A Method of Density Analysis for Chinese Characters

Jingwei Qu, Xiaoqing Lu, Lu Liu, Zhi Tang, and Yongtao Wang

Institute of Computer Science & Technology of Peking University, Beijing, China
{qujingwei,lvxiaoqing,liulu.pku,tangzhi,wyt}@pku.edu.cn

Abstract. Density analysis plays an important role in font design and recognition. This paper presents a method of density analysis for Chinese characters. A number of density metrics are adopted to describe the density degree of a character from both local and global perspectives, including center-to-center distance of connected components, gap between connected components, ratio of perimeter and area, connected components area ratio, and area ratio of holes. The experiment results demonstrate that the proposed method is effective in measuring the density of Chinese characters.

Keywords: density analysis, shape analysis, porosity, compactness, connected components.

1 Introduction

Density is a significant factor in the design, recognition, and other applications of Chinese characters as fonts. Three scenarios illustrate how density metrics benefit the design and application of fonts. First, the evaluation of font beauty is related to density, as shown in Fig. 1. The density adjustment of a single character results in different visual effects.

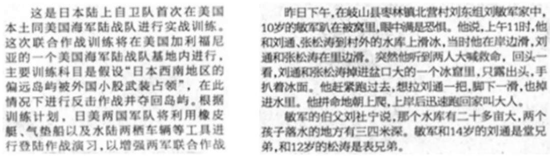
(a) (b)

Fig. 1. An example of different density of a same character in same style

Second, considering the aesthetic quality of a page, the density of different characters from the same font should not considerably vary, especially for Chinese characters whose stroke numbers differ significantly. In general, characters with many strokes tend to appear extremely dense on printed pages, whereas characters with relatively few strokes appear sparse. However, as Fig. 2 illustrates, some designs for a font without appropriate adjustments result in an unacceptable visual perceptions from the scale of a paragraph or a whole page.

Third, cross-cultural communication is becoming a staple to modern life at the present time, resulting in a large number of publications with mixed languages and characters. Density is an important factor for editors and designers to select fonts for different languages to achieve the consistent overall effect of layout.

C. Zong et al. (Eds.): NLPCC 2014, CCIS 496, pp. 54–66, 2014.
© Springer-Verlag Berlin Heidelberg 2014



这是日本陆上自卫队首次在美国本土同美国海军陆战队进行实战训练。这次联合作战训练将在美国加利福尼亚的一个美国海军陆战队基地内进行，主要训练科目是假设“日本西南地区的偏远岛屿被外国小股武装占领”，在此情况下进行反击作战并夺回岛屿。根据训练计划，日美两国军队将利用橡皮艇、气垫船以及水陆两栖车辆等工具进行登陆作战演习，以增强两军联合作战

(a)

昨日下午，在岐山县枣林镇北营村刘东组刘敏军家中，10岁的敏军趴在被窝里，眼中满是恐惧。他说，上午11时，他和刘通、张松涛到村外的水库上滑冰，当时他在岸边滑，刘通和张松涛在里边滑。突然他听到两人大喊救命，回头一看，刘通和张松涛掉进盆口大的一个冰窟里，只露出头，手扒着冰面。他赶紧跑过去，想拉刘通一把，脚下一滑，也掉进水里。他拼命地朝上爬，上岸后迅速跑回家叫大人。

敏军的伯父刘社宁说，那个水库有二十多亩大，两个孩子落水的地方有三四米深。敏军和14岁的刘通是堂兄弟，和12岁的松涛是表兄弟。

(b)

Fig. 2. Comparison between layout results of different designs for the same kind of fonts. (a) shows stronger consistency than (b) as considering more about density.

However, character density can only be estimated by human subjective judgments rather than by a quantification method in most font companies. At least three challenges exist. First, density degree is basically evaluated by human visual perception, which lacks reasonable visual models. Second, various factors influence shape density, thus requiring deep-seated shape analysis. Lastly, no evaluation method or common dataset is authorized to judge the validity of the density metric.

