DETECTION AND REPRESENTATION OF COMPLEX LOCAL FEATURES

Klas Nordberg and Robert Söderberg

Computer Vision Laboratory Department of Electrical Engineering Linköping University SE-581 83 LINKÖPING

ABSTRACT

This paper describes a method for extracting point features from an image, corresponding to corners and crossings of lines. The method is based on a local estimation of a 6×6 tensor which describes the parameters of a pair of line segments. By considering the rank of the tensor, it is possible to find points of interest. These points can then be further analyzed to provide detailed information about the configuration of the segments. The proposed method is intended for features which can be used for estimation of position and pose of 3D objects, e.g., for the purpose of grasping.

1. WHAT IS THE PROBLEM?

Detection of image features which are of higher complexity than lines or edges, such as corners, crossings, junctions, etc, can be done using rather simple methods, e.g., the so-called Harris corner detector, [1]. The downside of these methods are that the spatial localization of the corresponding feature is not very accurate and they are not very selective in the sense that they are likely to be sensitive also to noise, point patterns, etc. Furthermore, even though some methods use various representations for describing what phenomena is detected, these are not rich enough to distinguish between a corner and a T-junction, or describe the opening angle of a corner.

By finding interest points it is possible to train a system to estimate the pose of an object, [2], but in order to improve this process, we need point-like features which describe the actual image data in a richer way than just "here is something which may be interesting".

The problem discussed in this paper is the representation of multiple line or edge segments in a local region. We propose a representation which can be used for multiple segments in terms of their orientations and positions within the local region. This allows a representation of corner, crossing, and junctions, which both distinguishes between the different cases and describes the parameters of the segments. It is also shown that the representation can be estimated directly from image data using simple operations of convolution type.

2. IS THERE A SOLUTION?

We seek a representation which can be used for multiple line segments and which can describe their orientations and positions. Where do we start? Let us begin with the usual orientation tensor \mathbf{T} , a symmetric 2×2 matrix which can be used for describing the orientation of an edge segment:

$$\mathbf{T} = A \,\hat{\mathbf{n}}\hat{\mathbf{n}}^T.\tag{1}$$

 $\hat{\mathbf{n}}$ is the normal of the edge and A > 0 is related to the "strength" or contrast of the edge. Different methods for computing \mathbf{T} have been described for both 2D and 3D image data, see [3]. Taking the symmetry of into account, \mathbf{T} can be represented as a 3-dimensional vector in the 2D case.

T does not provide any information about the position of a specific edge. However, if **T** has been estimated at the point $\mathbf{x} = (x_1, x_2)^T$, the combination of \mathbf{x} and $\mathbf{T} = \mathbf{T}(\mathbf{x})$ represents the fact that *at* \mathbf{x} *there is a line segment with orientation given by* **T**. We refer to the tuple $\mathcal{O} = {\mathbf{x}, \mathbf{T}}$ as an *observation*. In the following we will develop a representation which can maintain several simultaneous observations. This is done using the following strategy: the representation of different parameters related to the same observation is constructed by taking their outer products, or tensor product " \otimes ", resulting in a matrix or tensor, and combinations of multiple observations are made by superposition.

In this presentation, multiple observations imply *independent* observations, by which we mean that they refer to segments which are not part of the same line. Consequently, the goal is to develop a representation which allows the separation of independent observations. Furthermore, in general all coordinates refer to a local coordinate system, usually related to the local region from which the representation is estimated. The goal is to define a representation such that the number of independent observations corresponds to the rank of the resulting matrix.

This work has been funded by the VISATEC project, IST-2001-34220.

The first step is to formally combine \mathbf{x} and \mathbf{T} into a single object by taking their outer product. Since \mathbf{x} is a 2-dimensional vector and $\mathbf{T}(\mathbf{x})$ is a 3-dimensional vector, their outer product $\mathbf{x} \otimes \mathbf{T}(\mathbf{x})$ can be represented as a 2 × 3 matrix **S**. Furthermore, two observations $\mathcal{O}_1 = {\mathbf{x}_1, \mathbf{T}(\mathbf{x}_1)}$ and $\mathcal{O}_2 = {\mathbf{x}_2, \mathbf{T}(\mathbf{x}_2)}$ are represented by

$$\mathbf{S} = \mathbf{x}_1 \otimes \mathbf{T}(\mathbf{x}_1) + \mathbf{x}_2 \otimes \mathbf{T}(\mathbf{x}_2).$$
(2)

