Tina Memo No. 2009-005
Literature Summary for PhD thesis

The Current Status of Pairwise Geometric Histograms

S. Coupe

Last updated
5 / 6 / 2009

WWW.TINA-VISION.NET

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

Introduction

PGHs were conceived in the late 1980s as the representational basis of a semi-(biologically)-realistic neural network-based object recognition system [1]. The concept was further developed throughout the next decade [2-11]; ultimately offering a solution to the problem of recognizing contour defined projected shapes in images. The concept was constrained by an assessment of the ways in which invariances can be used to restrict the number of distinct patterns required in order to learn and recognise a shape. This assessment concluded that an important factor in pattern recognition is the stability of the representation and that requiring this stability to be encoded in the representation (i.e. as probability distributions (see below)) precluded the construction of systems which were invariant to scale and out-of-plane rotation. PGHs were accordingly designed to fulfill the representational requirements of a real-world, view-based shape recognition system, thus offering a workable degree of invariance to the following factors:

- Illumination (in so far as using defining edge information in illuminated scenes provides)
- In-plane image orientation
- Image position
- Occlusion
- Background clutter
- Edge degradation (e.g. sensor noise, lack of contrast)
- Scale (although not directly, as described below)
- Scalability (able to support discrimination across very large shape data sets)

The PGH representation is proven to provide a complete and statistically optimal representation of 2D shape, suitable for use in a learning view-based 3D object recognition system [7]. We are aware of no other published representation of projected shape which possesses such characteristics. Indeed, PGHs can be seen as a tailored solution to the representation of projected, edge-defined shape.

This document goes on to give a detailed description of the PGH representation and previous associated research. Notably, there have been 3 previous PhD theses published regarding PGH development. The first, submitted by Evans in 1994 [4], essentially introduces the concept of the PGH, analysing the stability and utility of the representation, before quantifying its use for 2D recognition and specifying a learning framework for neural network-based 3D object recognition. The second, submitted by Ashbrook in 1998 [10], extends the previous work, introducing a probabilistic Hough transform for object pose determination, methodologies for the recognition of scaled shapes and analysis of the capacity of the representation. Finally, Ashbrook deviates from the problem of view-based recognition, introducing related methodologies for encoding surface shape from (range found) polygonised depth images. The third thesis, submitted by Aherne in 1998 [11], investigated use of a multi-objective genetic algorithm to optimise PGH parameterisation.

PGH Format

PGHs are defined for linear edge segments, encoding the relative (perpendicular) positions and orientations of any other linear edge segments in a specified local image region (Figure 1). Since any curved edge can be approximated, to a desired degree of precision, by line segments [7], PGHs can encode arbitrary local image shape in a form amenable to computerised analysis. Shape recognition is performed by searching for correspondence between sets of learned representative object histograms and image sampled ones.

As indicated in Figure 2, the use of perpendicular distance and orientation means that the sample line is effectively free to shift back and forth along the reference line, within the bounds of the sample zone without changing the histogram entries. Though a second orthogonal spatial ordinate could be encoded to unambiguously represent any edge configurations (essentially amounting to a fixed edge template), this would require a fixed point of reference, which simply cannot be assumed under typical viewing conditions for edge features (in accordance with the ‘aperture problem’). Moreover, the use of relative perpendicular distances is of key importance in accounting for fragmented portions of model lines. Edge fragmentation is commonplace across imaged objects and this potential for line breakage otherwise invalidates any approach based on distances from fixed reference points. By self-referencing the

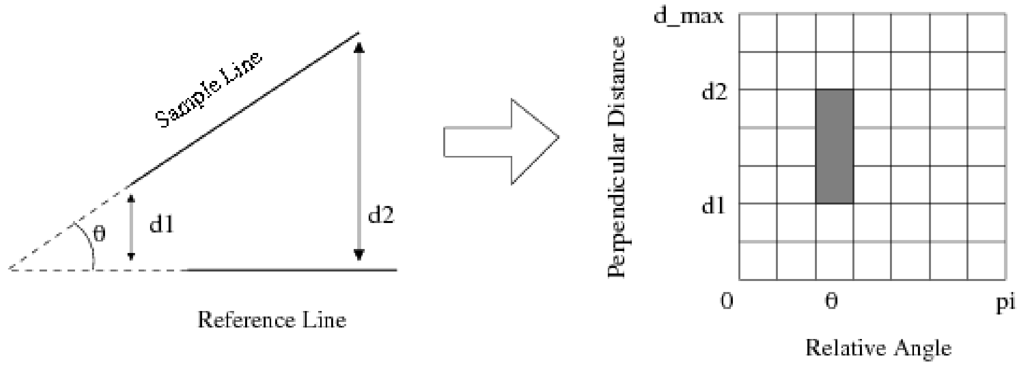


Figure 1: *The relationship between a pair of image lines can be detailed by the angle θ (defined at their point of intersection) and two perpendicular distances.*

reference line in each PGH, the relative contribution of any fragmented line segments can be reliably encoded. A 2D shape's set of defining PGHs allows for full, unambiguous shape, to be recovered [7] from a number of distinct geometric co-occurrences. This is because the data encoded is sufficient to solve for the spatial edge density distribution from a corresponding set of projection constraints. This completeness property validates the chosen representation.