This paper describes a feasible method for calculating the density of Chinese characters. We regard Chinese characters as shapes with multiple connected components and analyze several factors influencing the density of the characters, such as center-to-center distance of connected components, gap between connected components, ratio of perimeter and area, scale of connected components, and scale of holes.

The rest of this paper is organized as follows. Section 2 gives an overview of related work. Section 3 introduces some basic concepts. Section 4 details the proposed method about the measurement of density. Section 5 shows the experiment results. Finally, Section 6 draws the conclusions of this study and recommends future work.

2 Related Work

Research on character density has so far been few. However, density analysis has been involved in some studies on pattern recognition and image retrieval, which use similar concepts such as porosity and compactness. Some of the related studies are briefly examined in this section.

Song et al. [2] explicitly defined porosity as the scale of the gap between connected components based on mathematical morphology [7]. A closing operation was adopted on the object shape using a circular element with a different radius, analyzing the scale of the gap with a multi-scale. Finally, porosity was described by the diameter of the minimum circle connecting all connected components.

Bribiesca [3] used compactness to separately analyze 2D and 3D shapes. An object in the 3D domain, for example, has compactness that is classically related to the enclosing surface area and volume and can be measured by the ratio $(\text{area}^3)/(\text{volume}^2)$.

Liu et al. [8] proposed a concept of foreground pixel density, which is simply defined as the ratio of foreground pixels in an object with respect to the foreground pixels in the entire image. Liu et al. [10] employed the density distribution feature, which is defined as a matrix whose components determine the relative density of the foreground pixel in each small region.

Liu et al. [11] presented point–line distance distribution (PLDD) to detect arbitrary triangles, regular polygons, and circles, based on the common geometric property that the in-center of the shape is equidistant to the tangential lines of the contour points. Unlike SC [12] and IDSC [13], PLDD directly presented image features that are very similar to density by distance distribution.

The above analysis shows that the density of shapes with multiple connected components has not been particularly and extensively studied as a separate issue. Studies are merely involved in some image research, and the methods have limitations. For example, porosity, defined as the gap between connected components, merely employs the minimum gap, which is a simple description of the density. Compactness only qualitatively represents dispersion of a shape without considering the quantification of distance between pixels. Thus the description of the density is not comprehensive.

3 The Basic Concepts

Before discussing the algorithm proposed in this study, some basic concepts about density are necessary to be introduced.

3.1 Convex Hull

The convex hull of a shape is defined as its initial envelope, as shown in Fig. 3.

(a) (b)

Fig. 3. The effect of convex hull: (a) a Chinese character (b) corresponding convex hull

3.2 Hole

A hole of a connected component is a closed region whose gray level is obviously different from others. In a binary image, a hole is a black region in a white connected component, as shown in Fig. 4.

(a) (b)

Fig. 4. The effect of filling holes: (a) a Chinese character (b) the effect after filling holes

3.3 Morphologic Closing Operation

A shape's morphologic closing operation corresponds to (1) the morphologic dilation and (2) the morphologic erosion with Eq. (1) – (6) illustrated in Gonzalez et al [26]:

$$A \cdot B = (A \oplus B) \ominus B \tag{1}$$

$$A \oplus B = \left\{ z \middle| \left(\hat{B}\right)_z \cap A \neq \emptyset \right\} \tag{2}$$

$$A \ominus B = \{z|(B)_z \cap A^c = \emptyset\} \tag{3}$$

$$A^c = \{w|w \notin A\} \tag{4}$$

$$\hat{B} = \{w|w = -b, b \in B\} \tag{5}$$

$$(B)_z = \{c|c = b + z, b \in B\} \tag{6}$$

4 Density Analysis

Density is one of the important properties for shape analysis. A character, in this paper, is regarded as a shape composed of multiple connected components. By analyzing the elements and structures in characters, many factors contributing to the density are found. We select five dominant metrics for a more in-depth analysis and effective description of density: (1) center-to-center distance of connected components, (2) gap between connected components, (3) ratio of perimeter and area (4) connected components area ratio, and (5) area ratio of holes. More details about them and their relationships are provided in following subsections.