However, the current representation of \mathcal{O} leads to difficulties. First, A in (1) is related both to the local edge contrast and to the specific implementation of the method for estimating **T**. This means that given **S**, we are not able to unambiguously determine both **x** and **T** if A is unknown. Second, the idea that the rank of **S** should correspond to the number of independent observations fails. If the two observations describe edge segments of the same orientation, i.e., $\mathbf{T}(\mathbf{x}_1) \propto \mathbf{T}(\mathbf{x}_2)$, then **S** in (2) is of rank one. This also happens if \mathcal{O}_1 and \mathcal{O}_2 describe segments for which $\mathbf{x}_1 \propto \mathbf{x}_2$.

A solution to the first problem is to use a homogeneous representation of \mathbf{x} , for example:

$$\mathbf{x}_H = \begin{pmatrix} 1 & x_1 & x_2 \end{pmatrix}^T. \tag{3}$$

If we define $\mathbf{S} = \mathbf{x}_H \otimes \mathbf{T}$, it is always possible to determine \mathbf{x} in an unambiguous way given \mathbf{S} .

The second problem can be solved by mapping also T into a projective space, which is done in the following way: define the 3×2 matrix K as

$$\mathbf{K} = \mathbf{K}(\mathbf{x}) = \begin{pmatrix} -x_1 & -x_2\\ 1 & 0\\ 0 & 1 \end{pmatrix}.$$
 (4)

Given a line segment with normal vector $\hat{\mathbf{n}}$ at point \mathbf{x} , let us consider the 3-dimensional vector \mathbf{l}^H defined by

$$\mathbf{l}^{H} = \mathbf{l}^{H}(\mathbf{x}) = \mathbf{K}(\mathbf{x}) \,\hat{\mathbf{n}}(\mathbf{x}). \tag{5}$$

This vector is the *dual homogeneous coordinates* of the line with normal vector $\hat{\mathbf{n}}$ passing through the point \mathbf{x} . This can be seen by considering the product

$$d = \mathbf{y}_H^T \mathbf{l}^H = \hat{\mathbf{n}}^T (\mathbf{x} - \mathbf{y}), \tag{6}$$

where y_H is the homogeneous coordinates of an arbitrary point y. Clearly, d vanishes if and only if y is on the line. To improve our representation, we replace T by

$$\mathbf{S}_{02} = \mathbf{K} \, \mathbf{T} \, \mathbf{K}^T = A \, \mathbf{l}^H \otimes \mathbf{l}^H. \tag{7}$$

Notice that S_{02} is a function of x and that it is a symmetric 3×3 matrix, i.e., can be seen as a 6-dimensional vector.

Using $\mathbf{S} = \mathbf{x}_H \otimes \mathbf{S}_{02}$ as the representation of \mathcal{O} assures that \mathbf{S} is of rank one for a single observation and of rank two if it is the superposition of two observations. However,

this construction does not guarantee that the superposition of three independent observations results in a rank three matrix S. To reach this property, we can replace x_H by

$$\mathbf{S}_{20} = \mathbf{x}_H \otimes \mathbf{x}_H,\tag{8}$$

which we can see as a 6-dimensional vector. Finally, then, we arrive at a representation of a single observation \mathcal{O} according to

$$\mathbf{S} = \mathbf{S}_{22} = \mathbf{S}_{20} \otimes \mathbf{S}_{02},\tag{9}$$

which can be implemented as a 6×6 matrix.

The replacement of \mathbf{x}_H by \mathbf{S}_{20} also has the advantage of allowing us to estimate both first and second order moments of a segment. An edge segment does in practice never consist of a single point. Instead, for a particular segment we observe a set of points Γ which all have approximately the same orientation, given by \mathbf{T} . Since all these points are on the same line, given by \mathbf{I}^H , this means that \mathbf{S}_{02} in (7) is constant except for the variation in A which depends on \mathbf{x} . Consequently, the superposition of \mathbf{S}_{22} for all points in this set is of rank one:

$$\mathbf{S}_{22} = \left[\sum_{\mathbf{x}\in\Gamma} A(\mathbf{x}) \, \mathbf{x}_H \otimes \mathbf{x}_H\right] \otimes \mathbf{S}_{02}.$$
 (10)

Notice that the summation in the last expression results in weighted moments of order zero, one and two of the corresponding segment. This means that from S_{22} we can estimate both the centroid of a segments and its extension.