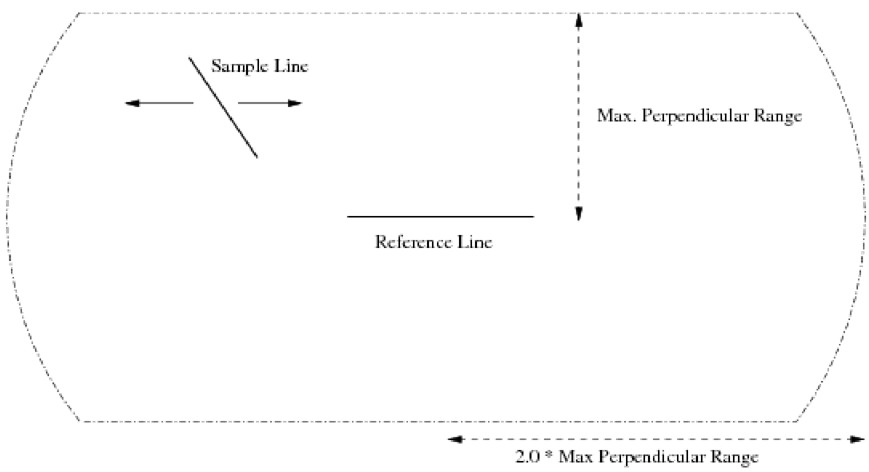


Figure 2: *According to the pairwise relationships detailed in Figure 1, sample lines are free to shift along an axis running parallel to the reference line. This is of key importance in encoding the relative contribution of fragmented portions of extended lines and extracting disjoint line segments across imaged scenes. The dotted border indicates the range through which lines are sampled for inclusion to the reference line's PGH.*

Rather than being a fixed solution to edge-pattern encoding, PGHs represent a family of related metrics, essentially offering a tradeoff between specificity and complexity. At one extreme, one-dimensional histograms can be used to measure the spread of angles without reference to relative distances. Although providing representational invariance to scale, excluding any relative distance information makes the representation too ambiguous for practical utility.

Further consideration is required regarding exactly how the relative angles and distances are encoded between image lines. A full analysis is provided in [7]. One factor is that lines may be directed according to the contrast polarity of each referenced pair, so that relative angles may be specified from 0.0 to 2π, or equivalently, $-\pi$ to π. Although contrast information may be available and pertinent to some recognition applications, this project is concerned with generic 3D shape recognition, for which contrast polarity is typically unreliable or unavailable. This is especially true for the recognition of plain surfaced 3D objects under arbitrary illumination and against arbitrary backgrounds. Fully directed line-based representations are therefore discounted from present consideration. Relative angles can instead be defined relative to the point of intersection of the 2 (extended) lines, as indicated in Figure 1. If the (extended) sample line intersects the reference line, the reference line is split at that point and the two parts are treated independently, with the resulting information being integrated in the final histogram.

The simplest form of distance-based histogram is a mirror symmetric one, which simply encodes relative angle (0-π) against unsigned perpendicular distance, thus offering invariance to any mirror symmetries across the reference line. While useful in certain constrained circumstances and offering a very compact representation, this form of histogram limits the specificity (sparseness) of the representation and generally introduces too much representational ambiguity. Alternatively, the angle between line segments (vectors directed from their extended intersection) can be extended full circle to 2π. This doubles the size of the histogram, making it more discriminatory. Equivalently, the relative angle can be fixed from 0 to π and the perpendicular distances can instead be signed according to which side of the reference line (directed away from the point of intersection) they lie.

The PGH representation can be made fully (image-plane-) rotation invariant, by essentially adding rotationally equivalent contributions from each side of the reference line to the same histogram bins. The reference line is assigned a direction pointing away from the point of intersection with each sample line, allowing a signed perpendicular axis to be defined and relative angle to be inferred, as indicated in Figure 3. Any intersecting lines are split at the point of intersection and are treated independently. While sample lines are otherwise free to displace laterally without affecting the representation (see Figure 2), this new constraint limits such displacements up to the point at which the intersection meets the reference line. This is because the polarity of part of the reference line will swap as the intersection point crosses it, so that histogram entries will switch into the opposite distance axis. Such rotational invariance cannot however be achieved without some loss of recognition specificity. To overcome this, if required, the reference line can be assigned an arbitrary direction, so that PGH entries are assigned relative to which side of the directed reference line they emanate from. The trade off here is that each image line will require analysis in each direction independently.

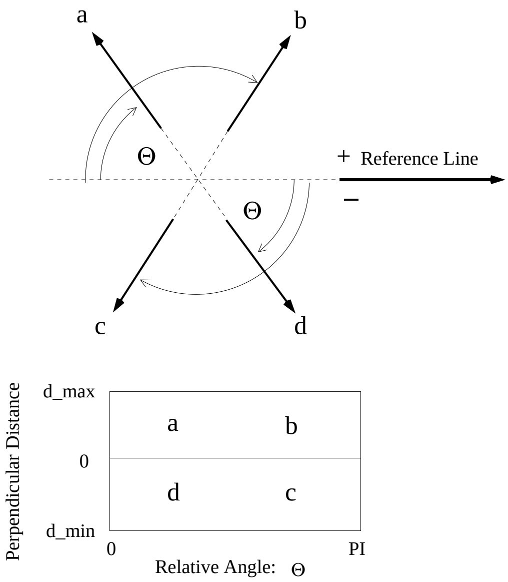


Figure 3: *For the image-plane-rotation invariant PGH format, the sample and reference lines are directed away from their point of intersection, with relative angles being entered into the histogram as indicated. If any extended sample lines intersect the reference line, the reference line is split at that point and each part is processed independently. The diagram can simply be rotated by 180 degrees to indicate the assignment of angles for reference lines pointing in the opposite direction.*

Although using a number of local non-colinear reference lines to represent a shape already enforces a global coherence constraint, the potential for low-level ambiguity can be lessened by categorising which side of the reference line each sample line is directed towards. Although this division is built into the rotationally invariant PGH, directed reference line-based PGHs can instead be duplicated, with each half corresponding to sample lines pointing to one side of the reference line as indicated in Figure 4. A continuous representational flow is maintained across the 2 representation spaces because the reference line is again split at any points of intersection, with each part being treated independently. This form of directed histogram represents the limit of information that may be conveyed by a PGH in the general case (i.e. ignoring the polarities of each PGH's sample lines).

Although previous research [4, 10] describes use of circular reference regions centered on the reference line's midpoint, it is now deemed more appropriate that these regions be extended lengthways to account for extended reference lines and possible fragmentation. A circular region of twice the diameter can therefore be defined, with a perpendicular cut-off at the prescribed pixel limit (see Figure 2).