4.1 Center-to-Center Distance of Connected Components (CCDCC)

To achieve an accurate description of density, the distribution of all connected components in a Chinese character is taken into account firstly. The distances between geometric centers of connected components can be used as an important feature. Hence, we propose the connected components center distance (CCDCC) to represent the layout information in a character. The CCDCC is defined as following:

$$d_{ij} = \sqrt{\left(\bar{x}_i - \bar{x}_j\right)^2 + \left(y_i - \bar{y}_j\right)^2} \tag{7}$$

$$D = \left\{d_{12}, d_{13}, \cdots, d_{ij}, \cdots, d_{n(n-1)}\right\}, |D| = C(n, 2) \tag{8}$$

where $(\bar{x}_i, \bar{y}_i)$ and $\left(\bar{x}_j, \bar{y}_j\right)$ are the i^{th} and j^{th} connected component's center, respectively, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots$, n, $i \neq j$, D is the set of center distances, and n is the number of connected components of a Chinese character.

Referring to the perspective that Liu et al. [8] and Bai et al. [9] proposed graph structure for image matching, we represent each of the connected components with its centroid. Then, a Chinese character is converted to a graph, all connected components mapping to the graph nodes. CCDCC is based on distances between each pair of the nodes.

The CCDCC metric, obviously, is sensitive to the areas of connected components. When the area is small, it performs well in revealing the distance between two connected components. On the contrary, if the regions of connected components enlarge in size, the distance between them is unsuitable to be described only with CCDCC. To eliminate the negative influence, we assign different weights to the CCDCCs, a minimal weight w_{min} given to the maximal center distance d_{max}, a maximal weight w_{max} given to the minimal center distance d_{min}, and an average weight w_{avg} given to the average center distance d_{avg}. The weighted CCDCC, WCCDCC, is obtained in Eq. (10),

$$d_{avg} = \frac{\sum d_{ij} - d_{max} - d_{min}}{|D| - 2} \tag{9}$$

$$WCCDCC = \frac{d_{max} * w_{min} + d_{min} * w_{max} + d_{avg} * w_{avg}}{L_{diagonal}} \tag{10}$$

where $\sum d_{ij}$ is the sum of all center distances and $L_{diagonal}$ is the diagonal length of bounding box of a character. Two examples are shown in Fig. 5.

(a) (b)

Fig. 5. An example of WCCDCC (a) WCCDCC = 0.1348 (b) WCCDCC = 0.5142

4.2 Gap between Connected Components (GCC)

Another feature of component distribution is the gaps between connected components. We improve the porosity metric [2] based on gaps for more accurate density measurement at local level.

We adopt a closing operation to measure the gap between two connected components. The closing operation uses a predefined structure element (eg. circle) at different scales (radius) to analyze the size of the gap. A morphologic closing operation is adopted to maintain the original shape compared to the morphologic dilation, and it is useful for gap junction between connected components. The size of gap is obtained when the circle radius are large enough to merge the two connected components.

Therefore, the minimum diameter of the circle that leads to joining two connected components together reveals the size of gap objectively, and then, this diameter is defined as the size of gap (gap_{ij}) between two connected components. A set GAP_SET composed of all the gap_{ij} is obtained in Eq. (11) and Eq. (12) based on the operation of morphologic closing predefined in Section 3.3.

$$gap_{ij} = 2 * min\{s \,|\, Compn(A_{ij} \cdot B_s) = 1\} \tag{11}$$

$$GAP_{SET} = \{gap_{12}, gap_{13}, \cdots, gap_{ij}, \cdots, gap_{n(n-1)}\}, |G| = C(n, 2) \tag{12}$$

where A_{ij} is a target shape composed of the i^{th} and j^{th} connected components, B_s is a predefined structure element (such as a circle) at scale s, Compn(X) returns the number of connected components in the shape X, gap_{ij} is the gap between the i^{th} and j^{th}

connected component, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n$, $i \neq j$, and n is the number of connected components in a Chinese character. Finally, an average value of gaps is calculated with Eq. (13).

$$gap_{avg} = \frac{\sum gap_{ij} - gap_{most} * n_{most} - gap_{max} * n_{max} - gap_{min} * n_{min}}{|G| - n_{most} - n_{max} - n_{min}} \tag{13}$$

where gap_{most}, gap_{max}, gap_{min} and gap_{avg} respectively denotes the mode gap value, maximum value, minimum value and the average in the *GAP_SET*, n_{most}, n_{max}, n_{min} and n_{avg} denote the numbers of corresponding gaps individually.