3. CAN IT BE ESTIMATED?

Given that we want to use S_{22} , (9), as a representation of an observation \mathcal{O} and superpositions of such terms for multiple observations, we now turn to the issue of how to estimate such a representation. One approach is based on first estimating a dense field of orientation tensors. Given a local region Ω , we can introduce a local coordinate system, and compute S_{22} as

$$\mathbf{S}_{22} = \sum_{\mathbf{x}\in\Omega} W(\mathbf{x}) \left[\mathbf{x}_H \otimes \mathbf{x}_H \right] \otimes \left[\mathbf{K}(\mathbf{x}) \ \mathbf{T}(\mathbf{x}) \ \mathbf{K}(\mathbf{x})^T \right] =$$
$$= \sum_{\mathbf{x}\in\Omega} W(\mathbf{x}) \ A(\mathbf{x}) \ \mathbf{S}_{20}(\mathbf{x}) \otimes \mathbf{S}_{02}(\mathbf{x}), \tag{11}$$

where W is a suitably chosen weighting function which we use to localize the estimated representation. Under the assumption that T is approximately zero except on the edges, it follows that S_{22} in (11) can be written as

$$\mathbf{S}_{22} = \sum_{k} \mathbf{S}_{20,k} \otimes \mathbf{S}_{02,k},\tag{12}$$

where the summation is taken over all segments in Ω and

$$\mathbf{S}_{20,k} = \sum_{\mathbf{x}\in\Gamma_k} W(\mathbf{x}) A(\mathbf{x}) \mathbf{x}_H \otimes \mathbf{x}_H, \qquad (13)$$

where Γ_k is the set of point related to segment k. Here, $\mathbf{S}_{02,k} = \mathbf{l}_k^H \otimes \mathbf{l}_k^H$, where \mathbf{l}_k^H is the dual homogeneous coordinates of segment k. Consequently, \mathbf{S}_{22} computed as in (11) results in a superposition of terms, each of the form described in (10) (with the additional weighting of W) and one term for each independent segment.

Furthermore, it should be clear from the previous presentation that S_{22} in (11) is computed by locally correlating the elements of **T** with polynomials of up to fourth order weighted by W. This local operation corresponds to the global operation of convolving the elements of **T** with these weighted polynomials. By choosing a separable W, e.g., a Gaussian, and because the polynomials are separable, it follows that these convolutions can be made in terms of linear combinations of separable convolutions.

4. DETECTION OF SECOND LEVEL FEATURES

Given that we can compute S_{22} at arbitrary points or even every point in an image, how can we use this data for image processing? One possible task which can be achieved is to find points or regions which contains two independent segments, e.g., corners or junctions. There are simpler methods which can do this, e.g., [1], but the proposed representation allows us also to determine the parameters of a junction, e.g., the position and orientation the corresponding segments, with a high accuracy.

To start with, we want to find points which are characterized by being close to two edge segments. If we have estimated \mathbf{S}_{22} for an image point, this means that we want to know if \mathbf{S}_{22} has rank two. This can be done in different ways, but the specific approach chosen here is to compute a Singular Value Decomposition (SVD) of the 6×6 matrix \mathbf{S}_{22} . If there are exactly two non-zero singular values, then the rank is two, and vice versa. However, in practice all singular values are non-zero which means that we need some method for estimating a qualitatively measure of "rank twoness" of \mathbf{S}_{22} . This can be done by taking the three largest singular values $\sigma_1, \sigma_2, \sigma_3$, and compute the following parameters:

$$t = \sigma_1 + \sigma_2 + \sigma_3, \qquad d = \sigma_1 \sigma_2 \sigma_3, q = \sigma_1 \sigma_2 + \sigma_2 \sigma_3 + \sigma_3 \sigma_1, \qquad c_2 = \frac{-9d + qt}{3d - 3qt + t^3}.$$
(14)

From this follows directly that $c_2 = 1$ if $\sigma_1 = \sigma_2 \neq 0$ and $\sigma_3 = 0$, and it vanishes if \mathbf{S}_{22} has rank one or if $\sigma_1 = \sigma_2 = \sigma_3 \neq 0$. Consequently, c_k represents a measure of confidence for rank k.

