Detection and Classification of Local Primitives in Line Drawings

Naeem A. Bhatti[1], Allan Hanbury[2]

[1] Computer Vision Lab
Vienna University of Technology
Favoritenstr. 9/183-2, 1040 Vienna, Austria
bhatti@caa.tuwien.ac.at

[2] Information Retrieval Facility
Donau City Str. 1, 1220 Vienna, Austria
a.hanbury@ir-facility.org

Abstract
The local primitives found in binary images are useful in the analysis and recognition of document and patent images. In this paper, an optimum detection of end points and junction points is obtained using morphological spurring and the granulometric curve of the image. A distance based algorithm is proposed to classify the local primitives found at the detected points. The size of the local region to classify a local primitive is determined granulometrically using the average thickness of lines found in the image. The classified primitives are quantized using a variant of local binary patterns. Ground truth is created and an analysis of the classification accuracy is performed. The values for all the parameters used in the proposed method are determined granulometrically which makes it scale invariant.

1. Introduction

Binary images such as technical drawings, diagrams, flowcharts etc. are used to express inventions and scientific developments in a visual form in documents. To deal with the infringement issues in patents, the retrieval of all the patent images similar to a query image is an important and challenging task. Different geometric shapes and their spatial arrangements form the content of a patent image. The intersection and crossing point of two or more lines constitutes a junction in a line drawing image. The local pattern formed by the intersection of at least two lines in different directions at a junction point is called the junction primitive (local primitive). Local primitives are useful in the analysis and recognition of document and natural images.

A few approaches are available in the literature which detect and classify the local primitives in binary images targeting different tasks. In an attempt to reduce the length of Freeman chain codes to a minimum in image representations, Hasan et al. [4] present a heuristic approach to parse all possible directional paths found in a 2D structure containing junction points. The proposed approach does not aim to detect and classify the primitive patterns found at the junction points. Cao et al. [2] proposed line primitive extraction based on line continuation and line smoothness features, however the proposed method is not directly applicable to extract the primitives found at the end points and junction

points. For a figure completion task, Mordohai et al. [7] developed a junction inference and classification method and take into account the orthogonal and parallel fields in a tensor voting framework. The developed method takes end points and only T and L types of junctions into consideration. Deschênes et al. [3] exploit local maxima of the line curvature estimates to extract the line junctions (L, T, Y and X junctions) and line termination points in gray level images. However, the proposed method works for already extracted lines in images so that the approach becomes independent of the line detection method. Sluzek [8] presented an approach to determine the geometry of a detected junction point in a contour image by detecting its projection pattern (1D profile) in a circular window. A thresholding operation is applied to determine the number of spikes which indicates the number of arms and the position of spikes indicates the geometry of the junction. The radius of the circular window is defined empirically from 15–30 pixels which means that the approach is not scale invariant.

In this paper, we present a new algorithm for the detection and classification of local primitives found at the end points and at the junction points called end point primitives (EPPs) and junction point primitives (JPPs) respectively in binary images found in patents. As the first step, an optimum detection of the junction points (JPs) and end points (EPs) is obtained using morphological spurring and the granulometric curve of the image. The classification of EPPs and JPPs is performed at the detected points in the spurred skeleton image in a square region. The size of the square region is obtained by taking into account the average thickness of lines obtained from the granulometric curve of the original image. The developed approach analyses the foreground pixels in the square region to detect the skeletal line segments (SL-segment) and their orientation using the distances of maximal distant pixels to the central pixel of the square region. Based on the computed distances, each SL-segment is quantized to one of the 8-directional segments defined with an angular difference of 45^o in between them. Local binary patterns of 8-bit length are used to describe a detected local primitive composed of maximum 8 SL-segments. The main contributions of the proposed method are: (1) use of the granulometric curve to determine the size of square regions, (2) distance based detection and classification of SL-segments which makes it computationally efficient, (3) the developed method is robust and scale invariant.