4.3 Ratio of Perimeter and Area (RPA)

The appearance of the outline also affects character density. Bribiesca [3] proposed a simple and effective measurement for the compactness of 2D and 3D shapes. We employ compactness to depict the density of Chinese characters in reference to the equation in 2D domain. Compactness for a 2D shape associates the perimeter with the area of the shape and can be measured by the ratio (perimeter^2) divided by area. The contact perimeter corresponds to the sum of the lengths of the segments shared by two adjacent pixels. The relation between the contact and the shape perimeters is represented by Eq. (14).

$$2P_c + P = 4LN \tag{14}$$

where P_C is the contact perimeter, P is the perimeter of the shape, L is the length of a side of the pixel, and N is the number of pixels. In Fig. 6, each square represents an image pixel, the solid lines denote the perimeter, and the dashed ones correspond to the contact perimeter, $N = 9$, $L = 1$, $P = 12$, $P_C = 12$.

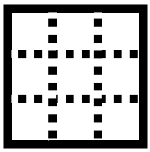


Fig. 6. A shape composed of 9 pixels

RPA of a Chinese character is calculated using Eq. (15).

$$C = \frac{n - P/4}{n - \sqrt{n}} \tag{15}$$

RPA can be applied not only to Chinese characters with multiple connected components, but also to those with single connected component. RPA reveals the distribution of all pixels of a character and partially describes its density. A higher pixel distribution results in a more concentrated image, as shown in Fig. 7.

(a) (b)

Fig. 7. An example of the RPA (a) RPA = 0.9922 (b) RPA = 0.9514

4.4 Connected Components Area Ratio (CCAR)

The scale of connected components refers to the fullness by which the strokes of a Chinese character cover the entire region. A character is the foreground, and the convex hull is the background. The ratio of the foreground area, CCAR, is used for calculating the scale of the connected components. As shown in Eq. (16), CCAR is calculated as the ratio of the sum area of all connected components in a Chinese character with respect to the area of the region enclosed by the entire convex hull:

$$CCAR = \frac{s_1 + s_2 + \cdots + s_i + \cdots + s_n}{S_{con}} \tag{16}$$

where s_i and S_{con} denote the ith area of the connected component and that of the convex hull in a Chinese character, respectively, i = 1, 2, …, n, and n is the number of connected components.

The higher the CCAR of a Chinese character, the denser is the image being perceived by human visual perception, as shown in Fig. 8.

(a) (b)

Fig. 8. An example of CCAR: (a) CCAR = 0.2979 and (b) CCAR = 0.4122

CCAR can be applied to both characters with multiple connected components and characters with single connected components.

4.5 Area Ratio of Holes (ARH)

The structure of strokes in a Chinese character also affects the human vision about density. The hole, predefined in Section 3.2, is one of the most important and relatively dominant parts of a character. We consider the scale of the holes and define it as ARH based on the algorithm in Soffer et al. [4] with Eq. (17)

$$ARH = \frac{S_{hole}}{S_{com}} \tag{17}$$

where S_{hole} is the area of all holes in a Chinese character, and S_{com} is the area of the character whose holes have been filled.

Though CCAR can reveal density differences to a certain extent, it is not always the best factor. For example, the CCAR of Fig. 9(a) and (b) are nearly the same, but (b) looks denser than (a).

(a) (b)

Fig. 9. An example of CCAR and ARH: (a) CCAR = 0.4171, ARH = 0.2195 and (b)CCAR = 0.4138, ARH = 0.0915

ARH reveals the relative size of the holes in a Chinese character. When ARH is high, the hole is large with respect to the character, and the character will seem loose and expanded to human vision. In the same example, Fig. 9 (a) has larger holes than (b), hence (b) appears denser.

