Invariant Analogical Image Representation and Pattern Recognition

Lowell Jacobsen and
Harry Wechsler

MLI 84-13

Pattern Recognition Letters 2 (1984) 289-299
North-Holland

September 1984

Invariant analogical image representation and pattern recognition

Lowell JACOBSON and Harry WECHSLER
Electrical Engineering Department, University of Minnesota Minneapolis, MN 55455, USA

Received 13 December 1983

Abstract: We describe an image representation that uniquely encodes the information in a gray-scale image, decouples the effects of illumination, reflectance, and angle of incidence, and is invariant, within a linear shift, to perspective, position, orientation, and size of arbitrary planar forms. Then we provide a theoretical basis for applying this representation to achieve invariant form recognition.

Key words: Image recognition, invariance, pattern recognition, perspective mapping, spatial frequency, Wigner distribution.

1. Introduction

The challenge of the visual recognition problem stems from the fact that the interpretation of a 3-D scene from a single 2-D image is confounded by several dimensions of variability. Such dimensions include uncertain perspective, position, orientation, and size (pure-geometric variability) along with sensor noise, object occlusion, and non-uniform illumination. Vision systems must not only be able to sense the identity of objects despite this variability, but they must also be able to explicitly characterize such variability. This is so because the variability in the image formation process (particularly that due to geometric distortion and varying angle of incident illumination) inherently carries much of the valuable information about the imaged scene. Consider human vision for a moment. In spite of the complications introduced by geometric distortion, it is precisely the 'unraveling' of such distortions that enables a human to readily perceive the three-dimensionality of any static 2-D image – be it a single face of a Necker cube, or a 15th century painting by Da Vinci. Indeed, humans seem capable of unraveling the physical and geometric distortions in an image almost as precisely as the physics and geometry of the world created them in the first place.

Contrasted with the apparent ease and elegance of human visual interpretation of scene geometry, current vision algorithms are clearly lacking. It is becoming increasingly clear that much of the blame lies with conventional image representations. Good image representations must satisfy a number of requirements which seem to be mutually incompatible. First, they should be simultaneously compact and complete in their representation of gray-scale image information. Compactness is synonymous with ease of computation and efficient use of memory. Completeness, on the other hand, implies that, if desired, one could fully reconstruct the original image from which the representation is derived. Secondly, image representations must provide good intra-object clustering and inter-object separability independent of image distortion while at the same time preserving information about pattern distortion. No conventional image representation satisfies all of the above conditions. Though many conventional representations claim compactness, most do not

make a credible attempt to decouple information about object identity from information about viewing geometry and illumination, nor do current representations fully exploit the abundance of valuable gray-scale information in an image. In contrast, invariant 'analogical' image representations, such as that advocated in this paper, can satisfy the above requirements.

2. Decoupling multiplicative processes in image formation

For the case of an ideally diffusing surface, an image can be modeled as a product of three independent signal components:

$$f(x, y) = i(x, y) \cdot r(x, y) \cdot \cos\alpha \qquad (1)$$

where $i(x, y)$ is the illumination, $r(x, y)$ is the reflectance, and α, $0 \leq \alpha \leq \pi/2$, is the angle of incidence of the illuminating light [14]. Let's assume that any additive noise that is present has small magnitude relative to the above three components of the signal. Then there is a well-known and effective method known as homomorphic filtering that allows one to individually filter out such multiplicative signal components when certain reasonable conditions are met. This method is briefly reviewed below.

Suppose one takes the logarithm of our function $f(x, y)$ as defined above. There results the so-called 'density image':

$$\ln f(x, y) = \ln i(x, y) + \ln r(x, y) + \ln \cos\alpha. \qquad (2)$$

Hence the product becomes a sum of three density components $\ln i$, $\ln r$ and $\ln \cos\alpha$. If these three additive density components have Fourier spectra which overlap very little in their regions of significant energy, then linear filtering can be used to extract any one of them. By taking the exponent of the extracted component, one obtains the corresponding multiplicative signal component which appeared in the original expression for the image $f(x, y)$. Homomorphic filtering has been applied to enhance imagery by selectively filtering out the slowly varying illumination component with very impressive results [15].

The success of homomorphic filtering clearly stems from the fact that the Fourier transform of a density image often has the effect of 'representationally' decoupling multipliciative image components. If image pattern recognition (rather than image filtering) could be based on such a Fourier representation, then perhaps the effects of surface reflectance, illumination, and angle of incidence could be decoupled. This in turn could lead to methods for form recognition that are insensitive to varying illumination conditions. Unfortunately, as is well known, neither the Fourier transform, nor its (phaseless) power spectrum, have proved to be especially useful for image pattern recognition. However, there are alternative representations – namely, simultaneous spatial/spatial-frequency representations – which, like the Fourier spectrum, can provide decoupling of multiplicative image density components and, at the same time, overcome the classical shortcomings of the Fourier spectrum as an image representation.

3. Simultaneous representations of space and spatial frequency

Vision researchers have traditionally emphasized the importance of either the spatial or the spatial-frequency domain, but not both. This should be contrasted with conventional representations of acoustic signals where simultaneous time/frequency representations (e.g. the spectrogram) have long been used. Nonetheless, vision researchers have very recently begun to express an active interest in simultaneous spatial/spatial-frequency image representations [2,8,9,10,11,12,13]. Such representations can provide for improved separability of information characterizing visually relevant patterns and they are also compatible with the representation of gray-scale characteristics ranging from textures to object contours [8]. Furthermore, as will be shown in Section 4, simultaneous spatial/spatial-frequency image representations can be used to decouple the effects of illumination, object reflectance, and angle of incidence.