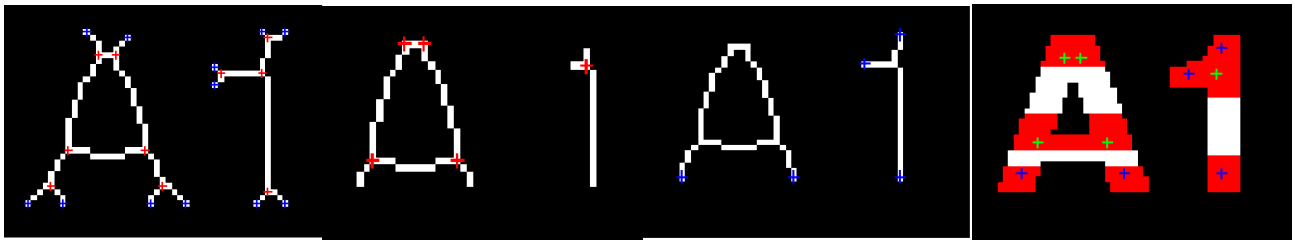
The proposed approach is described in the Section 2. The quantization of the junction primitives is described in Section 3. The experimental results are presented in Section 3. A conclusion is drawn in Section 4.

2. Proposed Method

In this section, we present the detection of junction and end points in binary images and then present the proposed method for the classification of primitives at the detected junction and end points.

2.1. Detection of Junction Points and End Points

The detection of junction points and end points is performed in a homotopic skeleton obtained by thinning using template based matching [9] which gives rise to false detections due to false skeletal lines as can be seen in the Figure 1(a). A morphological spurring operation removes the parasitic spurs in the binary image which are shorter than the number of iterations performed. An intersection of a spurred version of the skeleton image with the image containing the detection of junction points eradicates such false detections. The false detections are due to the unwanted skeletal lines introduced during the thinning process which have length approximately equal to the thickness of lines in the original image. The number of iterations of the spurring operation to get rid of these noisy skeletal



(a) (b) (c) (d)

Figure 1. (a) Noisy thinned image with JPs and EPs detected (b) GI_{JP} times spurred image with JPs detection (c) EP detection in GI_{EP} times spurred skeletal image (d) JP and EP detection in original image with ground truth marked as red.

lines can be obtained by quantifying thickness of lines. In this approach, we use the granulometric curve of original image to quantify the thickness of lines. Granulometries have been successfully used to extract the dominant width of the objects found in binary images [10]. The granulometry of an image is computed by the morphological opening of the image with a square structuring element (SE) of increasing size. Let I be the original image and $\gamma_k(I)$ be the series of morphologically opened images given by:

$$\gamma_k(I) = I \text{ O } B_k \tag{1}$$

where B_k with $k = 1, 2.....$ is the family of SEs used. The granulometry curve is obtained by plotting the volume (sum of all pixel values) of the images $\gamma_k(I)$ against B_k. A granulometric curve is often interpreted by computing the discrete derivative and plotting it against k, called the pattern spectrum:

$$PS_k(I) = \sum_{i \in I} (\gamma_k(I) - \gamma_{k+1}(I)) \tag{2}$$

where $PS_k(I)$ is the value of bin k of $PS_k(I)$. The patent binary images can contain lines of variable thickness which makes the use of the index of the global maximum as line thickness unreliable. To take into account the variability of the line thickness, we take the weighted average of the pattern spectrum:

$$Th_{lines} = \frac{\sum_k PS_k(I) * k}{\sum_k PS_k(I)} \tag{3}$$

to obtain the thickness of lines Th_{lines} which can be used to determine the number of iterations for the spurring operation to remove noisy detections and size of the square region as explained in Section 2.2. The optimum detection for JPs is obtained at the integer part of Th_{lines} taking $GI_{JP} = \lfloor Th_{lines} \rfloor$ (the values for Th_{lines} obtained are floating point) shown in Figure 1(b), however the optimum detection for EPs is obtained at a value computed as $GI_{EP} = \lceil Th_{lines}/2 \rceil$ shown in Figure 1(c). The combined optimum detection is shown in the Figure 1(d) with ground truth marked as red.

