Rough Geometry and Its Applications in Character Recognition

Xiaodong Yue and Duoqian Miao

Department of Computer Science & Technology, Tongji University Shanghai, 201804, PR China yswantfly@yahoo.com.cn miaoduoqian@163.com

Abstract. The absolutely abstract and accurate geometric elements defined in Euclidean geometry always have lengths or sizes in reality. While the figures in the real world should be viewed as the approximate descriptions of traditional geometric elements at the rougher granular level. How can we generate and recognize the geometric features of the configurations in the novel space? Motivated by this question, rough geometry is proposed as the result of applying the rough set theory to the traditional geometry. In the new theory, the geometric configuration can be constructed by its upper approximation at different levels of granularity and the properties of the rough geometric elements should offer us a new perspective to observe the figures. In this paper, we focus on the foundation of the theory and try to observe the topologic features of the approximate configuration at multiple granular levels in rough space. Then we also attempt to apply the research results to the problems in different areas for novel solutions, such as the applications of rough geometry in the traditional geometric problem (the question whether there exists a convex shape with two distinct equichordal points) and the recognition work with principal curves. Finally, we will describe the questions induced from our exploratory research and discuss the future work.

Keywords: Rough sets, rough geometry, geometric invariants, equichordal points, principal curves.

1 Introduction

As the belief of Zadeh that there exits the information granularities in many areas with different forms (see, *e.g.*, [27,28]), human problem solving always involves the ability of perception, abstraction, representation and understanding of real world problems at different levels of granularity. Through the research on the basic issues of "Granular Computing" (see, *e.g.*, [8,9,10,11]), the complex or uncertain problem is tried to be transformed among the different granular spaces for seeking the proper solutions (see, *e.g.*, [21,22]).

Like the research focus of granular computing theory, the multilevel methods of analyzing the image content in both spatial and frequency domain have been the hotspots in the research area of pattern recognition (see, *e.g.*, [1,7]). The

analysis of the objects contained in the digital images from different levels or views can not only improve the recognition efficiency, but also help to understand the images' content more effectively, and this recognition process may be more coincident with the human intelligence. In recent years, rough set theory has been applied to the research of image analysis and processing as a granular computing model to provide the hierarchical methods (see, *e.g.*, [2,15,16]).

Especially, some popular issues in the research area of image analysis, such as the recognition methods of off-line handwriting characters, usually pay attention to the geometric features of the objects (see, *e.g.*, [14,29]), but the geometric elements analyzed in practical applications often have different representations from those in the traditional geometry. In other words, the figures in the real world should be viewed as the approximate descriptions of traditional geometric elements at rougher granular level. How can we generate and recognize the geometric features of the configurations in the novel space?

Euclidean geometry has been the most popular measurement tool in the past thousands years, but the absolutely abstract and accurate geometric elements defined in Euclidean geometry always have lengths or sizes in reality. For example, the points and straight lines in a digital image are sized rather than abstract and the sizes of these geometric elements depend on the resolution of this digital image. In another view, the Euclidean points lying in the region of a pixel are indiscernible and equivalent in the digital image, then the partition of the Euclidean points can be obtained from this equivalence relation and the pixels can be considered as the equivalence classes of the Euclidean points. So rough geometry is proposed as the result of applying the granular computing scheme and rough set theory to the traditional geometry [12].

Fig. 1.1 shows a Euclidean straight line in digital space, and Fig. 1.2 indicates that the representation of this line in the digital space is formed by the pixel set that covers it. The pixel set can also be viewed as the union of the equivalence classes which have nonempty intersections with the straight line, namely, the pixel set constructs the upper approximation of the Euclidean straight line under the partitions in digital space.

As a matter of fact, most geometric configurations in the real world are the approximate representations of Euclidean geometric elements, and the geometric properties of these approximate configurations are often different from those of the corresponding Euclidean geometric elements. The new properties of the



Fig. 1. A Euclidean Straight Line in Digital Space

geometric configuration constructed by upper approximation at different granular levels should offer us a new perspective to observe the geometric elements.