3.1. The Wigner distribution (WD)

Two simultaneous spatial/spatial-frequency

representations have recently received much attention. These are the Gabor function representation [7,12,13] and the Wigner distribution [2,8,9,10, 11,12,20]. (Note that we are here concerned only with 4-D directionally selective representations; this excludes, for example, Marr's 3-D DOG representation [14] and others like it which do not have a fourth dimension covering spatial-frequency angle.) As discussed elsewhere [11], Wigner distribution provides higher simultaneous resolution than is possible using the Gabor functions. In fact as discussed in [6], every simultaneous representation ever proposed can be expressed in terms of averages of the Wigner distribution over its independent spatial and spatial-frequency variables. Like the Fourier transform, the Wigner distribution (WD) is not, in general, a computable function since this would require the evaluation of infinite integrals. However, just as the Fourier transform has proved to be an elegant and convenient transformation with which to handle many problems in the spatial-frequency domain, the WD is an invaluable tool for problems *simultaneously* involving the spatial and spatial-frequency domains. As with the Fourier transform, the WD can be used in practice by employing an approximation obtained via finite integration windows. A particularly attractive approximation to the WD, denoted as the 'composite pseudo Wigner distribution' (CPWD) has been introduced [11].

The definition of the Wigner distribution is given in the appendix and a number of useful properties can be found in [10]. We note here that the WD is a pure-real function (i.e., it has no associated phase), and that it is a reversible, hence complete representation as well. As indicated in the appendix, the WD of an image $f(x, y)$ is denoted as $W_f(x, y, u, v)$, where x, y are the spatial variables of the image, and u, v are the two spatial-frequency variables over which the Fourier transform is defined.

4. The WD and multiplicative signal component separability

Recall our brief discussion of homomorphic filtering and multiplicative signal component separability. The essence of the approach was to transform the image density function (i.e., the logarithm of the sensed image function) into the spatial-frequency domain in order to decouple the image components due to illumination, reflectance, and angle of incidence. To simplify our calculations, we assume a two component model including only illumination and reflectance:

$$f(x, y) = i(x, y) \cdot r(x, y). \tag{3}$$

Taking the logarithm,

$$\ln f(x, y) = \ln i(x, y) + \ln r(x, y), \tag{4}$$

and computing the WD of both sides yields

$$\begin{aligned} W_{\ln f}(x, y, u, v) = {} & W_{\ln i}(x, y, u, v) \\ & + W_{\ln r}(x, y, u, v) \\ & + 2\, W_{\ln i,\, \ln r}(x, y, u, v), \end{aligned} \tag{5}$$

where (x, y) and (u, v) specify the spatial and spatial-frequency domains, respectively. The third component, representing the cross-Wigner distribution [6] occurs because computing the WD involves a nonlinear correlation operation. Nevertheless, if the auto-WDs of the functions $\ln i$ and $\ln r$ do not overlap substantially in space and spatial frequency, then approximations to the cross-WD contribution will in practice contain negligible energy. (This non-obvious fact follows from considering the definition of the CPWD [11].) Therefore, as with the Fourier transform, if regions of significant spectral energies of $W_{\ln i}$ and $W_{\ln r}$ are disjoint, then separability of the multiplicative signal components will have been achieved. These observations generalize to the case where angle of incidence effects are included; then the right hand side of the above equation would include the auto-WD of the density function of this component along with its cross-WDs with the illumination and reflectance density functions. The Wigner distribution, like the Fourier transform, can therefore decouple the effects of illumination, reflectance, and angle of incidence when the spectra of these components are mutually disjoint. We describe in the remainder of this paper how the Wigner distribution can be employed to define a unique image representation that, in addition to providing decoupling of multiplicative image com-

ponents, also provides invariance to geometric distortions introduced by the imaging process.

5. Invariant form recognition in the frontoparallel plane

As discussed in the introduction, good image representations should not only decouple the effects of illumination, reflectance and angle of incidence, but they should also allow objects to be recognized irrespective of the a priori unknown geometric distortions introduced by the image formation process. We begin this section by discussing only methods for obtaining invariance to linear transformations of non-occluded, planar gray-scale patterns in frontoparallel view. The more general problem including perspective distortion will be considered in a later section. One particular approach for obtaining invariance to linear transformations makes use of the complex-logarithmic (CL) conformal mapping. Some researchers [16,17,18,19] have advocated use of the representation derived by CL conformally mapping the gray-scale image function itself. Others [3,4,5] have suggested that the CL conformally mapped Fourier power (or magnitude) spectrum should be used. In fact, neither representation seems entirely adequate. The former representation (CL mapped image function) is indeed invariant, within a linear shift (WALS-invariant), to rotation and scaling of the image about a single image point. However, such a representation is not invariant to translation of an image and the effects of illumination and reflectance are not in any way decoupled from one another. The second representation mentioned above (CL conformally mapped power spectrum of an image), is strictly invariant to translation and WALS-invariant to rotation and scaling of an image, but it does not uniquely represent an image since Fourier phase information is discarded. As alluded to earlier in our discussion of multiplicative signal separability, the loss of phase information makes Fourier spectra especially ill-suited for imagery containing clutter or multiple objects to be recognized. Since the WALS-invariance properties of the above-mentioned representations arise exclusively from using the CL conformal mapping, one is naturally led to consider whether this mapping could be used to develop considerably more robust image representations (perhaps simultaneous spatial/spatial-frequency representations) which are similarly WALS-invariant to rotation and scale changes. We will shortly show that this is possible, but first let's review the mathematical formulation of the CL conformal mapping.