Because PGHs are particular to single straight line segments, 2D shapes are represented redundantly by sets of overlapping localised PGHs. Such a part-based representation is however essential in terms of recognizing occluded objects, as any competent recognition system should be able. Otherwise, if only full models are used, any very localised correspondences will likely be obscured against coincidental feature matches across other complete object models and scene clutter. Although single histograms will be susceptible to the ambiguity issues just discussed (parallel displacement), fragmented portions of objects will still be represented by local clusters of PGHs, albeit reduced numbers, so that these ambiguities can be disregarded due to the recognition constraints imposed by the other reference lines.

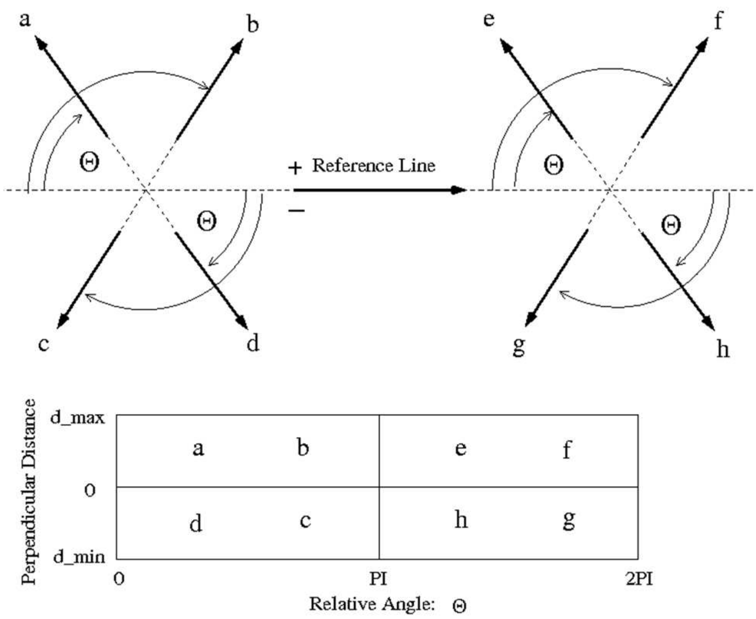


Figure 4: *For the directed reference line-based PGH format, the reference line is assigned a fixed direction so that each image line requires matching in each direction. In the simplest case, only sample lines a-d and the PGH from 0-π are required. To enhance the discriminability of the representation, the adjacent PGH ranging from π to 2π can be used to separate sample lines for which the reference line points towards their point of intersection (e-h). The reference frame can again be rotated by 180 degrees to indicate the same PGH bin assignments for reference lines pointing in the opposite direction.*

Considering that many PGHs may be required to represent an object, the issue of feature saliency was discussed in [4] with regard to streamlining the recognition process. Instead of registering a PGH for each and every edge feature, a subset of key features could instead be used, according to some saliency criterion. Unfortunately, it is very difficult to define a measure of saliency in the general case and there can never be a guarantee that any significantly reduced subset of features will be sufficient for recognition under difficult viewing conditions. Line length, for example, was considered in [4], but it was observed that for many shapes, the spatial distribution of the longest features would be very uneven.

Previous research [4] has further discussed the possibility of recognizing 2D shapes via single global histograms that accumulate the histogram evidence from each line segment. Although these histograms prove capable of recognition of fixed segmented 2D shapes, they are very sensitive to factors such as scene clutter, are devoid of any completeness properties and therefore offer very little practical utility in terms of general shape recognition.

Although the mirror and rotationally invariant PGHs offer advantages in terms of recognition processing costs, the enhanced discriminative powers of the bigger directed histograms have proved advantageous in previous research regarding 2D object recognition. The purported completeness property of the representation is also only strictly applicable to directed PGHs.

PGH Bin Entry

Beyond whichever PGH format is adopted for recognition tasks, consideration is required regarding the number of histogram bins used to sample the angle and distance axes. As explained below, this is related to the measured information available in the data, which is encoded via the degree of cross-bin blur applied to entries. At a more practical level, PGH parameter selection essentially becomes a trade-off between specificity and stability. A key consideration regarding histogram quantisation is that of application. If the computer vision task concerns very fine scale discrimination between similar shapes, then a high level of detail will require encoding. From a computational perspective, it is however desirable to limit the number of histogram bins, so that an excessive amount of memory and bin comparisons are avoided [1].

Since we are dealing with multi-view-based vision, it is also desirable to allow some form of generalisation to limit the number of views required and to account for any measurement errors such as those introduced from line quantisation or sensor error. It is also critical that the representation is stable, so that small changes in input result in similar smooth variation in histogram form. This is especially important when attempting to map out continuous shape manifolds for view interpolation. Although histogram binning inherently provides some degree of blurring, a further stage of cross-bin blurring is required to avoid discontinuous shifts in the representation due to bin quantisation. This strategy is justified by a probabilistic interpretation of histogram structure as representation of the level of uncertainty in local geometric structure (see below). Higher levels of bin blurring therefore allow the representation and recognition of deformable objects, although at the expense of decreased discrimination. Though the visual change induced by an out-of-plane rotation is a form of deformation, the recognition of more general forms of deformable object is beyond the scope of the current research.

The process of PGH bin blurring is described in [4]. First considering the angle axis, it was observed that the required degree of blurring is related via error propagation to the degree of curve linearisation imposed by the system. If the curve linearisation factor is low, so that relatively few lines are used to represent curves, then a higher degree of blurring will be required to account for shifts in relative angles due to arbitrary curve-line partitioning. A Gaussian distribution approximation was temporarily implemented. A wrap around function was also introduced, so that any blurring at extremal bins (around 0 or π) was wrapped around to the bins at the opposite side of the angle axis. Similarly, it was noted that determining the distribution of perpendicular distance bins was problematic. In this case, a simple rectangular blurring function was implemented for each entry.