The rest of this paper is organized as follows. Section 2 focuses on the theoretic foundation of rough geometry and based on that, we try to study the geometric properties and observe the variation of the topologic features of the approximate configuration at different granular levels in rough space. Then in Section 3, we also attempt to apply the research results to the problems in different areas for novel solutions, such as the applications of rough geometry in the traditional geometric problem (the question whether there exists a convex shape with two distinct equichordal points) and the recognition work with principal curves will be mainly introduced. Finally, we will describe the questions induced from our exploratory research and discuss the future work in Section 4.

2 Rough Geometry

2.1 Rough Sets

The rough geometric space is constructed based on the foundation of rough set theory (see, e.g., [17,18,19]), so the notions of rough sets which are related to rough geometry will be firstly recalled as shown below.

Information system IS is a pair S = (U, A), where U is a non-empty finite set of objects and A is a non-empty finite set of attributes. Each subset of attributes $B \subseteq A$ determines a binary indiscernibility relation $IND_S(B) : IND_S(B) \Rightarrow$ $\{(x, y) \in U \times U | \forall a \in B, a(x) = a(y)\}$. Because the binary relation $IND_S(B)$ is reflexive, symmetric and transitive, $IND_S(B)$ is an equivalence relation and defines a partition on the universe U.

Given an equivalence relation R, the equivalence class of the element $x \in U$ under the partition induced from R consists of all objects $y \in U$ such that xRy, which is defined as $[x]_R = \{y | y \in U \land xRy\}$, the objects in an equivalence class are indiscernible from each other. The equivalence class of any object x under the partition formed by the indiscernibility relation $IND_S(B)$ $(B \subseteq A)$ is usually denoted by $[x]_B$ for simplicity.

In an information system S = (U, A), the subset of objects $X \subseteq U$ can be described by the attributes subset $B \subseteq A$, i.e. X can be approximated using only the information contained in B by constructing the B-lower and B-upper approximations, which are denoted by $\underline{B}X$ and $\overline{B}X$ respectively, where $\underline{B}X = \{x | [x]_B \subseteq X\}$ and $\overline{B}X = \{x | [x]_B \cap X \neq \emptyset\}$. $\underline{B}X$ and $\overline{B}X$ can also be viewed as the intension and extension of the concept represented by X.

2.2 Rough Space and Rough Configuration

In rough geometry, rough set theory is combined with the traditional geometry and the figures are represented by the equivalence classes and set approximation. These approximate representations are the new geometric elements of distinct features in different space. In the following paragraphs, the fundamental concepts about the new space and configuration approximation defined in rough geometry will be introduced.

Let φ is a mapping from a real number field \Re to a subset of real numbers \Re' . For $\forall x, y \in \Re$, if $x \leq y \Rightarrow \varphi(x) \leq \varphi(y)$, φ is called a monotone increasing mapping from \Re to \Re' , the binary relation induced from the mapping $\varphi : E_{\varphi} = \{(x, y) \in \Re \times \Re | \varphi(x) = \varphi(y)\}$ is obviously an equivalence relation on \Re . Let \Re^n is an n-dimension Euclidean space, the indiscernibility relation " \approx_{φ} " defined by E_{φ} in \Re^n is as follows: two n-dimension points (x_1, \ldots, x_n) and (y_1, \ldots, y_n) are indiscernible iff $(x_1, y_1) \in E_{\varphi}, (x_2, y_2) \in E_{\varphi}, \ldots, (x_n, y_n) \in E_{\varphi}$, i.e. $(x_1, \ldots, x_n) \approx_{\varphi} (y_1, \ldots, y_n) \Leftrightarrow (x_i, y_i) \in E_{\varphi}, (i = 1, 2, \ldots, n)$. The relation \approx_{φ} is reflexive, symmetric and transitive, thus it is an equivalence relation and determines a partition in \Re^n , in other words, the points of an n-dimension space can be divided into the corresponding regions through the mapping φ .