5.1. *The complex logarithmic conformal mapping*

We choose to represent points in the Cartesian plane by

$$(x, y) = (\mathrm{Re}(z),\ \mathrm{Im}(z)),$$

where $z = x + jy$. Thus we can write

$$z = r \cdot \exp\{j\theta\}, \tag{6}$$

where

$$r = |z| = (x^2 + y^2)^{1/2},$$

and

$$\theta = \arg(z) = \arctan(y/x).$$

Now the CL mapping is simply the conformal mapping of points z onto points w defined by

$$w = \ln(z) = \ln(r \cdot \exp\{j\theta\}) = \ln r + j\theta. \tag{7}$$

Thus, points in the target domain are given by

$$(\ln r, \theta) = (\mathrm{Re}(w),\ \mathrm{Im}(w)).$$

The effect of this mapping is shown in Fig. 1. Logarithmically spaced concentric rings and radials of uniform angular spacing are mapped into uniformly spaced straight lines. More generally, after CL mapping, rotation and scaling about the origin in the Cartesian domain correspond to simple linear shifts in the θ and $\ln r$ directions, respectively.

5.2. *The conformally mapped Wigner distribution*

Assume we are given an arbitrary image density function $\ln f(x, y)$ and also a function $\ln g(x, y)$ which is obtained by arbitrarily scaling and rotating $\ln f(x, y)$ about the origin $(x, y) = (0,0)$. Then the corresponding Wigner distributions $W_{\ln g}$

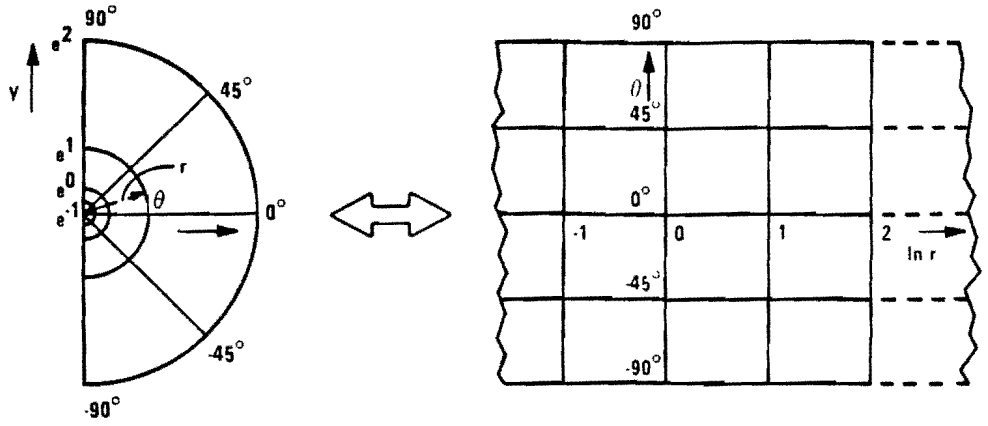


Fig. 1. The complex-logarithmic conformal mapping of the Cartesian half-plane.

and $W_{\ln f}$ are related by

$$W_{\ln g}(x, y, u, v) = k^2 W_{\ln f}\left(\frac{x\cos\phi + y\sin\phi}{k}, \frac{-x\sin\phi + y\cos\phi}{k}, \frac{u\cos\phi + v\sin\phi}{1/k}, \frac{-u\sin\phi + v\cos\phi}{1/k}\right), \tag{8}$$

where k and ϕ denote the scale factor and rotation angle, respectively, relating $\ln g(x, y)$ to $\ln f(x, y)$. Now if one CL conformally maps both the spatial and spatial-frequency domains of $W_{\ln f}$ and $W_{\ln g}$ with respect to the origin $(x, y, u, v) = (0, 0, 0, 0)$, then the corresponding CL mapped WDs $\tilde{W}_{\ln f}$ and $\tilde{W}_{\ln g}$ are related as follows:

$$\tilde{W}_{\ln g}(z', \theta', w', \xi') = k^2 \tilde{W}_{\ln f}(z' - \ln k, \theta' - \phi, w' + \ln k, \xi' - \phi), \tag{9}$$

where k and ϕ are as defined above,

$$z' = \ln(x^2 + y^2)^{1/2}, \qquad \theta' = \arctan(y/x),$$
$$w' = \ln(u^2 + v^2)^{1/2}, \qquad \xi' = \arctan(v/u).$$

Hence we see that the 4-D CL conformally mapped WD is WALS-invariant to rotation and scaling of the image $f(x, y)$ about the origin $(x, y) = (0, 0)$. Note that $W_{\ln g}$ can only translate with respect to $W_{\ln f}$ along a 2-D hyperplane

$$\{(z', \theta', w', \xi') | z' = -w', \theta' = \xi'\}$$

in the 4-D representational space. That there are only two degrees of translation freedom in the 4-D space, corresponding to the separate effects of rotation and scaling, is to be expected. This is because the spatial and spatial-frequency domains of the WD undergo reciprocal scaling and equal rotation in response to scaling and rotation, respectively, of the underlying image.