2.2. Classification of the Local Primitives at End Points and Junction points

In a skeleton image, a junction point primitive can be defined as a pattern formed by 2 or more line segments which meet or cross each other at a junction point with an angular difference $> 0^o$ and $\neq 180^o$. An end point primitive is composed of a single line segment in any direction. The proposed approach is intended to quantize all the detected primitives (JPPs and EPPs) to the primitives of skeletal line segments meeting or crossing in 8 angular directions in 360^o with 45^o angular difference to each other. In comparison to the skeleton image shown in Figure 1(a), the spurred skeleton images (SSI) obtained by GI_{JP} and GI_{EP} iterations are free from skeletonization noise as can be seen in Figure 1(b) and (c). In order to avoid false detections, the classification of primitives is performed in the spurred skeleton image (SSI) by GI_{EP} iterations instead of SSI by GI_{JP} iterations as $SSI_{GI_{JP}} \subset SSI_{GI_{EP}}$.

The number of skeletal line (SL) segments (junction branches) at a given junction point are unknown and can be detected in a defined local region. In the proposed approach, the size of the square region at a point is defined by $WS \times WS$ in pixels and $WS = 2n + 1$, where n is an integer number and the size of the square region is: 3×3, 5×5 ... for $n = 1, 2, ...$ An SL segment in the square region of interest in SSI_{GIEP} is defined as a line of foreground pixels with a predominant direction.

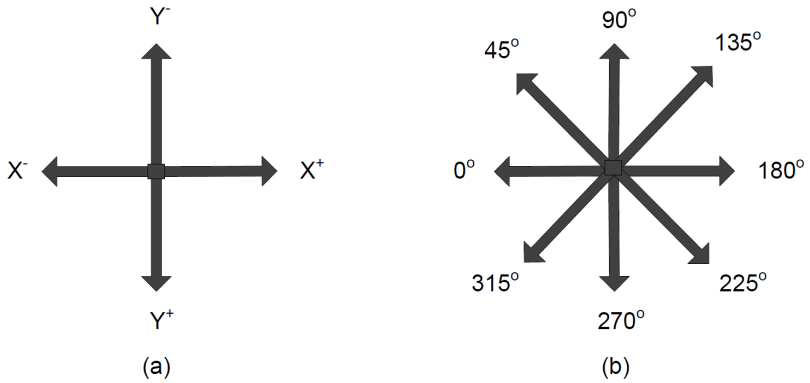


Figure 2. (a) X and Y space (b) Quantization of Skeletal Line segments to reference line segments in 8-directions.

Let Ω be the set of foreground pixels in a region of interest S and L_x, L_y be the vectors of their x and y locations in a 2-dimensional space. Let the region of interest S be a Cartesian space as shown in Figure 2(a). In this space, the coordinate locations of 4 foreground pixels $P_i \in \Omega$, where $i = 1...4$ with the maximum distances in Y^+, Y^-, X^+, X^- directions are found. The distance of each of the P_i from the central pixel of the region S is obtained as: $dx_i = L_x(P_i) - Cp_x$ and $dy_i = L_y(P_i) - Cp_y$. where dx_i and dy_i are the distances in the x, y directions respectively with $i = 1...4$. The Cp_x and Cp_y denote the central pixel coordinates in the square region S in the x and y directions respectively. Each of the computed distances dx_i and dy_i specify the distance of the SL-segment from the X and Y axes respectively. A SL-segment can be considered as a straight line with origin (Cp_x, Cp_y) and end point (dx_i, dy_i). The existing SL-segment can lie in any of the four quadrants of X-Y space and is to be quantized to one of the closest segments in the 8-directional space shown in Figure 2(b). Instead of using the slope of an SL-segment to quantize it in 8-directional space, the proposed approach uses the computed distances dx_i and dy_i which takes less computations as compared to the calculation of slope and distance for each SL segment making it computationally efficient. Another reason for using the distances in deciding about the orientation of an SL-segment in 8-directional space is to give importance to horizontal and vertical lines which exist more in number than the slanting lines in the binary line drawing images considered.