Definition 1. Rough Space. Let \approx_{φ} is an equivalence relation in n-dimension space \Re^n , the set of all equivalence classes formed by the partition induced from \approx_{φ} , which denoted by $\Re^n / \approx_{\varphi}$ is called a rough space.

Definition 2. Rough Point. Let $\Re^n / \approx_{\varphi}$ is a rough space induced from the equivalence relation \approx_{φ} in an n-dimension space \Re^n , an element in the rough space, i.e. an equivalence class under the partition of relation \approx_{φ} is called a rough point in the space $\Re^n / \approx_{\varphi}$.

Definition 3. Rough Configuration. Let $\Re^n / \approx_{\varphi}$ is a rough space induced from the equivalence relation \approx_{φ} in an n-dimension space \Re^n , a subset of rough points in the rough space is called a rough configuration in the space $\Re^n / \approx_{\varphi}$.



Fig. 2. Upper Approximations of A Straight Line in Different Rough Spaces

Definition 4. Rough Subspace. Let \Re^n / \approx_1 and \Re^n / \approx_2 are two rough spaces, if every rough point of \Re^n / \approx_1 is contained in a rough point of \Re^n / \approx_2 , the space \Re^n / \approx_1 is called a rough subspace of \Re^n / \approx_2 , i.e. \Re^n / \approx_1 is a rough subspace induced from \Re^n / \approx_2 , and \Re^n / \approx_2 is called the upper space of \Re^n / \approx_1 . The relation between such two spaces can be denoted by $\Re^n / \approx_1 \leq \Re^n / \approx_2$.

Furthermore, a special rough space $\Re^n / =$ is defined as $\{\{(x_1, \ldots, x_n)\}|(x_1, \ldots, x_n) \in \Re^n\}$, i.e. every equivalence class in the space $\Re^n / =$ contains only one Euclidean point. If the difference between one Euclidean point and the set containing only this point is ignored, the rough space $\Re^n / =$ is just the n-dimension

Euclidean space \Re^n . Obviously, $\Re^n / =$ is a rough subspace of any rough space $\Re^n / \approx_{\varphi}$.

Definition 5. Transformation of Upper and Lower Approximation. Let S is a rough configuration in space \Re^n / \approx_1 , the upper approximation and lower approximation of S in another rough space \Re^n / \approx_2 are denoted by $U_{\approx_2}(S)$ and $L_{\approx_2}(S)$ respectively, which defined as $U_{\approx_2}(S) = \{P \in \Re^n / \approx_2 | P \cap \overline{S} \neq \emptyset\}$, $L_{\approx_2}(S) = \{P \in \Re^n / \approx_2 | P \subseteq \overline{S}\}$, where \overline{S} is the union of all elements of configuration S in space \Re^n / \approx_1 . Fig. 2 indicates the transformation of upper approximation of a Euclidean straight line in rough spaces \Re^n / \approx_1 and \Re^n / \approx_2 , where $\Re^n / \approx_1 \leq \Re^n / \approx_2$.

2.3 Geometric Invariants of Upper Approximation Transformation

German mathematician Felix Klein had given the most general definition of "geometry" as the research on geometric invariants under a group of transformations, such as the projective geometry focuses on the geometric invariants under the projective transformation. In rough geometry, we will pay attention to the geometric invariants of configurations under upper approximation in rough spaces, i.e. the invariant properties of approximate configuration at different granular levels.

In the following paragraphs, some geometric invariants of upper approximation transformation in rough spaces will be introduced and the concepts such as "rough line segment", "rough convex" and "equal rough line segments" will be further represented in the proofs of the corresponding properties. The monotone mapping $\varphi_{\delta} : R \to Z, x \to \lfloor x/\delta \rfloor$, where $\delta \in R^+$ will be adopted to construct the rough spaces in this section, R, R^+ and Z are the real number field, positive real number field and integer field respectively, and $\lfloor x \rfloor$ is the operator for returning the greatest integer less than or equal to real number x. (see Fig. 3).