In the current state of operation, bin entries, in each regard, are blurred in the form of isosceles trapeziums, i.e. rectangular central regions with sloped sides, with associated widths being specified in each case. This is evidently a simple yet effective solution to the problem. Although inherently related to histogram bin width, factors affecting choice of width of blur are described in the following section describing the stability of the representation.

The overriding interpretation for PGH construction is that the histogram is a quantitative representation of the frequency of geometric co-occurrence of edge features, with the constraint that the sum of multiple entries from a fragmented line must be directly proportional to the total entry for a corresponding un-fragmented line. Fundamentally, PGHs can be considered as the integration of equally weighted entries from individual edges (as might be supported in biological systems), but their invariance to line fragmentation allows us to construct them more rapidly (on a serial computer) from linear approximations to edge boundaries. As a final point, a system which were to construct the histograms from individual edge samples would automatically generate blurring, as described above, as a consequence of measurement noise. These histograms would also possess Poisson sampling characteristics, which forms the basis of the matching scheme now to be described.

The Bhattacharyya Match Metric

As histogram formation is defined as a counting process, normalised PGHs can be regarded as sampled conditional probability distributions (of geometric co-occurrence); with the idealised view being that each bin is equivalent to a Poisson distributed random variable [7, 12]. Standard statistical tests can therefore be drawn upon as the basis of histogram similarity for recognition purposes. The standard method for histogram comparison is the χ^2 (chi-squared) test, which is defined as a sum of squared differences operator, normalised by the expected measurement error. As discussed in detail in related work [12], the χ^2 operator is an approximation to 'Fisher's exact' test, and only typically valid across small differences in pattern space. Alternatively, the Bhattacharyya metric is proposed as being a more appropriate means by which to compare two histogram distributions because it embodies a square root transform [12]. Only with such a variance normalising transform can a measure of similarity be reliably and uniformly assessed across large distances in pattern space as a Euclidean distance.

The Bhattacharyya metric performs a dot product of the square rooted PGHs (a suitable construction for computation in neuronal tissue). We can intuitively consider this as returning the cosine of the angles between the two sampled vector spaces, i.e. 1.0 for identical (unit-normalised) histograms. Notably, maximising the Bhattacharyya metric is essentially equivalent to minimising the Matusita distance measure and offers a computationally efficient

[1] Typically, with images of approximately 0.5 million pixels, PGHs are sampled with a 50 pixel perpendicular bound in each direction, with sufficient discriminatory resolution and stability being provided by splitting each distance axis into 10 bins and the angle axis (from 0 to π) into 32 bins.

and stable means of assessing PGH similarity.

$$D_{Bhattacharyya} = \sum_i \sqrt{a_i}\sqrt{b_i}$$

However, the Bhattacharyya metric is specifically appropriate for assessing bin-to-bin similarities, for which sampling noise is the primary source of interference (cases of scene clutter and occlusion are discussed below). Since visualised PGHs will typically distort laterally with in-depth model rotation, this means that although two PGHs may be topologically similar, despite blurring and bin quantisation, their corresponding Bhattacharyya overlap scores may be relatively very low. I.e. any zero valued entries in the histograms will eliminate any non-zero values in the other PGH's corresponding bins. Thus, no distinction is made between systematically spatially distorted and completely disparate shapes. In order to get the best out of this matching process it is therefore necessary to model view induced (correlated) changes in histograms appropriately. This emphasises the significance of an interpolation procedure for PGH reconstruction throughout view-space, which stands as a key issue in current research.

Stability of the PGH Representation

One immediate potential concern with the PGH representation relates to the use of line segments to approximate curved edges. Curve linearisation is performed by iteratively subdividing the curve into linear segments, so that the ratio of perpendicular separation from the midpoint of each line to the curve to the length of the line is kept below some preset threshold. Although the linearisation factor can be increased so that there is effectively no difference between the curved and linearised representations at the levels concerned, it is desirable to limit the number of lines and PGHs used for computational tractability. It is therefore critical that the PGH encoded shape representation is largely invariant to shifts in the placement of linear sections across curves, as would be the case when linearising arbitrary fragments of curves in images. Such shifting would alter the orientation of the reference line and any related bin assignments. As discussed, the process of bin widening and entry blurring can be used to compensate for such factors, as proven experimentally in [4]. Essentially, a trial and error process is used to set a curve linearisation factor appropriate for the specific form of PGH used in the specific recognition application. There will otherwise be a limit of linearisation at which PGHs become negligibly indistinguishable, according to Bhattacharyya-based similarity. By constructing any reference models at this scale of linearisation, we are able to limit any effects of relative difference due to line segmentation, although at the potential cost of having to generate a lot of PGHs for curved regions. The bearing of these factors on determination of angle-bin blurring is further discussed in [7].

The other issues that may affect the stability of the PGH representation for recognition are shape fragmentation (e.g. due to a lack of contrast or occlusion) and sensor error. Given the proposed use of the Bhattacharyya metric for PGH similarity evaluation, we are able to directly evaluate any such effects in terms of their effect on the match score.

Edge fragmentation is very common in typical images of objects. Aside from occlusion factors, this is typically due to a lack of contrast across certain edge regions as a result of conspiring illumination or background conditions. Evans' thesis [4] quantitatively illustrates that the Bhattacharyya similarity metric (D) degrades linearly and thus very stably, relative to the proportion of remaining shape

$$D \rightarrow D(1-k)$$

Where k is the proportion of missing data, i.e. a fixed quantity. We therefore expect that occlusions and missing regions of an object will not invalidate the use of this similarity measure in competitive matching strategies. Particularly if an effort is made to robustly integrate the information from multiple histogram matches in an inclusive (i.e. non categorical) manner. The linear nature of this response can be seen once again as being related to simpler template based counting strategies, albeit with additional edge orientation sensitivity.