4.6 Overall Density Metric

With the above extracted five density features, a density descriptor can be directly obtained by the five-dimension vector $V_{density}$ expressed as Eq. (18).

$$V_{density} = (WCCDCC, GAP, RPA, CCAR, ARH) \tag{18}$$

However, the five factors have varying importance. CCAR and RPA are only effective in local regions. They are used as weights to adjust other global factors. More specifically, to quantify the density of a character, we give different gaps obtain in section 4.2 with different weights. CCAR and RPA are adopted to calculate weights in Eq. (19) to enhance the gap value, because they have closer relationships with gaps,

$$GAP = gap_{most} * w_{most} + gap_{max} * w_{max} + gap_{min} * w_{min} + gap_{avg} * w_{avg} \tag{19}$$

where gap_{most}, gap_{max}, gap_{min} and gap_{avg} respectively denotes the mode gap value, maximum value, minimum value and the average; w_{most}, w_{max}, w_{min} and w_{avg} are their weights; GAP is the enhanced gap value, and an example is shown in Fig. 10.

We thus propose a combined form density metric. While GAP is calculated based on CCAR and RPA, the final density is the combination of the other three metrics, namely, ARH, WCCDCC, and GAP. We normalize GAP by $L_{diagonal}$ as Eq. (20) to derive GAP_{norm}. The overall density metric is defined in Eq. (21).

$$GAP_{norm} = \frac{GAP}{L_{diagonal}} \tag{20}$$

$$Density = (WCCDCC + GAP_{norm} + ARH) / 3 \tag{21}$$

(a) (b)

Fig. 10. An example of GAP (a) GAP = 22 (b) GAP = 64

5 Experiments

5.1 Setup and Datasets

The experiment environment included an Intel Core i3 (3.30 GHz) processor with 4.00 G RAM and Windows 8 operating system, as well as Matlab 2013a.

We access four Chinese character databases, with each corresponding to a typeface, namely, Song, Fangsong, boldface, and regular script. Each database also contains 6,715 Chinese characters that are 128 x 128 in size.

5.2 Experiment Results

Comparison Results. We calculate the five features and the density of each Chinese character in the four databases and compare them. The samples of comparison results are shown in Table 1.

Table 1. The samples of comparison on the five features of four different typefaces

	Song	Fangsong	Boldface	Regular Script
(a)				
WCCDCC	0	0	0	0
GAP	2.5	2.5	2.5	2.5
RPA	0.8943	0.9306	0.9766	0.9619
CCAR	0.5399	0.9159	0.9690	0.8449
ARH	0	0	0	0
Density	0.0072	0.0079	0.0075	0.0073
(b)				
WCCDCC	0.3154	0.2938	0.3329	0.3150
GAP	8	8	8	12
RPA	0.9333	0.9111	0.9632	0.9353
CCAR	0.2874	0.2459	0.4413	0.3062
ARH	0.2595	0.2207	0.1381	0.1305
Density	0.2077	0.1883	0.1734	0.1734
(c)				
WCCDCC	0.3443	0.3017	0.3457	0.3295
GAP	57.7750	21.7273	59.1364	77.8000
RPA	0.9125	0.8940	0.9469	0.9238
CCAR	0.2845	0.2605	0.3913	0.3035
ARH	0	0	0	0
Density	0.2312	0.1456	0.2375	0.2751

Table 1. (*continued*)

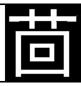
(d)				
WCCDCC	0.2765	0.2792	0.2871	0.2895
GAP	23.7000	21.7000	26.4000	19.8000
RPA	0.9456	0.9352	0.9701	0.9564
CCAR	0.2888	0.2704	0.4118	0.3695
ARH	0.5340	0.5122	0.4197	0.4067
Density	0.3185	0.3113	0.2907	0.2764
(e)				
WCCDCC	0.3290	0.3129	0.3347	0.3096
GAP	49.0909	17.2364	55.2000	47
RPA	0.9211	0.8932	0.9491	0.9252
CCAR	0.3349	0.2620	0.4601	0.3263
ARH	0	0	0	0
Density	0.2715	0.1438	0.2346	0.143

Table 1 shows that a Chinese character has similar and different feature values relative to the typeface. For (a), when the number of the character strokes is small and the structure is simple, the density has no obvious differences between the four typical typefaces. However, from (b) to (e), as the number of strokes increases and the structure becomes more complex, the difference in the typefaces become more evident. Therefore, for most Chinese characters, density varies with different typefaces.