Fig. 3. Monotone Mapping $\varphi_{\delta} : R \to Z$

The mapping φ_{δ} divides R into a queue of intervals such as $\ldots [-2\delta, -\delta)$, $[-\delta, 0)$, $[0, \delta)$, $[\delta, 2\delta) \ldots$ The rough space $\Re^n / \approx_{\varphi_{\delta}}$ induced from the indiscernibility relation $\approx_{\varphi_{\delta}}$, which is defined by the equivalence relation $E_{\varphi_{\delta}}$, is denoted by $SPACE(\delta)$. We can see that $SPACE(\delta_1) \leq SPACE(\delta_2)$ iff δ_2 is a multiple of δ_1 , $\Re^n / =$ is denoted by SPACE(0), and 0 is considered to be divisible by any real number. Furthermore, the upper approximation of configuration S in $SPACE(\delta)$ will be denoted as $U_{\delta}(S)$ rather than $U_{\approx_{\varphi_{\delta}}}(S)$ for simplicity.

Theorem 1. Let $SPACE(\delta) \leq SPACE(\delta_1)$, a rough point in $SPACE(\delta)$ is still a rough point through the transformation of upper approximation from $SPACE(\delta)$ to $SPACE(\delta_1)$. *Proof.* Let $SPACE(\delta) \leq SPACE(\delta_1)$, P is a rough point in $SPACE(\delta)$, because δ_1 is divisible by δ , so the upper approximation of this point $U_{\delta_1}(P)$ is still a rough point in space $SPACE(\delta_1)$, see Fig. 4.



Fig. 4. Rough Points in SPACE(0.5) and SPACE(1.5)

Theorem 2. Let $SPACE(\delta) \leq SPACE(\delta_1)$, the relative location of two rough points in rough space $SPACE(\delta)$ is invariant in space $SPACE(\delta_1)$ through the proper transformation of upper approximation.

Proof. Let $SPACE(\delta) \leq SPACE(\delta_1)$, $P(i_1, j_1)$ and $Q(i_2, j_2)$ are two rough points in $SPACE(\delta)$, and $i_1 \leq i_2$ (maybe $i_1 \geq i_2$ or $j_1 \leq j_2$ or $j_1 \geq j_2$), from Def. 2, the upper approximations of the two points in $SPACE(\delta_1)$, i.e. $U_{\delta_1}(P) = (l_1, t_1)$ and $U_{\delta_1}(Q) = (l_2, t_2)$, also have $l_1 \leq l_2$ ($l_1 \geq l_2$ or $t_1 \leq t_2$ or $t_1 \geq t_2$).

Definition 6. Rough Line Segment. Let S be a subset of rough points in $SPACE(\delta)$, if there exists at least one Euclidean line segment l such that $S = U_{\delta}(l)$, S is called a rough line segment in $SPACE(\delta)$.

Theorem 3. Let $SPACE(\delta) \leq SPACE(\delta_1)$, a rough line segment in $SPACE(\delta)$ is still a rough line segment through the proper transformation of upper approximation from $SPACE(\delta)$ to $SPACE(\delta_1)$.

Proof. Let $SPACE(\delta) \leq SPACE(\delta_1)$, S is a line segment in space $SPACE(\delta)$, $S' = U_{\delta_1}(S)$ is the upper approximation of S in $SPACE(\delta_1)$, according to Def. 1 and Def. 6, the Euclidean line segment $l: S = U_{\delta}(l)$ also has $S' = U_{\delta_1}(l)$, so there must exist a Euclidean line segment whose upper approximation is S' in $SPACE(\delta_1)$.

Definition 7. Rough Convexity

Let S be a rough configuration in $SPACE(\delta)$, if $(i_0, j_0) \in S$, let $S_T(i_0) = max\{j|(i_0, j) \in S\}$, $S_R(j_0) = max\{i|(i, j_0) \in S\}$, $S_B(i_0) = min\{j|(i_0, j) \in S\}$, $S_L(j_0) = min\{i|(i, j_0) \in S\}$. As illustrated in Fig. 5.1, $i_0 = 3$, $j_0 = 2$, $S_T(i_0) = 3$, $S_B(i_0) = 0$, $S_R(j_0) = 6$, $S_L(j_0) = 1$.