Taking into account the 8-directional space, SL-segments lying in the range of an orientation band will be quantized to the reference orientation of the specific band. To specify the range of an orientation band and to take into account the rotation variance and skeletonization noise, we introduce a quantization parameter:

$$\begin{cases} if\ (dx_i >= -q\ \&\ dx_i <= +q)\ \ dx_i = 0 \\ if\ (dy_i >= -q\ \&\ dy_i <= +q)\ \ dy_i = 0 \end{cases} \tag{4}$$

which means that an SL-segment lying at a maximum deviation of q pixels on both sides of its reference position can be detected correctly. The total range value ($-q$ to $+q$) for the quantization parameter q represents the *orientation band* of an SL-segment. The orientation band varies with the size of square region and the value of q parameter is determined by $q = \lfloor n/2 \rfloor$. This value defines the

Table 1. Classification accuracy obtained at the manual and granulometrically obtained size of square regions in ten images

Index	M_{WS}	$\%M_{Acc}$	$GIJP_{WS}$	$\%GIJP_{Acc}$	$\%D_{GIJP}$	$GIEP_{WS}$	$\%GIEP_{Acc}$	$\%D_{GIEP}$
1	3	88.3	5(5.43)	84.8	3.5	3	88.3	0
2	1	96.0	4(4.99)	87.7	8.3	3	85.3	10.7
3	5	75.4	6(6.98)	72.4	3	4	74.4	1
4	5	89.2	5(5.27)	89.2	0	3	82.8	6.4
5	4	81.7	6(6.55)	76.9	4.8	4	81.7	0
6	2	84.2	4(4.37)	79.8	4.4	3	83.6	0.6
7	5	71.2	6(6.05)	68.1	3.1	4	70.0	1.2
8	3	87.6	9(9.56)	69.4	18.2	5	83.1	4.5
9	3	80.3	6(6.14)	75.7	4.6	4	76.0	4.3
10	3	86.0	6(6.17)	78.7	7.3	4	85.9	0.1

size of the orientation band according to the size of square region around a reference orientation for $n = 2, 3, ...$ whereas for $n = 1$, q will be zero. We explore the effect of the quantization parameter in Section 3. where it is found that the best value is $q = \lfloor GI_{EP}/2 \rfloor$ at the size of the square region $WS = 2 \times GI_{EP} + 1$.

To detect and quantize the existing unknown SL segments in S, the computed distances dx, dy are analyzed. Each of the computed distances dx_i, dy_i for $P_i \in \Omega$, where $i = 1...4$ is analyzed to detect all of the existing SL segments in S in a clock-wise direction and are quantized to the defined 8-directional SL segments by analyzing the distances. This produces a vector D of length 8 with component i corresponding direction $45j^o$ and the existing SL segments in S with a 1 and the SL segments which do not exist with a 0. We explore the effect of the size of square region on the average classification accuracy of the local primitives (EPPs and JPPs) ranging from 1...10 and it is found that the optimum classification accuracy is obtained at a size of $GI_{EP} \times GI_{EP}$ which is determined using the granulometric curve of the original image. The classification results are shown and explained in the section 3.

3. Quantization of Junction Primitives by LBPs

Considering the vector D as a binary number, the 8-SL segments detected at a point are represented by 8-bits in 8-dimensional vector and represents a junction point primitive or end point primitive. The detected primitives are binary coded in an 8-directional clockwise way. Interpreting the vector D as an N-bit binary number, 2^N distinct primitives can be described in a Local Binary Patterns fashion [5]. In turn, a binary image can be represented by a 2^N-dimensional vector where each bin of the vector represents the number of occurrences of 2^N primitives in it. The junction primitives with 2-SL segments having an angular difference of exactly 180^o with each other do not fulfill the definition of a junction point primitive and would be 4 in number in 8-directional space. Eradicating these primitives, we are left with $M = 2^N - 4$ distinct JPPs in which the end point primitives composed of 1-SL segment are also included. The number of distinct EPPs is 8 which are represented by each dimension of the 8-dimensional vector D successively. Excluding the 8 EPPs, the number of distinct JPPs is $T = M - 8$.

