates of these image and real-world points are determined empirically. In this implementation, the system has been over-determined by using thirty six points (since exact solutions to this problem were found to be ill-conditioned) and a least-square-error estimate of the polynomial coefficients computed.

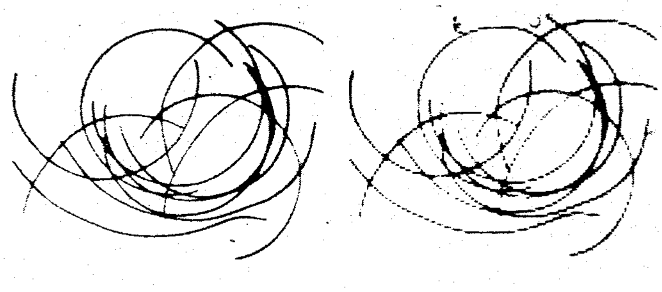
To model the robot non-linearity the robot has been programmed to identify the thirty six points itself; the manipulator maps out thirty six points as a six by six square grid. An image is generated of the resulting (almost square) grid and displayed on an in-house Vicom image processor monitor. The corresponding image point coordinates are then determined interactively by using a cursor. Once the polynomial coefficients have been computed they are saved on file for subsequent use by the robot vision suite of programs.

3.0 IMPLEMENTATION.

The present implementation is configured as a development system and not as a final industrial prototype. A Vicom image processing unit is used for image acquisition and the images are passed to a multi-user Vax 11/780 via a DMA link for processing and analysis. The robot used is a Smart Arm 6R/600 DC servo-controlled manipulator with five degrees of freedom, supervised by an Acorn Atom; the appropriate control signals are sent by the Vax to the Atom via an RS232C serial link. No special lighting is used, the only light source is overhead room strip-lighting. All algorithms have been developed for use with images of a spatial resolution of 128x128; since the Vicom digitises images at a resolution of 512x512 pixels, the acquired image is transformed to a 128x128 image by local averaging.

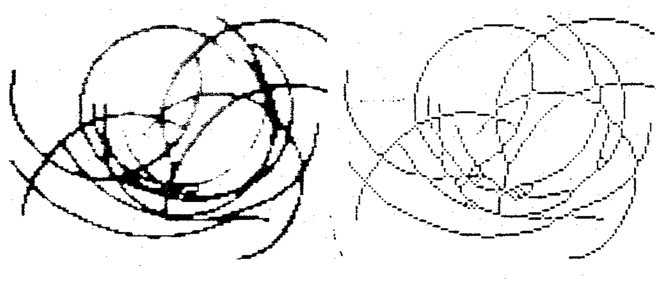
4.0 RESULTS

A typical scene is shown in diagram 2(a) while diagram 2(b) shows the corresponding 128x128 resolution grey-scale image. The segmented binary image obtained using an automatically determined threshold is shown in diagram 3(a) and the thinned image is shown in diagram 3(b). The selected grasp point is also illustrated in diagram 3(b).



(a) (b)

Diagram 2. (a) Typical scene. (b) 128x128 grey-scale image.



(a) (b)

Diagram 3. (a) Thresholded image. (b) Thinned image.

The average time taken to determine a feasible grasp point is 6.5 seconds; the average sensing process times are summarised in Table 1. Note: When calculating the average time taken to determine a grasp point the preliminary automatic threshold detection time is excluded. In addition, the Vicom-to-Vax transfer time and 512-to-128 resolution conversion time would not be applicable in a target system and, as such, these times are not included either.

Process	Average Time (sec.)
Image acquisition	0.10
Transfer to Vax	3.87
Generate 128x128 image	4.95
Threshold selection	45.80
Thresholding	0.23
Thinning	5.86
Image analysis	0.28

TABLE 1 - Average Sensing Process Times.

5.0 CONCLUSIONS.

The average sensing cycle time, with the current implementation, is lengthy. It remains to be seen what speeds will be obtained with a dedicated target system; consideration may have to be given to implementing low-level front-end processing in hardware.

Given the assumptions about scene complexity (small, well-scattered, well-illuminated wires no more than one of two layers deep, and a clearly visible contrasting background), the binary image techniques detailed here are entirely appropriate. However, once the organisation of the wires becomes more complex, with many layers of wires occluding both themselves and the background, the required information may no longer be extracted with these techniques. Efficient techniques for grey-scale image analysis are currently being researched to cater for such situations.

References.

BALL82 Ballard, D.H. and Brown, C.M.. "Computer Vision", Prentice-Hall, New Jersey, 1982.

BLUM62 Blum, H., "An associative machine for dealing with the visual field and some of its related properties", Biol. Prot. and Synth. Syst. 1, pp.244-260.

BLUM67 Blum, H., "A transformation for extracting new descriptors of shape", in "Models for the Perception of Speech and Visual Form", Wather-Dunn, W. (Ed.), MIT Press, Cambridge, Mass., 1967, pp.362-380.

DIXO78 Dixon, R.N. and Taylor, C.J., "Automated Asbestos Fibre Counting", in "Machine-aided Image Analysis", Gardener, W. (Ed.), 1978, pp178-185.

KASV78 Kasvand, T., "Experiments on Automatic Extraction of Paper Pulp Fibres", Proceedings of the Fourth International Joint Conference on Pattern Recognition, Nov. 7-10, 1978, pp.958-960.

KATZ65 Katz, Y.H., "Pattern Recognition of Meteorological Satellite Cloud Photography", Proc. Third Symposium on Remote Sensing of the Environment pp.173-214, Institute of Science and Technology, University of Michigan, February 1965.

LEVI70 Levi, G. ad Montanari, U., "A Grey-Weighted Skeleton", Information and Control, 17, 1970, pp.62-91.

MARR80 Marr, D., and Hildreth, E. "Theory of Edge Detection", Proc. Royal Soc. London, B207, 1980, pp.187-217.

MOTZ35 Motzkin, Th., "Sur quelques Proprietes caracteristiques des ensembles bornes non convexes", Atti. Acad. Naz. Lincei, 21, 1935, pp.773-779.

PAVL78 Pavlidis, T., "A Review of Algorithms for Shape Analysis", Computer Graphics and Image Processing, Vol. 7, 1978, pp.243-258.

PRAT78 Pratt, W.K.. "Digital Image Processing", Wiley-Interscience, New York, 1978.

ROSE82 Rosenfeld, A. and Kak, A., "Digital Picture Processing", Vol. 2., Academic Press, 1982.

SERR82 Serra, J., "Image Analysis and Mathematical Morphology", Academic Press, 1982.

SIMO80 Simons, G.L., "Robots in Industry", NCC Publications, Manchester, England, 1980.

TAMU78 Tamura, H., "A comparison of line-thinning algorithms from digital geometry view-point", Proc. 4th. Int. Joint Conf. on Pattern Recognition, 1978, pp.715-719.

TSUJ82 Tsuji, S., Yachida, Y., Guo, H., and Chou, Y., "Wiresight: Robot Vision for determining Three-dimensional Geometry of Flexible wires", Proc. 6th. Int. Conf. Pattern Recognition, Vol. 1, 1982, pp.220-222.

WESZ78 Weszka, J., "A Survey of Threshold Selection Techniques", Computer Graphics and Image Processing, vol. 7, 1978, pp.259-265.