A rough configuration S is called upper-convex, if for any pair of points $P = (i_P, j_P)$ and $Q = (i_Q, j_Q)$ in S (suppose $i_P \leq i_Q$), there exists at least one line

segment L in $SPACE(\delta)$ passing through P and Q such that $S_T(i) \ge L_T(i)$ for any $i_P \le i \le i_Q$, see Fig. 5.2.

As shown in Fig. 5.3, a rough configuration S will be called lower-convex, if for any pair of points $P = (i_P, j_P)$ and $Q = (i_Q, j_Q)$ in S (suppose $i_P \leq i_Q$), there exists at least one line segment L in $SPACE(\delta)$ passing through P and Qthat satisfies $S_B(i) \leq L_B(i)$ for any $i_P \leq i \leq i_Q$.

Similarly, a rough configuration S is called right-convex, if for any pair of points $P = (i_P, j_P)$ and $Q = (i_Q, j_Q)$ in S (suppose $j_P \leq j_Q$), there exists at least one line segment L in $SPACE(\delta)$ passing through P and Q that satisfies $S_R(j) \geq L_R(j)$ for any $j_P \leq j \leq j_Q$, and left-convex can be defined as follows, if for any pair of points $P = (i_P, j_P)$ and $Q = (i_Q, j_Q)$ in S (suppose $j_P \leq j_Q$), there exists at least one line segment L in $SPACE(\delta)$ passing through P and Qthat satisfies $S_L(j) \leq L_L(j)$ for any $j_P \leq j \leq j_Q$.



Fig. 5. Convex Configurations in SPACE(0.5)

Theorem 4. A upper-convex rough configuration in a rough subspace $SPACE(\delta)$ is still upper-convex in the upper space $SPACE(\delta_1)$ of $SPACE(\delta)$ through the proper transformation of upper approximation, the similar results can be obtained on lower-convex, left-convex and right-convex configurations.

Proof. Let S is an upper-convex configuration in $SPACE(\delta)$, from Def. 7, for any pair of points $P = (i_P, j_P)$ and $Q = (i_Q, j_Q)$ in S (suppose $i_P \leq i_Q$), there exists a line segment L : PQ in $SPACE(\delta)$ such that $S_T(i) \geq L_T(i)$, $i_P \leq i \leq i_Q$. Thus for any $i, (i_P \leq i \leq i_Q)$, there must exist a point $Z = (i, j_Z)$ satisfying

$$Z \in S \land j_Z \ge L_T(i) \tag{1}$$

Let $SPACE(\delta) \leq SPACE(\delta_1), S' = U_{\delta_1}(S)$ is the upper approximation of S in $SPACE(\delta_1)$, we can also get $L' = U_{\delta_1}(L), P' = U_{\delta_1}(P), Q' = U_{\delta_1}(Q)$ and $Z' = U_{\delta_1}(Z)$. Let $\varphi_{\delta'}(x) : x \to \lfloor x/\delta' \rfloor$ is the mapping from $SPACE(\delta)$ to $SPACE(\delta_1)$, in which

$$\delta' = \begin{cases} \delta_1/\delta, & if \quad \delta > 0\\ \delta_1, & if \quad \delta = 0 \end{cases}$$
(2)

Because $\varphi_{\delta'}(x)$ is monotone increasing, we have

$$i_{P'} = \varphi_{\delta'}(i_P) \le i' = \varphi_{\delta'}(i) \le i_{Q'} = \varphi_{\delta'}(i_Q) \tag{3}$$

$$j_{Z'} = \varphi_{\delta'}(j_Z) \ge \varphi_{\delta'}(L_T(i)) = L'_T(i') \tag{4}$$

From (3) and (4), we know that for any i', $i_{P'} \leq i' \leq i_{Q'}$, there is a point $Z' = (i', j_{Z'})$ such that

$$Z' \in S' \land j_{Z'} \ge L'_{T}(i') \tag{5}$$

As mentioned above, it can be inferred that for the pair of points P' and Q' in S', there exists a line segment L': P'Q' in $SPACE(\delta_1)$ such that $S'_T(i') \geq L'_T(i')$ for any i' ($i_{P'} \leq i' \leq i_{Q'}$). So S' is still upper-convex in $SPACE(\delta_1)$. Through the similar proof, the results on lower-convex, left-convex and right-convex configuration can also be obtained.