Suppose that $f(x, y) = i(x, y) \cdot r(x, y)$ and that $g(x, y)$ corresponds to $r(x, y)$ rotated and scaled about the origin by angle ϕ and factor k, respectively, while $i(x, y)$ remains unchanged. Then if $\tilde{W}_{\ln i}$ and $\tilde{W}_{\ln r}$ have good separability, both before and after the linear spatial transformation of $r(x, y)$, we obtain by using Eqs. (5) and (9),

$$\tilde{W}_{\ln g}(z', \theta', w', \xi') = k^2 \tilde{W}_{\ln r}(z' - \ln k, \theta' - \phi, w' + \ln k, \xi' - \phi) + \tilde{W}_{\ln i}(z', \theta', w', \xi'). \tag{10}$$

Therefore, subject to reasonable restrictions, different multiplicative components of an image can undergo independent translations within the 4-D rotation and scale WALS-invariant representation introduced above.

It should be recalled that the CL mapped WD is WALS-invariant to rotation and scale changes on-

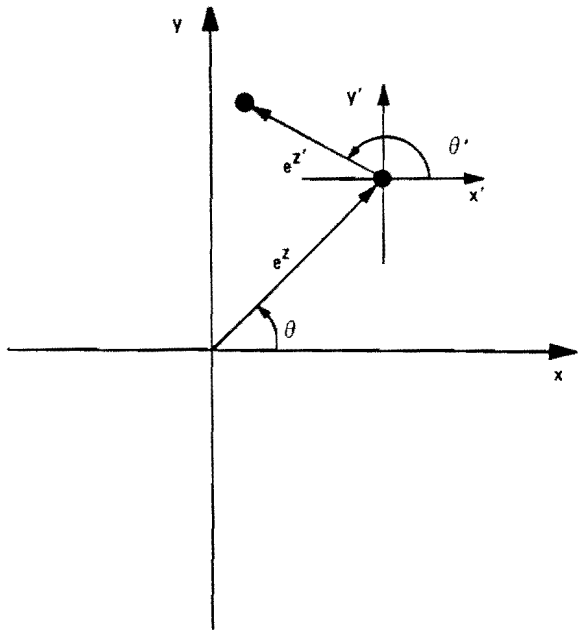


Fig. 2. The coordinates used in multipositional CL mapping of the WD.

ly about a single image point, namely the spatial domain origin $(x, y) = (0, 0)$. Therefore it has previously been suggested [9] that the 'fixation point' (i.e. the origin) be sequentially moved about to obtain rotation and scale invariant representation of each frontoparallel planar form in an image. Alternatively, at the cost of adding two additional dimensions to the proposed representation, one can obtain a new representation which provides WALS-invariance with respect to translation, as well as rotation and scale changes. Referring to Fig. 2, one simply constructs the 6-D representation

$$\hat{W}_f(z', \theta', w', \xi'; z, \theta),$$

obtained by performing spatial CL conformal mapping about each point $(x, y) = (e^z \cos\theta, e^z \sin\theta)$ in the spatial domain. (As before, spatial-frequency domain conformal mapping is done only about the origin $(u, v) = (0, 0)$.) It is interesting to note that the idea of performing multipositional conformal mappings was first proposed