Scene clutter is another commonly occurring factor that the representation must be able to cope with. Such clutter can manifest itself as interference from other objects in the scene or other lighting induced artifacts such as shadows. Evans' thesis [4] indicates the resilience of PGHs to these factors by measuring the effect on recognition of adding increasing levels of randomly oriented spurious line segments across reference images. Experiments showed that matching of geometric feature distributions is theoretically robust to the presence of arbitrary spurious image lines. It was however observed that such artificial *line noise* was unrealistic and that actual line interference may be much more correlated with image structure, e.g. shadows. Because of the ambiguity of the PGH representation with regard to parallel localisation of features relative to the reference line, any spurious local features that are parallel to template lines should however present more of an adverse effect on match scores by overly weighting any specific histogram entries. The robustness of the representation should however be enough to still support reliable recognition. In later work [9], match scores (D) were shown to degrade as:

$$D \to \frac{D}{1+a}$$

Where a is the proportion of uncorrelated scene clutter. Again, for a fixed pattern of background clutter, this process is found to introduce approximately fixed reductions in match score, which do not invalidate their use during competitive matching strategies. The use of localised reference regions help to further reduce the effects of any such interference [9].

Finally, Evans thesis [4] discusses any adverse effects that sensor error may have on match reliability. Having noted that actual sensor error, i.e. image noise, would typically only affect detected edge locations by very small amounts, up to approximately 0.5 pixels, the analysis considers the cumulative effects on the detected positions and orientations of edge segments from additional factors such as line fragmentation and curve approximation (as just discussed). It is observed that these effects are proportional to line separation and inversely proportional to histogram resolution and the level of blurring used. Without such quantitative encoding of the level of uncertainty in relational geometry, it is reiterated that match scores will fall off sharply and irregularly.

Recognition across Scale

As should now be clear, because of the incorporation of distance measurements, PGHs are directly suited to the recognition of (2D) projected shape at predefined image scales. The scale invariant, one-dimensional, orientation-only encoding PGH has otherwise been shown to be too ambiguous for practical use [7]. A recognition system must however be able to recognise objects as they move closer to or further from the camera, at a range of projected scales. PGHs are partially invariant to shifts in scale and will accordingly degrade relatively smoothly and consistently, so that collections of adjacently scaled histograms can be used to offer a degree of invariance across extensions of scale space.

In the extreme, the relative scale of an imaged object may occur from a pixel to the full size of the image, ignoring partial object visibility at larger scales. For a range of such smaller scales, an object's representation will be composed of clusters of pixels, e.g. 5*5, 15*15, which may offer little or no discernible information regarding unambiguous object identity. At lower resolutions, especially for intricately structured objects, any edge-based descriptors may be vastly different from those at full resolution, as sets of edges are effectively blurred together in pixel bins. Furthermore, any errors on edge-based measurements will become more pronounced as scale reduces. For the moment, research is focused on identifying objects at a range of scales for which objects' edge-based appearances are relatively stable, e.g. for projected object diameters of 50 through to 200 pixels for relatively simple object structures. The use of discrete models through scale space does however offer the potential for adaptation of the models at each scale according to any feature visibility constraints.

Changes in the scale of a PGH encoded edge pattern are represented as uniform compression or stretching of distance entries. Ashbrook's thesis [10] and [8] provide a thorough analysis of these issues, illustrating that PGHs form smooth although globally non-linear trajectories throughout pattern space relative to image scaling. It is further shown that these shape specific trajectories can be approximately modelled by a sample of consecutively scaled PGHs. This is performed by taking a PGH at the lowest required scale, then iteratively adding new reference nodes at the maximum inferred scale difference, so that a tolerable error is continuously maintained. Any fall offs in recognition accuracy at mid-sample points can be accounted for by the noise models and associated blurring functions discussed in previous sections. Although it was initially proposed that scale be inferred directly from the scale of the nearest corresponding scaled PGH [10], it was noted that a uniform scale error would be introduced. Instead, the scale-oriented methodologies were proposed as input to a probabilistic Hough transform procedure, to support assessment of object scale.

The Probabilistic Hough Transform (PHT)

Typically, an object recognition application will return a list of hypotheses of possible model to scene correspondences. Given determination of which edge lines correspond to which objects', it is relatively trivial to infer the corresponding image position, orientation and scale of any so discovered object from line correspondences and optimisation of any transformation parameters. This allows for the position of the object to be estimated relative

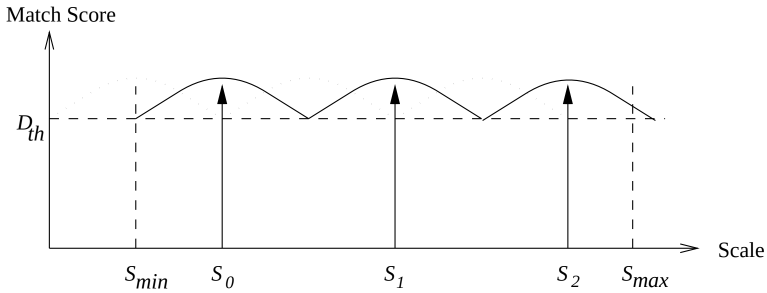


Figure 5: *The above diagram indicates how individual histograms may be combined to offer invariance across a predefined range of scales up to a tolerable degree of error.*

to the camera and, potentially, for any recognised edges to be removed from subsequent scene analysis. However, as discussed in [4], the problem is rarely this simple and there will typically be much ambiguity regarding which scene edges relate to which objects at which parameterisations. This is typically due to interference from noise, occlusion and background clutter. For these reasons, the generalised Hough Transform was proposed in Evans' thesis as a robust means to identify and parameterise well-supported model match hypotheses in noisy conditions, albeit without any initial appreciation of scaling.

The generalised Hough transform was introduced to the computer vision community in 1981 [13]. The procedure essentially operates by amassing discrete votes for each model match hypothesis, according to the hypothesised location and orientation of the considered model (at a fixed scale for now). Consistent sets of votes from related edge segments should therefore produce readily detectable peaks in representation space that can be used to indicate the most likely parameterisations of any model matches. The relative intensity of the peak will be indicative of the amount of support for that hypothesis. Spurious edge matches should otherwise distribute their votes randomly into a noise field that can be disregarded from further consideration.