Figure 1 shows a test image together with the corresponding value of c_2 , weighted with the norm of S_{22} . In the latter image, bright areas indicate tensors with a high confidence of rank two. As expected these areas are close



Fig. 1. Test image, Each pixel is a measurement of the corresponding tensors confidence of being rank two, c_2 .

to corners and local regions which contains two line segments. By searching, e.g., local maxima in this image we can detect points of interest. Notice that the bright areas are always found inside the corners and with the maximal value at some distance from a corner if it is sharper. This indicates that the rank measure is not only dependent on the number of line segments but also on the distance to the corner or junction and the angle between the corresponding line segments.

5. ANALYSIS OF LOCAL FEATURES

A detailed presentation of the analysis of the S_{22} tensor is found in [4]. The basic idea to be used here is that once we have found a point of interest, we try to estimate the corresponding values of S_{20} and S_{02} for each of the segments. This can be done by computing a SVD of S_{22} , which we already have in order to obtain c_2 .

$$\mathbf{S}_{22} = \mathbf{U} \Sigma \mathbf{V}^T = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T.$$
(15)

It must then be the case that $S_{02,1}$ and $S_{02,2}$ are linear combinations of v_1 and v_2 , i.e.

$$\mathbf{S}_{02,k} = \alpha_{1,k} \, \mathbf{v}_1 + \alpha_{2,k} \, \mathbf{v}_2. \tag{16}$$

Each $S_{02,k}$ is a 6-dimensional vector which can be rearranged into a 3×3 matrix, which in addition should be of rank 1. This last property can be used to determine the coefficients $\alpha_{i,k}$ in a straightforward way, at least to the order of a common scaling. Since S_{02} is an element of a projective space, this scaling does not introduce any ambiguities in the interesting parameters. Here, this is done by setting the coefficient of the first order term in the characteristical polynomial of $S_{02,k}$ to zero and solve for $\alpha_{i,k}$. Once a transformation **A** on the right singular vectors has been established, which maps them into tensors $S_{02,k}$, there is a corresponding transformation on the left singular vectors which keeps S_{22} invariant.

$$\mathbf{S}_{22} = (\mathbf{U} \Sigma \mathbf{A}^{-1}) (\mathbf{A} \mathbf{V}^T).$$
(17)

Fig. 2. Detection of two local segments applied on a synthetic image. Notice that the position and extension of the segments allow us to discriminate between, e.g., corners and T-junctions

The results are estimates of $S_{02,1}$ and $S_{20,2}$, in the form described in (13). Since these tensors contain first and second order moments of each segment, the tensors $S_{20,k}$ can be further analyzed to give the centroid, extension and the orientation of each segment. For the case of two segments, we can then visualize the content of a tensor S_{22} in the form of two line segments with orientation, centroid and extension according to the estimated parameters.

Figure 2 and 3 illustrate the results of performing the computations outlined above. As seen, the pairs of line segments drawn on top of the images lie very close to the real segments in the images.

6. CONCLUSIONS AND DISCUSSIONS

This paper presents a representation of local complex features for 2D images in form of a 6×6 tensor S_{22} . The rank of S_{22} equals number of independent edge segments up to rank three which can be used to detect regions which contains two segments. Further analysis of S_{22} can be made for providing information about the centroid, extension and orientation of each segment. This allows us to distinguish between corners, T-junctions or crossings, which can be useful in a pose estimation process where as much information as possible is needed for building high level features, [5]. A general process for doing this is to group lower level fea-



Fig. 3. The algorithm applied on a selection of boxes.

tures and we believe that this can be done in a more robust way by using the rich information extracted from the features described here, thereby reducing the risk of grouping features on the object of interest with the background.

7. REFERENCES

- C. G. Harris and M. Stephens, "A combined corner and edge detector," in *4th Alvey Vision Conference*, September 1988, pp. 147–151.
- [2] David G. Lowe, "Object recognition from local scaleinvariant features," in *Proc. ICCV'99*, 1999.
- [3] B. Jähne and H. Haussecker, Eds., Computer Vision and Applications, Academic Press, 2000, ISBN 0-12-379777-2.
- [4] K. Nordberg, "A fourth order tensor for representation of orientation and position of oriented segments," Tech. Rep. LiTH-ISY-R-2587, Dept. EE, Linköping University, SE-581 83 Linköping, Sweden, Februari 2004.
- [5] G. H. Granlund, "The complexity of vision," *Signal Processing*, vol. 74, no. 1, pp. 101–126, April 1999, Invited paper.