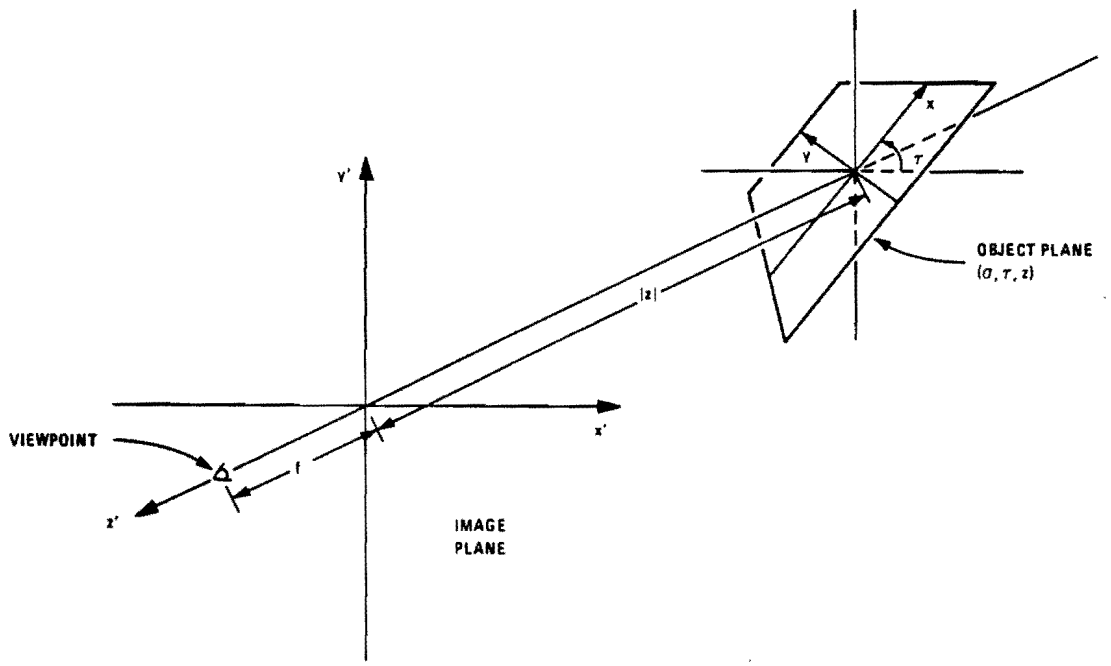


Fig. 3. The geometry relating the object and image planes.

as part of a model for the function of the primary visual cortex in mammals [17,18].

6. Obtaining invariance to perspective distortion

Our discussion has previously assumed imagery of planar, non-occluding forms in some frontoparallel plane. We now relax these conditions to include non-frontoparallel views and describe how one can obtain invariance to perspective distortions of gray-scale planar forms. Given the geometry of Fig. 3, we assume some arbitrary object plane (σ, τ, z) where σ, τ, and $|z|$ specify respectively the plane's slant, tilt, and line-of-sight distance from the viewpoint. Let an arbitrary point in the given object plane, whose object plane coordinates are (x_0, y_0), have world coordinates (x_0', y_0', z_0') in the (x', y', z')-space. Then from the geometry of Fig. 3, it is easy to see that our arbitrary point's object plane and world coordinates are related by:

$$\begin{bmatrix} x_0' \\ y_0' \\ z_0' \end{bmatrix} = \begin{bmatrix} x_0\cos\tau - y_0\sin\tau\cos\sigma \\ x_0\sin\tau + y_0\cos\tau\cos\sigma \\ z - y_0\sin\sigma \end{bmatrix}. \tag{11}$$

From the familiar formula governing perspective projection onto a plane [1], a point having world coordinates (x_0', y_0', z_0') projects to a point $(x', y', 0)$ in the image plane given by

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \frac{f}{f - z_0'}\begin{bmatrix} x_0' \\ y_0' \end{bmatrix}. \tag{12}$$

Now using Eq. (11) to substitute for x_0', y_0', and z_0' in the above equation, we obtain the transformation from object plane coordinates (x_0, y_0) to image plane coordinates $(x', y', 0)$:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \frac{f}{f - z + y\sin\sigma}\begin{bmatrix} x\cos\tau - y\sin\tau\cos\sigma \\ x\sin\tau + y\cos\tau\cos\sigma \end{bmatrix}, \tag{13}$$

where we have dropped the subscripts on the object plane coordinates since these were arbitrarily chosen. Now we wish to invert the above relation to obtain an expression for the object plane coordinates (x, y) in terms of its corresponding image plane coordinates (x', y'). After some algebraic manipulations, Eq. (13) becomes

$$\begin{bmatrix} \cos\tau & \left(-\sin\tau\cos\sigma - \dfrac{x'\sin\sigma}{f}\right) \\ \sin\tau & \left(\cos\tau\cos\sigma - \dfrac{y'\sin\sigma}{f}\right) \end{bmatrix} \cdot \begin{bmatrix} \dfrac{x}{f-z} \\ \dfrac{y}{f-z} \end{bmatrix} = \begin{bmatrix} \dfrac{x'}{f} \\ \dfrac{y'}{f} \end{bmatrix}. \tag{14}$$

Finally, using Cramer's rule to solve this system for $x/(f-z)$ and $y/(f-z)$, we get