In the initially considered case of 2D recognition [10], 3 parameters are required for object parameterisation; 2 for location and 1 for relative image plane orientation. Votes are cast from pairs of edge segments for particular models. An object's centroid is typically used to detail the location of the object. Although such a point can only be represented along a line running parallel to the reference line in PGH form (Figure 2), since we are using pairs of lines, supposedly from the same model at the same scale, the centroid's position can be determined at their point of intersection. The relative orientation parameter can be subsequently inferred from the associated orientation of the hypothesised model. These methods were proven able to support reliable determination of model match parameterisations without need for subsequent global model optimisation in noisy, cluttered images, although only with regard to fixed scale object recognition.

Ashbrook went on to propose the use of a Probabilistic Hough Transform (PHT) [10, 8], serving to account for any inferred errors on the estimated location and orientation parameters. Relating the techniques to those of maximum likelihood statistics, this would come to introduce a far more robust and accurate means of inferring the parameterisations of any valid model match hypotheses. Indeed, the process is essentially equivalent to performing a robust least squares fit of the projected model data. Crucially, the PHT also incorporated scale, returning the most likely inferred scales of any parameterised model matches detected in the scene.

The Probabilistic Hough Transform (PHT) process is discussed in detail in [10]. In brief, a PHT is performed for a specific model, given enough support from the PGH matching process. A (2D) location PHT is initially performed. A single entry (a conditional probability) is initially made in the PHT for each pair of scene lines that are in reasonable agreement regarding the position and scale of the model with regard to any associated measurement errors. As previously discussed, each line will constrain the model's centroid to lie along a straight line in the image, so that each pair of lines hypothesises the position of the centroid at the point of intersection of these constraint lines. The errors on the positions of any line endpoints are assumed to be Gaussian distributed, so that error propagation can be used to infer the likely error on the point of intersection. The residual errors from a real-world scene are shown to validate this process. Noting that the segmentation and scale errors are independent, the segmentation error function may then be convolved with the scale error function allowing the probabilistic entry to the location PHT to be determined.

As detailed in the previous section, sets of consecutively scaled PGHs are used to represent edge distributions through scale space. Because the errors are constant for each scaled PGH region, in effect, a rectangular bound is placed on the location of the hypothesised centroid position. Since the scale of the model must be the same relative to each line segment, the position of the centroid is further constrained to lie on an equal constraint line passing through this (skewed) rectangular region (Figure 5). The error distribution relating to line segmentation can therefore be convolved with this constraint, allowing the most likely position of the hypothesised centroid of the scaled object to be probabilistically determined for entry to the location PHT.

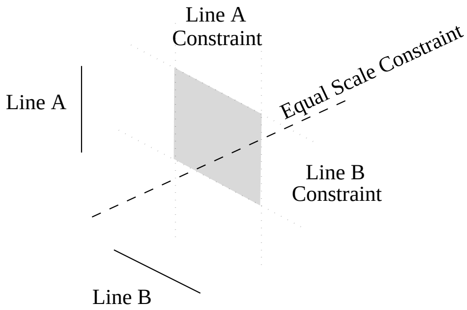


Figure 6: *The possible positions of a shape's origin relative to two reference lines are constrained to lie on the dotted line within the shaded region [8].*

Once the occurrence and location of a hypothesised model match are determined from a PHT, subsequent, single parameter PHTs are used to explicitly determine the scale and orientation of the match. Votes are cast from each scene line in accordance with the proposed model match. The scale parameter is simply determined from the perpendicular distance from the model line to the centroid, relative to the same distance in the image frame. The orientation is voted for according to orientation differences between scene lines and corresponding model lines.

The PHT has been shown to be a robust method of detecting, localising and parameterising any 2D model occurrences in cluttered images. The method is able to cope with multiple identical objects in a scene, but does however strictly require rigid 2D shape templates. Any deviations from rigidity will produce fragmented localisation peaks and sub-optimal object detection. In terms of scale-based analysis, current research is instead investigating the potential of sampling the scene at multiple scales, in a manner similar to SIFT [14], rather than having to store multiple scaled entries for each object view.

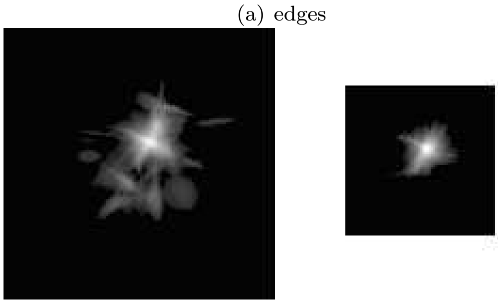
Capacity of the PGH Representation

The suitability of the PGH representation for application to real-world-oriented computer vision recognition tasks also critically depends on the capacity of the representation, in terms of its ability to differentiate distinct shapes across very large datasets. As discussed previously, the capacity of the representation will be proportional to the resolution of the histogram and the levels of blurring used. It is however, regardless, very difficult to define such a measure of capacity. It has otherwise been shown [9] that processing requirements scale no worse than linearly with the number of stored models. Recognition reliability can even be shown to improve with larger model databases because votes from any spurious scene features' PGHs become more diluted in representation space, so that any consistent model match hypotheses stand out more clearly.