Definition 8. Configuration Intersection. Let S_1 and S_2 are two rough configurations in $SPACE(\delta)$, if $S_1 \cap S_2 \neq \emptyset$, it is considered that S_1 and S_2 intersect in $SPACE(\delta)$. Let P is a rough point and S is a rough configuration in $SPACE(\delta)$, if $P \in S$, it is considered that S passes the point P.

Theorem 5. Let $SPACE(\delta) \leq SPACE(\delta_1)$, S_1 and S_2 are two rough configurations in $SPACE(\delta)$, if S_1 and S_2 intersect in $SPACE(\delta)$, $U_{\delta_1}(S_1)$ and $U_{\delta_1}(S_2)$ must intersect in the upper space $SPACE(\delta_1)$. Suppose S is a rough configuration in $SPACE(\delta)$, if S passes the rough point P in $SPACE(\delta)$, $U_{\delta_1}(S)$ must pass the point $U_{\delta_1}(P)$ in $SPACE(\delta_1)$.

Theorem 6. Let $SPACE(\delta) \leq SPACE(\delta_1)$, configurations S_1 and S_2 are symmetric about the origin point or coordinate axis in $SPACE(\delta)$, their upper approximations $U_{\delta_1}(S_1)$ and $U_{\delta_1}(S_2)$ in $SPACE(\delta_1)$ are still symmetric about the same element through the proper transformation.

Definition 9. Rough Distance. Let $P, Q \in SPACE(\delta)$ are two rough points, the upper approximation of the distance between P and Q is defined as

$$U_{\delta}(P,Q) = \varphi_{\delta}(max\{|AB||A \in P, B \in Q\}) \times \delta + \delta$$
(6)

and the lower approximation of the distance is correspondingly defined as

$$L_{\delta}(P,Q) = \varphi_{\delta}(\min\{|AB||A \in P, B \in Q\}) \times \delta \tag{7}$$

i.e. the distance approximation between the two points in rough space is constructed from the maximum and the minimum distance between the Euclidean points contained in the rough points. The closed interval $d_{\delta}(P,Q) = [L_{\delta}(P,Q), U_{\delta}(P,Q)]$ is considered as the roughness range of the distance between points P and Q in $SPACE(\delta)$.

See Fig. 6, the maximal and the minimal Euclidean distance between two rough points P = (1,1) and Q = (5,3) in SPACE(0.5) are the distances between Euclidean pints C, D and E, F respectively.

 $\begin{array}{l} U_{0.5}(P,Q) = \varphi_{0.5}(max\{|AB||A \in P, B \in Q\}) \times 0.5 + 0.5 = \varphi_{0.5}(|CD|) \times 0.5 + \\ 0.5 = 3, \ L_{0.5}(P,Q) = \varphi_{0.5}(min\{|AB||A \in P, B \in Q\}) \times 0.5 = \varphi_{0.5}(|EF|) \times 0.5 = \\ 1.5, \ \text{thus the roughness range of the distance between } P \ \text{and} \ Q \ \text{is} \ d_{0.5}(P,Q) = \\ [1.5, 3]. \end{array}$



Fig. 6. Rough Distance Between P and Q in SPACE(0.5)

Definition 10. Equal Distance. Let *T* is an index set, given a set of rough point pairs $\{(P_t, Q_t) | t \in T\}$ in $SPACE(\delta)$, the distance of a pair (P_t, Q_t) is the rough distance between P_t and Q_t , the distances of the pairs in the set $\{(P_t, Q_t) | t \in T\}$ are considered equal, iff $\bigcap_{t \in T} d_{\delta}(P_t, Q_t) \neq \emptyset$.