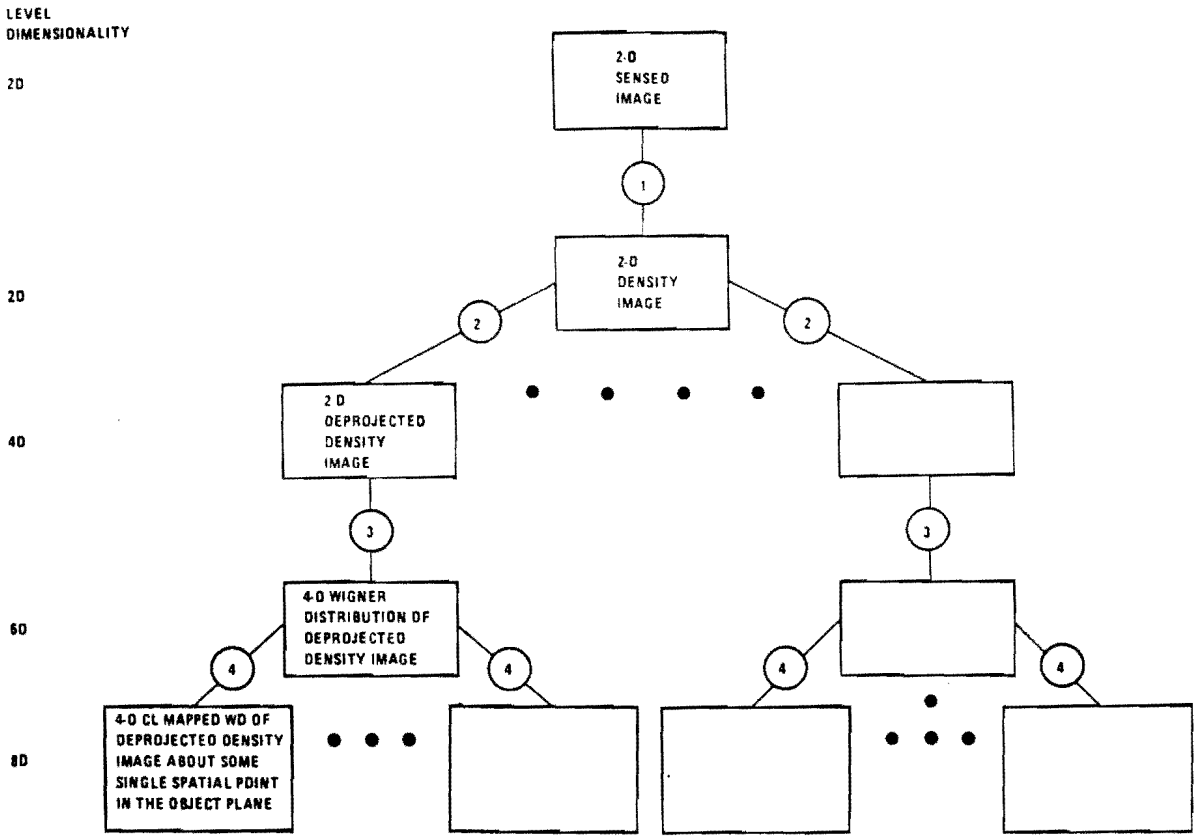
$$\begin{bmatrix} \dfrac{x}{f-z} \\ \dfrac{y}{f-z} \end{bmatrix} = \frac{1}{\Delta \cdot f}\begin{bmatrix} (x'\cos\tau + y'\sin\tau)\cos\sigma \\ (-x'\sin\tau + y'\cos\tau) \end{bmatrix} \tag{15}$$

where

$$\Delta = \cos\sigma + \frac{\sin\sigma}{f}(x'\sin\tau - y'\cos\tau).$$

It is apparent from the above relation that the inverse perspective mapping gives the object plane coordinates only to within a scale factor $(f-z)$, the distance from the viewpoint to the intersection between object plane and line of sight. In fact the distance-normalized coordinates $[x/(f-z), y/(f-z)]$ characterize *all* object planes having a common value of slant and tilt. The above relationship therefore defines a family of 'deprojective' transformations parametrized on the set of all ordered pairs (σ, τ) of slant and tilt.

The procedure for getting projection invariance is as follows: Given a density image $\ln f(x', y')$ (whose formation is constrained as previously specified), one applies to it all possible deprojective transformations, i.e., one transformation for each ordered pair (σ, τ), $0 \le \sigma < 90°$, $0 \le \tau < 360°$. This yields a 4-D representation which we will denote as $\ln f^\circ(x, y; \sigma, \tau)$. The important observation to make is that every planar form in the scene that produced the original image will appear in frontoparallel view within one of the deprojected images composing $\ln f^\circ(x, y; \sigma, \tau)$. Hence we say that $\ln f^\circ(x, y; \sigma, \tau)$ is WALS-invariant to all



STEP	TYPE OF TRANSFORMATION	NAME OF TRANSFORMATION	EFFECT OF TRANSFORMATION
1	FUNCTIONAL	LOGARITHM	CHANGES PRODUCT OF SIGNAL COMPONENTS INTO A SUM
2	GEOMETRIC	DEPROJECTION	PROVIDE INVARIANCE TO PLANAR PERSPECTIVE
3	FUNCTIONAL	WIGNER DISTRIBUTION	PROVIDE SEPARATION OF MULTIPLICATIVE SIGNAL COMPONENTS PROVIDE A LOCAL 2D TEXTURE REPRESENTATION
4	GEOMETRIC	CL CONFORMAL MAPPING	PROVIDE ROTATION, SCALE, AND POSITION INVARIANCE WITHIN EACH OBJECT PLANE

Fig. 4. The computation tree for the 8-dimensional representation $\hat{\hat{W}}_{\ln f}(z', \theta', w', \xi'; z, \theta, \sigma, \tau)$.

possible planar projections.

Now for each of the deprojected images obtained by the above procedure, that is, each $\ln f^{\circ}(x, y; \sigma_0 \cdot \tau_0)$ for some $(\sigma, \tau) = (\sigma_0, \tau_0)$, one merely computes the representation

$$\hat{W}_{\ln f}(z', \theta', w', \xi'; z, \theta)$$

that was introduced earlier. This leads to an 8-dimensional representation

$$\hat{\hat{W}}_{\ln f}(z', \theta', w', \xi'; z, \theta, \sigma, \tau)$$

which provides WALS-invariance to perspective, position in 3-D space, and also orientation and size of forms in the distance-normalized object plane coordinate system. A computation tree summarizing the steps required to get the 8-D representation is shown in Fig. 4.

7. Invariant pattern recognition

To review our progress thus far, we now have an 8-dimensional representation which is WALS-invariant to all common geometric distortions of rigid planar forms and which, given reasonable assumptions, is also invariant to non-uniform illumination. We briefly describe next how such a representation can be used to actually perform image pattern recognition.