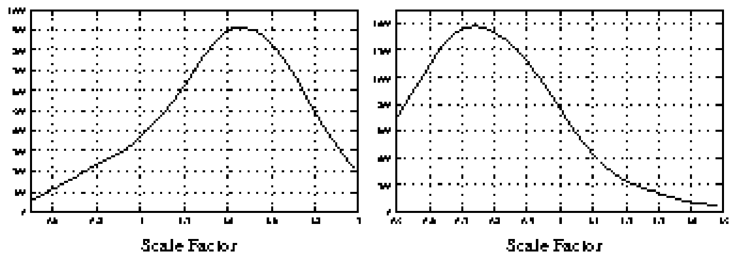
One approach to capacity estimation is to estimate the area of a typical circular hypersphere surface patch representing a distinctive view and accounting for errors, relative to the surface area of the positive quadrant of the hypersphere - according to standard PGH dimensions. The problem with this approach is that projected structural edge patterns are likely to be very correlated and the hypersphere is not likely to be uniformly populated. Accordingly, associated research has focused on identifying the 'effective dimensionality' of the representation [9, 10]. Deriving results from mismatch probability estimation curves, the effective local dimensionality of PGH encoded data is estimated at 16. This result is used to estimate that PGHs are typically capable of storing between 10^8 and 10^{13} distinct histograms. Given these numbers it is relatively easy to see why histogram matches are likely to be quite informative, even when match scores are degraded by significant quantities of clutter and occlusion.

Although the proposed fragmented pattern encoding scheme (e.g. sampling PGHs up to a 50 pixel perpendicular distance from the reference line) does mean that there are likely to be localised recognition ambiguities, these bounds on the capacity of the representation suggest that overall, multi-segment matching strategies should be robust to ambiguity in terms of application to real-world oriented learning recognition tasks.

(a) edges



(b) hough space



(c) scale estimate

Figure 7: *The above diagrams (taken from [8]) show 2D probabilistic Hough transform results for the expected locations of the 2D projected models highlighted in the corresponding images (at scales of 1.5 and 0.75). Although the peak positions (bright white highlights) are clearly identifiable, the images indicate the effects of variable sensitivity through scale space. The lower 2 diagrams show the corresponding 1D Hough transforms used to determine the scales of the inferred model matches.*

PGH-Based 3D Object Recognition (Previous Research)

It should now be clear that PGHs are a valid solution for representing 2D projected shape as the basis of a multi-view-based real-world-oriented 3D object learning recognition system. In the simplest terms, given that we now have a system able to recognise views of objects, each object of interest could be finely and uniformly sampled around its view-sphere and a nearest neighbour matching strategy could be implemented to recognise any likely model occurrences. Although this would constitute a 3D object recognition system, there are many issues that require further attention in support of a more practically viable and physiologically plausible view-based recognition system. Notably, an inordinate number of views would be required by such a system at the accuracy levels required, whereas, many views may be redundant, supporting potential application of interpolation mechanisms to optimise the required number of views.

A PGH is essentially a high-dimensional vector, with dimensionality equivalent to the number of histogram bins. As such, the stability of the representation means that clustered views of an object should describe smooth low-dimensional shape manifolds in this high-dimensional vector space. This is strictly only true for objects' aspects, for which a continuous set of visible features is maintained. Otherwise, discontinuous sets of these manifolds will be formed. By further normalising any PGHs, we can constrain any such shape manifolds to lie on the surface of (the positive quadrant of) the unit hypersphere. This processing allows objects to be represented as (possibly sets of) continuous hypersurfaces, which can be extracted and analysed as much lower dimensional shape manifolds. The problem at hand is therefore appropriate learning, modelling and recognition of these object-specific, possibly fragmented, shape manifolds.

The utility of self organising neural networks has previously been investigated to perform this PGH to object identity mapping [1, 3, 4]. Appreciating that some objects' hypersurfaces may intersect and may be susceptible to interference from noise, the proposed neural network architecture was composed to return probabilistic estimates of object recognition hypotheses, according to approximation of Bayesian a posteriori probabilities. A 3 layer network was proposed for use.

A shape representation layer is initially used by the network to automatically distribute reference units around the domain of hypersurfaces according to the distribution of PGH vectors sampled during training. Resultantly, a Voronoi tessellation of learned hypersurfaces is formed (a nearest neighbour mapping), so that the best matching node will be 'fired' upon presentation of a novel PGH vector in a 'winner-takes-all' manner. The expected output firing rate of a node is thus the probability of the node being the best description of the data. A Bhattacharyya distance metric is used as the basis of the discriminant function (in accordance with an assumption of frequency coding) and each node's weighted vector parameters are updated in a probabilistic manner, to account for those of the newly matched vector during training. This training process reduces both noise in stored patterns and the problems of missing or additional responses, via a process of 'resonance' (as per Grossberg's physiologically motivated ART network). In this way, subsequent responses of the network are robust to effects of occlusion and clutter in the viewed scene. A second layer (computationally identical to the first) is then used to learn the pattern of responses from the first, integrated over the features of the entire scene, in a manner consistent with a requirement of invariance under temporal partitioning. Again, the response is a probabilistic 'winner-takes-all'one.

The third layer of the proposed neural network is an object recognition one, which, in the simplest case, has a node for each object presented during training. Since the hypersurfaces of different objects' parts may intersect, shape units in the first shape representation layer may be representative of a number of different objects. The object recognition layer therefore essentially serves to learn the probabilities that different objects may be being viewed, given a firing pattern of nodes in the previous shape representation layer. This is achieved by connecting each shape representation node with each object node and, for each such connection, learning how frequently each object may win through labelled training data (this assumes that enough training data is learned to reflect these distributions). The estimated conditional probability that an object C is being viewed in the image data D ($P(C|D)$) can then be determined over the training data according to the ratio of the number of times the shape node j responded to an object $P(C|j)$ to the number of times that the node won $P(j|D)$. This probability value can then be assigned to the connection for subsequent output. Resultantly, presentation of a novel PGH vector to the network will result in output of a set of probabilities that each object learned in the network is being sampled, i.e.

$$P(C|D) \; = \; \sum_j P(C|j)P(j|D)$$

an approach known as probability recoding. The algorithms presented in [1] were designed to allow the robust estimation of $P(j|D)$ from the accumulated responses to a set of individual histograms. Given the outlined neural network framework, subsequent analysis [4] indicated that the accuracy of the representation will be relative to the number of nodes used in the representation layer and the number and typicality of the example PGHs used in training the network. It is suggested that the number of nodes used can be optimised (essentially by trial and error) to limit any recognition ambiguity across the objects considered [4]. The utility of the neural network system was however successfully demonstrated in this regard and performance accuracy was related to the number of nodes required by the network in order to accurately compute $P(C|D)$. It was found that the early layers of the network needed to converge to stable representations in order that later layers could be adapted to their task. In general, the full machinery of an adaptive system was therefore considered an unnecessary overhead for investigation of the remaining issues, provided that issues of occlusion and clutter were avoided during construction. Later work therefore dispensed with the network architecture and concentrated on the form of histograms required in order to unambiguously represent edge- based shape. Basic recognition experiments were run on a selection of wireframe models [4], which although indicative of PGH behaviour, as discussed, were not suited to general 3D object recognition tasks.