Clustering Results. We use the K-means algorithm to cluster Chinese characters with boldface based on the feature vectors comprising the five features, namely, WCCDCC, GAP_{norm}, RPA, CCAR and ARH. The experiment results demonstrate, that at K = 3, the differences between clusters are the most evident. Tables 2 and 3 summarize the experiment results and the corresponding character examples for all clusters.

Table 2. The clustering results based on the feature vectors

	WCCDCC	GAP_{norm}	RPA	CCAR	ARH
(a)	0.3383	0.4808	0.9519	0.4136	0.0732
(b)	0.2945	0.0725	0.9606	0.4276	0.1314
(c)	0.3161	0.2580	0.9543	0.4160	0.0776

Table 3. Some examples in each cluster

(a)		二			三			为			心	
		六			办			示			兰	
(b)		一			上			片			腱	
		髑			廛			幢			篝	
(c)		业			代			泯			粉	
		参			悴			臂			战	

As shown in Tables 2 and 3, the strokes of characters in cluster (a) are simple and are less compact compared with the other two clusters. Characters in cluster (b) are composed of two types: (1) single connected components characters and (2) characters with more complex strokes. Accordingly, the values of the features are either maximum or minimum among the three clusters and are most dense. The characters in cluster (c) do not have very evident differences. Some look dense, while others look less compact. The values of their features are the middle among the three clusters.

6 Conclusion and Future Work

This study focuses on the density of Chinese characters. We propose five metrics, namely, CCDCC, GAP, RPA, CCAR and ARH, to describe the density from the pixel, outline, and component levels. By combining the five metrics, both local and global information of the connected components are considered. The experiment results demonstrate that our method can not only discriminate density differences between different typefaces for the same Chinese character, but also depict density differences between characters with the same typeface. Future studies can explore more density factors to describe density and incorporate related knowledge such as psychology. In addition, we look forward to applying the density analysis to benefit more research fields.

Acknowledgments. This work is supported by National Natural Science Foundation of China under Grant 61300061 and Beijing Natural Science Foundation under Grant 4132033.