a. Ground Truth Verification. The aim is to verify the classification accuracy of end point and junction point primitives by the proposed method. An approach used by Bowyer et al. [1] is followed

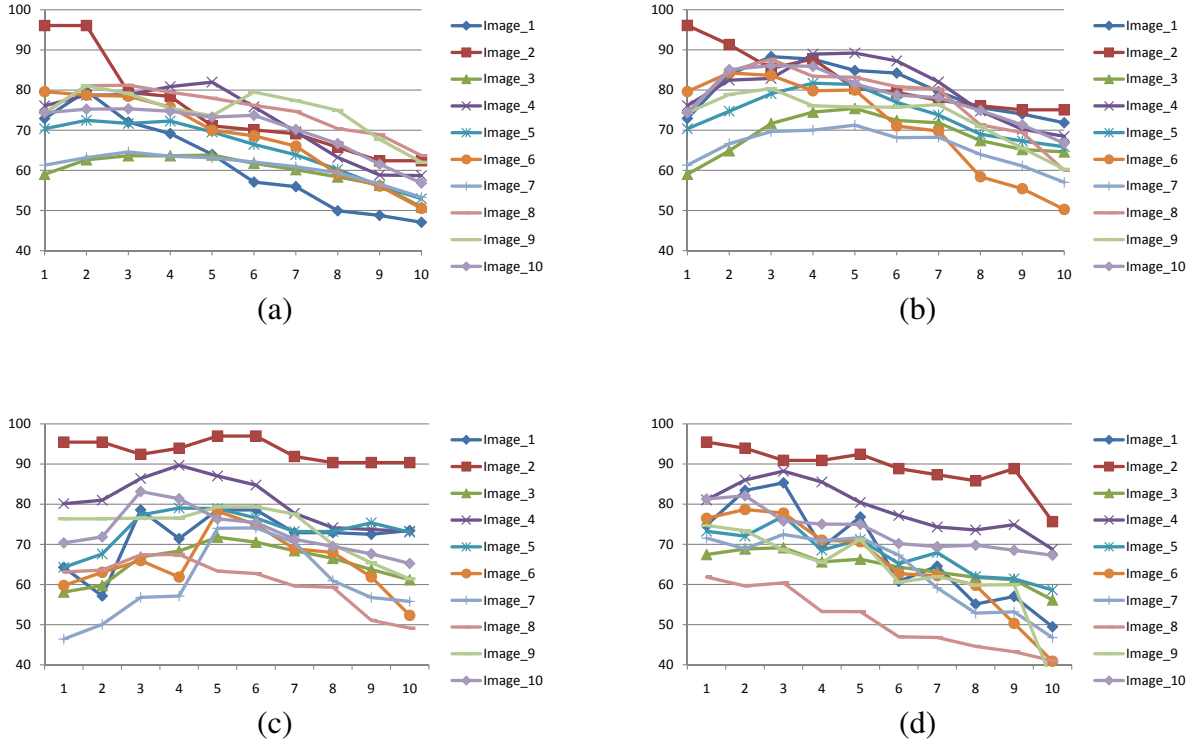


Figure 3. Effect of the quantization parameter q on average classification accuracy of all primitives with (a) $q = 0$ **(b)** $q = \lfloor n/2 \rfloor$ **(c)** $q = \lfloor n/2 \rfloor - 1$ **(d)** $q = \lfloor n/2 \rfloor + 1$

in which a ground truth based on human judges is created in ten selected images [1] and used to determine the classification accuracy of all types of primitives. During the manual creation of ground truth, each primitive is assigned a decimal value calculated based on its composition in the 8-dimensional vector following the Local Binary Patterns. The criterion to compute the classification accuracy given in [6] is used:

$$A_{cc} = 100 \times \frac{N_m}{N_g} \tag{5}$$

where A_{cc} is the classification accuracy N_g denotes the total number of primitives marked in the ground truth and N_m denotes the total matches of the classification algorithm and ground truth.