Related Research

The only other prominent research associated with PGHs is that of Hancock and Huet [15]. This work aimed to optimise the use of PGHs for retrieval of 2D line patterns from large databases. Their proposed representation maintained the use of relative pairwise lengths and angles and the use of the Bhattacharyya metric for histogram comparison, but opted for a scale invariant histogram format using a single global histogram for each template.

Observing that global histograms were however prone to saturation, thus making unambiguous recognition impossible in the general case, entries to the global histogram would instead be limited to the nearest (typically 5 or 6) local linear edge features. Because of this 'gating', it was proposed that scale invariance would be achieved. However, the associated research was based around matching fixed templates of black logos against plain white backgrounds and very similar aerial road photos, for which the generalised 3d object recognition issues of occlusion, background clutter and variable illumination were very limited or not encountered at all. For example, any scheme based upon selecting the nearest neighbours to an edge feature in an image will be hampered by broken edges, as so often occur in typical images of 3D objects, or the addition of background clutter, which would invalidate the solution for objects' extremal edge features against noisy backgrounds. Furthermore, by using a single histogram, the completeness properties of the PGH representation are invalidated. Whilst abstracting the recognition scheme proposed in the original work by Thacker et al. [2-11] in an attempt to offer a more compact and less specific representation, the work opens up susceptibility to ambiguity.

In the decade following the initial publications of the PGH method, numerous publications have suggested techniques which recognise objects using local information relating to orientation encoded in histograms. What this later work lacks is an approach which integrates measurement uncertainty with representation, a theory of statistical matching, or notions of information as embodied here by the idea of completeness. Analysis suggests that the original research's proposed scheme remains a perfectly valid, complete and optimal one for representing and recognizing the projected shapes of objects in arbitrary scenes via PGHs.

Conclusions

The Pairwise Geometric Histogram (PGH) representation has been shown to possess all the characteristics required for application to a multi-view-based, real-world-oriented object recognition system. PGHs are the result of an engineering effort orchestrated to that end and it is suggested that no alternative (essentially different) representation may exist under the constraints proposed. However, to date, no opportunities to actually implement a multi-view-based recognition system based upon PGHs have been forthcoming.

Although neural networks hold great promise in terms of generalised unsupervised learning, they are notoriously fiddly to train and implement and much more research is required before a finalised practical solution to the problem is realised. In current related research, although the same underlying processes are effected, a more direct and controlled form of 3D object learning is to be implemented.

[1] N. A. Thacker and J. E. W. Mayhew
Designing a Network for Context Sensitive Pattern Classification
Neural Networks, 3, 3, pp. 291-299, 1990

[2] A. C. Evans, N. A. Thacker and J. E. W. Mayhew
The Use of Geometric Histograms for Model-Based Object Recognition
Proc. BMVC93, pp. 429-438, 1993

[3] A. C. Evans, N. A. Thacker and J. E. W. Mayhew
A Practical View-Based 3D Object Recognition System
Proc. Third International Conference on Artificial Neural Networks
pp. 6-15, 1993

[4] A.C. Evans
Geometric Feature Distributions for Shape Representation and Recognition
PhD Thesis, Department of Electronic & Electrical Engineering,
The University of Sheffield, 1994

[5] P. A. Riocreux, N. A. Thacker and R. B. Yates
An Analysis of Pairwise Geometric Histograms for View-Based Object Recognition
Proc. BMVC94, pp75-84, 1994

[6] A. P. Ashbrook, N. A. Thacker and P. I. Rockett
Multiple Shape Recognition Using Pairwise Geometric Histogram Based Algorithms
Proc. IPA95, 1995

[7] N. A. Thacker, P .A. Riocreux and R. B. Yates
Assessing the Completeness Properties of Pairwise Geometric Histograms
Image and Vision Computing, 13, 5, pp. 423-429, 1995

[8] A. P. Ashbrook, N. A. Thacker, P. I. Rockett and C. I. Brown
Robust Recognition of Scaled Shapes Using Pairwise Geometric Histograms
Proc. BMVC95, pp. 503-512, 1995

[9] A. P. Ashbrook, N. A. Thacker and P. I. Rockett
Pairwise Geometric Histograms: A Scalable Solution for the Recognition of 2D Rigid Shape
Proc. SCIA95, pp. 271, 1995

[10] A. P. Ashbrook
Pairwise Geometric Histograms for Object Recognition: Developments and Analysis
PhD Thesis, Department of Electronic & Electrical Engineering, The
University of Sheffield, 1998

[11] F. J. Aherne
Towards an Automatic Object Recognition Scheme
PhD Thesis, Department of Electronic & Electrical Engineering, The
University of Sheffield, 1998

[12] N. A. Thacker
The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data
Kybernetika, 34, 4, pp. 363-368, 1997

[13] D.H. Ballard
Generalizing the Hough Transform to Detect Arbitrary Shapes
Pattern Recognition, 13, 2, pp. 111-122, 1981

[14] D. Lowe
Object Recognition from Local Scale-Invariant Features
International Conference on Computer Vision, Greece, pp. 1150-1157, 1999

[15] B. Huet and E. R. Hancock
Line Pattern Retrieval Using Relational Histograms
IEEE Transactions on Pattern Analysis and Machine Intelligence, 21,
12, pp. 1363-70, 1999