References

1. Schomaker, L., de Leau, E., Vuurpijl, L.: Using Pen-Based Outlines for Object-Based Annotation and Image-Based Queries. In: Huijsmans, D.P., Smeulders, A.W.M. (eds.) VISUAL 1999. LNCS, vol. 1614, pp. 585–592. Springer, Heidelberg (1999)
2. Song, J.-G., Lu, X.-Q., et al.: Envelope Extraction for Composite Shapes for Shape Retrieval. In: Proceedings of the Internationnal Conference on Pattern Recognition, Tsukuba Science City, Japan, pp. 1932–1935 (2012)
3. Bribiesca, E.: An Easy Measure of Compactness for 2D and 3D Shapes. Pattern Recognition 41, 543–554 (2008)
4. Soffer, A., Samet, H.: Negative Shape Features for Image Databases Consisting of Geographic Symbols. In: 3rd International Workshop on Visual Form Capri (1997)
5. Rosenfeld, A., Pfaltz, J.L.: Distance Functions on Digital Pictures. Pattern Recognition 1(1), 33–61 (1968)
6. Haralick, R.M.: A Measure for Circularity of Digital Figures. IEEE Trans. Syst. Man Cybernet 4(4), 394–396 (1997)
7. Hu, R.-X.: A Perceptually Motivated Morphological Strategy for Shape Retrieval. In: Huang, D.-S., Gan, Y., Gupta, P., Gromiha, M.M. (eds.) ICIC 2011. LNCS, vol. 6839, pp. 105–111. Springer, Heidelberg (2012)
8. Liu, L., Lu, Y., et al.: Near-duplicate Document Image Matching: a Graphical Perspective. Pattern Recognition (2013)
9. Bai, X., Yang, X.W., et al.: Learning Context-sensitive Shape Similarity by Graph Transduction. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(5) (2010)
10. Liu, H., Feng, S.Q., Zha, H.B., et al.: Document Image Retrieval Based on Density Distribution Feature and Key Block Feature. In: Proceedings of International Conference on Document Analysis and Recognition, pp. 1040–1044 (2005)
11. Liu, H.M., Wang, Z.H.: PLDD: Point-lines Distance Distribution for Detection of Arbitrary Triangles, Regular Polygons and Circles. J. Vis. Commun. Image R (2013)
12. Belongie, S., Malik, J., Puzicha, J.: Shape Matching and Object Recognition Using Shape Context. IEEE Transactions Pattern Analysis and Machine Intelligence 24(4), 509–522 (2002)
13. Ling, H.B., Jacobs, D.W.: Shape Classification Using the Inner-distance. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(2), 286–299 (2007)
14. Guru, D.S., Nagendraswamy, H.S.: Symbolic Representation of Two-dimensional Shapes. Pattern Recognition 28, 144–155 (2007)
15. Siddiqi, K., Shokoufandeh, A., Dickinson, S.J., Zucker, S.W.: Shock Graphs and Shape Matching. International Journal of Computer Vision 35(1), 13–32 (1999)
16. Sebastian, T.B., Klein, P.N., Kimia, B.B.: Recognition of Shapes by Editing Shock Graphs. In: IEEE International Conference on Computer Vision, pp. 755–762 (2001)
17. Shi, C., Xiao, J., Jia, W., Xu, C.: Automatic generation of chinese character based on human vision and prior knowledge of calligraphy. In: Zhou, M., Zhou, G., Zhao, D., Liu, Q., Zou, L., et al. (eds.) NLPCC 2012. CCIS, vol. 333, pp. 23–33. Springer, Heidelberg (2012)
18. Marr, D., Vision, A.: A Computational Investigation into the Human Representation and Processing of Visual Information. Freeman and Company, WH San Francisco (1982)
19. Lai, P.K., Pong, M.C., Yeung, D.Y.: Chinese Glyph Generation Using Character Composition and Beauty Evaluation Metrics. In: International Conference on Computer Processing of Oriental Languages (ICCPOL), Honolulu, Hawaii, pp. 92–99 (1995)

20. Yu, K., Wu, J., Zhuang, Y.: Skeleton-based Recognition of Chinese Calligraphic Character Image. In: Huang, Y.-M.R., Xu, C., Cheng, K.-S., Yang, J.-F.K., Swamy, M.N.S., Li, S., Ding, J.-W. (eds.) PCM 2008. LNCS, vol. 5353, pp. 228–237. Springer, Heidelberg (2008)
21. Bribiesca, E.: Measuring 2-D Shape Compactness Using the Contact Perimeter. Computers & Mathematics with Applications 33(11), 1–9 (1997)
22. Bribiesca, E.: A Measure of Compactness for 3D Shapes. Computers & Mathematics with Applications 40(10), 1275–1284 (2000)
23. Bogaert, J., Rousseau, R., Van Hecke, P., et al.: Alternative Area-perimeter Ratios for Measurement of 2D Shape Compactness of Habitats. Applied Mathematics and Computation 111(1), 71–85 (2000)
24. Del Bimbo, A., Pala, P.: Image Indexing Using Shape-based Visual Features. In: Proceedings of the 13th International Conference on Pattern Recognition, vol. 3, pp. 351–355. IEEE (1996)
25. Herrmann, P., Schlageter, G.: Retrieval of Document Images Using Layout Knowledge. In: Proceedings of the Second International Conference on Document Analysis and Recognition, pp. 537–540. IEEE (1993)
26. Gonzalez, R.C., Woods, R.E., Eddins, S.L.: Digital Image Processing Using Matlab. Princeton Hall Pearson Education Inc., New Jersey (2004)