b. Discussion of Results. During the creation of ground truth in ten selected images, a primitive found at a point (end and junction point) is assigned a value mentioning its type according to the number of SL-segments it is composed of in an 8-directional space . An average classification accuracy for all existing types of primitives in an image (the selected ten images contain primitives composed of 1-4 skeletal lines) is obtained by the proposed method. On the classification accuracy of primitives, the effect of two parameters: the size of square region manually increased from 1 to 10 and the quantization parameter q (Equation 4) at 4 values is explored and shown in the Figure 3. Regarding the size of the square region defined by $WS \times WS$ (as explained in the section 2.2.), the best classification accuracy $\%M_{Acc}$ obtained at the manual size of square region M_{WS} and the classification accuracies $\%GIJP_{Acc}$, $\%GIEP_{Acc}$ at the granulometrically obtained respective sizes of the square regions $GIJP_{WS}$, $GIEP_{WS}$ are compared and their respective differences $\%D_{GIJP}$, $\%D_{GIEP}$ to $\%M_{Acc}$ are shown in the Table 1. It is found that an optimum classification accuracy is obtained at a square region

[1] http://mklab.iti.gr/content/patent-database

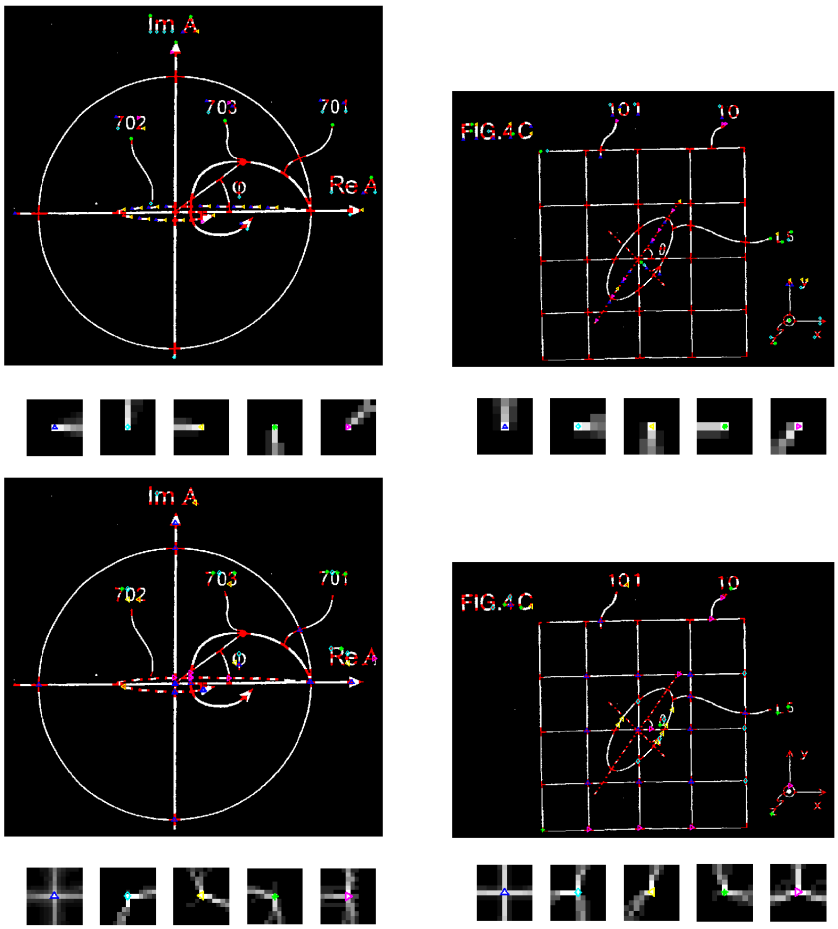


Figure 4. Two of the ten images with ground truth points and superimposed detected EPPs in above images and JPPs in lower images with five highest occurrences. The shapes of the similar classified primitives for each corresponding image are shown in small gray images in skeletonization space.

of size $GIEP_{WS}$ defined by $WS = 2 \times GI_{EP} + 1$ which makes the size of square region to detect and classify primitives scale invariant. The classification of primitives is performed in the skeletonization space where a primitive is composed of single pixel skeletal lines which are broken and noisy. The optimum size of square region $GIEP_{WS}$ becomes significant enough in the skeletonization space to detect the all SL-segments comprising a primitive. A square region of size less than the optimum makes the orientation band of the SL-segment shorter and a square region of size more than the optimum will contain longer SL-segments which are broken due to the skeletonization noise and widens the orientation band effecting the classification accuracy.