The memory prototype pattern characterizing some arbitrary planar form is just the 4-D CL conformally mapped WD of a frontoparallel view of that form, where spatial domain mapping has been performed about some arbitrary point (see Fig. 5). If that same planar form is ever 'seen' again, its presence is detected by mathematically correlating the 4-D prototype with the previously defined 8-D image representation $\hat{W}_{\ln f}$ for all linear shifts in a 6-D hyperspace (not an 8-D space – recall from Section 5 the coupling of the spatial and spatial-frequency domains of the CL conformally mapped WD under image rotation and scaling). If the resulting 6-D correlation function exceeds threshold somewhere, then recognition is achieved. The location of the suprathreshold peak in the 6-D correlation space furthermore specifies the object position (in distance-normalized coordinates) within the object plane, the distance-normalized size, the orientation within object plane and finally, the slant and tilt of the object plane within which the planar form lies – all relative to the fixation point, size, and orientation of the pattern (in frontoparallel view) from which the matching template was formed.

8. Conclusions

We have described in this paper an approach to invariant image pattern recognition which utilizes an 8-dimensional analogical image representation. This representation decouples common dimensions of variability in the image formation process to reveal particular 4-D canonical patterns that characterize arbitrary planar gray-scale forms invariant to imaging geometry and scene illumination. Canonical patterns that are embedded within the 8-D representation can be detected by mathematically correlating the 8-D representation

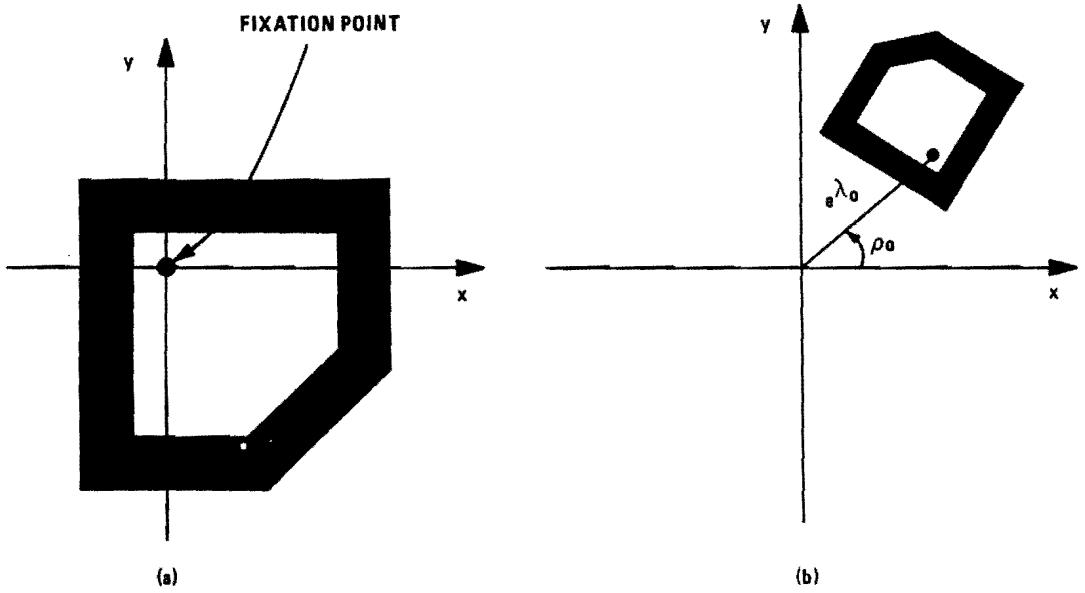


Fig. 5 (a) The 'fixation point' about which CL mapping is performed when deriving the template. (b) The identified form whose distance-normalized position in its object plane (σ_0, τ_0, z) is specified in polar coordinates by $(e^{\lambda_0}\cos\varrho_0, e^{\lambda_0}\sin\varrho_0)$, and whose distance-normalized size and orientation, relative to the template form, is denoted by k_0 and φ_0, respectively.

with corresponding 4-D canonical patterns stored in visual memory. Furthermore, once a given canonical pattern has been identified, a number of geometric attributes of the corresponding planar object in a scene are specified by the location of the suprathreshold peak in the corresponding correlation function.

Work is under way to investigate the performance and computational feasibility of the methods described in this paper. In particular, the authors will be using a discretely sampled approximation to the proposed 8-D representation. Though the amount of computation required is substantial, combinatorial explosion is not a problem since each dimension of the 8-D representation need only be encoded at a small sampling of discrete points. This follows from the fact that five of the eight dimensions correspond to finite angular axes, and the other three dimensions are logarithmic distance axes. Furthermore, referring to the computation tree for the 8-D representation (Fig. 4), it should be apparent that computation of each 4-D function found at any leaf node of the tree can be carried out independent of all other such nodes. A leaf node representation provides WALS-invariance to rotation and scaling about a single point within a single object plane. Therefore, a reasonable strategy would be to seek recognition of planar forms by sequentially shifting 'attention' from one leaf node to the next. This would eliminate the need to compute the entire 8-D representation in parallel. It could also lead to efficient strategies for handling image frame sequences by exploiting the context of earlier frames to guide an attentional mechanism.