We explore the effect of quantization parameter at four different values on the classification accuracy of primitives with a manual increment in size of square region from 1 to 10 as shown in the Figure 3. An optimum classification accuracy is obtained at $q = \lfloor n/2 \rfloor$ where n is the manual increment in size of the square region is shown in Figure 3 (b). The effect on classification accuracies obtained at $q = 0$ and $q = \lfloor n/2 \rfloor - 1$ and $q = \lfloor n/2 \rfloor + 1$ are shown in the Figures 3 (a), (c) and (d) respectively, which confirms the optimality of the quantization parameter. The optimum classification accuracy is obtained at size of square region defined by $WS = 2 \times GI_{EP} + 1$ where the q-parameter is defined by $q = \lfloor GI_{EP}/2 \rfloor$ and the value GI_{EP} is obtained granulometrically which makes the proposed approach parameter free and scale invariant. The visual results for the 1-SL primitives found at end points and 2, 3, 4-SL primitives found at junction points with 5 highest occurrences superimposed on

the marked ground truth in red and their respective shapes in the skeletonization space in gray images are shown in Figure 4.

4. Conclusion

In this work a novel algorithm to detect and classify the primitives found at end points and junction points in binary images found in documents and patents is presented. The proposed approach first performs an optimum detection of end and junction points using morphological spurring and the granulometric information of the image. The primitives found at the detected points are classified based on the different line segments comprising the primitive using a distance based approach. The developed method is computationally efficient and robust. The extent of the local regions where a primitive is classified are obtained granulometrically by taking into account the average thickness of lines which makes the developed method parameter free and scale invariant. The proposed approach classifies all types of primitives in 8-directional space with high accuracy which is confirmed by manually created ground truth.

References

[1] K. Bowyer, C. Kranenburg, and S. Dougherty. Edge detector evaluation using empirical roc curves. In *CVPR*, volume 1, page 359, 1999.

[2] Ruini Cao and Chew Lim Tan. Line primitive extration by interpretation of line continuation,. In *International Conference on Control, Automation, Robotics and Vision (ICARCV)*, 2000.

[3] F. Deschênes and D. Ziou. Detection of line junctions and line terminations using curvilinear features. *Pattern Recogn. Lett.*, 21(6-7):637–649, 2000.

[4] Haswadi Hasan, Habibollah Haron, Siti Zaiton Hashim, and Fakhrul Syakirin Omar. Logical heuristic algorithm in extracting 2d structure thinned binary image into freeman chain code (fcc). In *International Visual Informatics Conference (IVIC)*, pages 770–778, 2009.

[5] M. Heikkila, M. Pietikainen, and C. Schmid. Description of interest regions with local binary patterns. *Pattern Recognition*, 42(3):425–436, March 2009.

[6] Farzin Mokhtarian and Farahnaz Mohanna. Performance evaluation of corner detectors using consistency and accuracy measures. *Computer Vision and Image Understanding*, 102(1):81 – 94, 2006.

[7] Philippos Mordohai and Gerard Medioni. Junction inference and classification for figure completion using tensor voting. In *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW)*, volume 4, pages 56–64, 2004.

[8] Andrzej Sluzek. A local algorithm for real-time junction detection in contour images. In *CAIP '01*, pages 465–472, 2001.

[9] P. Soille. *Morphological Image Analysis: Principles and Applications*. Springer, 2nd edition, 2003.

[10] L. Vincent. Fast grayscale granulometry algorithms. In Jean Serra and Pierre Soille, editors, *ISMM'94, Mathematical Morphology and its Applications to Image Processing*, pages 265–272. Kluwer, 1994.