Appendix: Definition of the WD

Assume now an arbitrary, possibly complex-valued, image function $f(x, y)$, defined over the Cartesian coordinates $x, y, -\infty < x < \infty, -\infty < y < \infty$. The WD of $f(x, y)$ is defined as:

$$W_f(x, y, u, v) = \int\int_{-\infty}^{\infty} f(x+\tfrac{1}{2}\alpha, y+\tfrac{1}{2}\beta) f^*(x-\tfrac{1}{2}\alpha, y-\tfrac{1}{2}\beta) \cdot \exp\{(\alpha u+\beta v)\}\, d\alpha\, d\beta$$

$$= \int\int_{\infty}^{\infty} R_f(x, y, \alpha, \beta)\exp\{-j(\alpha u+\beta v)\}\, d\alpha\, d\beta \quad (A1)$$

where

$$R_f(x, y, \alpha, \beta) = f(x+\tfrac{1}{2}\alpha, y+\tfrac{1}{2}\beta) f^*(x-\tfrac{1}{2}\alpha, y-\tfrac{1}{2}\beta),$$

and * denotes complex conjugation. The WD is similarly defined in terms of the Fourier transform of $f(x, y)$ as

$$W_f(x, y, u, v) = \frac{1}{4\pi^2}\int\int_{-\infty}^{\infty} F(u+\tfrac{1}{2}\eta, v+\tfrac{1}{2}\xi) F^*(u-\tfrac{1}{2}\eta, v-\tfrac{1}{2}\xi) \cdot \exp\{j(\eta x+\xi y)\}\, d\eta\, d\xi$$

$$= \frac{1}{4\pi^2}\int\int_{\infty}^{\infty} S_f(u, v, \eta, \xi)\exp\{j(\eta x+\xi y)\}\, d\eta\, d\xi \quad (A2)$$

where

$$S_f(u, v, \eta, \xi) = F(u+\tfrac{1}{2}\eta, v+\tfrac{1}{2}\xi) F^*(u-\tfrac{1}{2}\eta, v-\tfrac{1}{2}\xi).$$

References

[1] Ballard, D.H., and Brown, C.M. (1982). *Computer Vision*. Prentice-Hall, Englewood Cliffs, NJ.

[2] Bamler, R., and Glunder, H. (1983). Coherent-optical generation of the Wigner distribution function of real-valued 2D signals. In: *Proc. 10th Int. Optical Computing Conf.*, Cambridge, MA, 117–121.

[3] Brousil, J.K., and Smith, D.R. (1967). A threshold logic network for shape invariance. *IEEE Trans. on Elec. Computers* 16, 818–828.

[4] Casasent, D., and Psaltis, D. (1975). Position, rotation and scale invariant optical correlation. *Appl. Optics* 15, 1795–1799.

[5] Cavanagh, P. (1978). Size and position invariance in the visual field. *Perception* 7, 167–177.

[6] Claasen, T.A.C.M., and Mecklenbrauker, W.F.G. (1980). The Wigner distribution – a tool for time-frequency analysis. Parts I-II. *Philips J. Res.* 35, 217–250, 276–300, 372–389.

[7] Gabor, D. (1946). Theory of communication. *J. IEE (London)* 93, 429–457.

[8] Jacobson, L., and Wechsler, H. (1982). The Wigner distribution as a tool for deriving an invariant representation of 2-D images. In: *Proc. IEEE Conf. Pattern Recognition and Image Processing*, Las Vegas, NV, 218–220.

[9] Jacobson, L., and Wechsler, H. (1982). The Wigner distribution and its usefulness for 2-D image processing. In: *Proc. 6th. Int. Conf. Pattern Recognition*, Munich, West Germany.

[10] Jacobson, L., and Wechsler, H. (1982). A paradigm for invariant object recognition of brightness, optical flow and binocular disparity images. *Pattern Recognition Letters* 1, 61–68.

[11] Jacobson, L., and Wechsler, H. (1983). The composite pseudo-Wigner distribution. In: *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Processing*, Boston, MA.

[12] Kulikowski, J.J., et al. (1982). Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex. *Biol. Cybernetics* 43, 187–198.

[13] Marcelja, S. (1980). Mathematical description of the responses of simple cortical cells. *J. Opt. Soc. Am.* 70, 1297–1300.

[14] Marr, D. (1978). *Vision*. Freeman, San Francisco, CA.

[15] Oppenheim, A.V., et al. (1968). Nonlinear filtering of multiplied and convolved signals. *Proc. IEEE* 56, 1264–1291.

[16] Schenker, P.S., et al. (1981). New sensor geometries for image processing: Computer vision in the polar exponential grid. In: *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Atlanta, Georgia, 1144–1148.

[17] Schwartz, E.L. (1977). Spatial mapping in the primate sensory projection: Analytic structure and relevance to perception. *Biol. Cybernetics* 25, 181–194.

[18] Schwartz, E.L. (1981). Cortical anatomy, size invariance, and spatial frequency analysis. *Perception* 10, 455–468.

[19] Weiman, C.F.R., and Chaikin, G. (1979). Logarithmic spiral grids for image processing and display. *Comput. Graphics and Image Processing* 11, 197–226.

[20] Wigner, W. (1932). On the quantum correction for thermodynamic equilibrium. *Phys. Rev.* 40, 749–759.