Theorem 7. Let $SPACE(\delta) \leq SPACE(\delta_1)$, T is an index set, $\{(P_t, Q_t) | t \in T\}$ is a set of rough point pairs in $SPACE(\delta)$, $U_{\delta_1}(P_t)$ and $U_{\delta_1}(Q_t)$ are the upper approximations of P_t and Q_t $(t \in T)$ in $SPACE(\delta_1)$, if the distances of all rough point pairs in $\{(P_t, Q_t) | t \in T\}$ are equal, the distances of the pairs in set $\{(U_{\delta_1}(P_t), U_{\delta_1}(Q_t)) | t \in T\}$ are still equal in $SPACE(\delta_1)$ through the proper transformation.

Proof. Let $SPACE(\delta) \leq SPACE(\delta_1), (P, Q)$ is a rough point pair in $SPACE(\delta)$, we have

$$max\{|A, B||A \in P, B \in Q\} \le max\{|A, B||A \in U_{\delta_1}(P), B \in U_{\delta_1}(Q)\}$$
(8)

$$\min\{|A, B| | A \in P, B \in Q\} \ge \min\{|A, B| | A \in U_{\delta_1}(P), B \in U_{\delta_1}(Q)\}$$
(9)

Because $\varphi_{\delta}(x) = \lfloor x/\delta \rfloor$ is monotonically increasing,

$$\varphi_{\delta}(\max\{|A, B||A \in P, B \in Q\}) \times \delta + \delta \leq \\\varphi_{\delta}(\max\{|A, B||A \in U_{\delta_{1}}(P), B \in U_{\delta_{1}}(Q)\}) \times \delta + \delta$$
(10)

As $SPACE(\delta) \leq SPACE(\delta_1)$ and δ_1 is divisible by δ , thus

$$\varphi_{\delta}(\max\{|A, B||A \in U_{\delta_{1}}(P), B \in U_{\delta_{1}}(Q)\}) \times \delta + \delta \leq \varphi_{\delta_{1}}(\max\{|A, B||A \in U_{\delta_{1}}(P), B \in U_{\delta_{1}}(Q)\}) \times \delta_{1} + \delta_{1}$$

$$(11)$$

From (10) and (11), we have $U_{\delta}(P,Q) \leq U_{\delta_1}(U_{\delta_1}(P), U_{\delta_1}(Q))$. Similarly,

$$\varphi_{\delta}(\min\{|A, B||A \in P, B \in Q\}) \times \delta \geq$$

$$\varphi_{\delta}(\min\{|A, B||A \in U_{\delta_{1}}(P), B \in U_{\delta_{1}}(Q)\}) \times \delta \geq$$

$$\varphi_{\delta_{1}}(\min\{|A, B||A \in U_{\delta_{1}}(P), B \in U_{\delta_{1}}(Q)\}) \times \delta_{1}$$
(12)

So $L_{\delta}(P,Q) \geq L_{\delta_1}(U_{\delta_1}(P), U_{\delta_1}(Q))$, from the formulas above, we can infer that $d_{\delta}(P,Q) \subseteq d_{\delta_1}(U_{\delta_1}(P), U_{\delta_1}(Q)).$

Given a set of rough point pairs $\{(P_t, Q_t) | t \in T\}$ in $SPACE(\delta)$, and all distances of these pairs are equal, according to Def. 10, $\bigcap_{t \in T} d_{\delta}(P_t, Q_t) \neq \emptyset$, because $\bigcap_{t \in T} d_{\delta}(P_t, Q_t) \subseteq d_{\delta_1}(U_{\delta_1}(P_t), U_{\delta_1}(Q_t))$ $(t \in T)$, therefore $\bigcap_{t \in T} d_{\delta_1}(U_{\delta_1}(P_t), U_{\delta_1}(Q_t)) \supseteq \bigcap_{t \in T} d_{\delta}(P_t, Q_t) \neq \emptyset$. It follows that the distances between the upper approximations $U_{\delta_1}(P_t)$ and $U_{\delta_1}(Q_t)$ $(t \in T)$ are equal in $SPACE(\delta_1)$.

2.4 Problems and Possible Improvement

Mapping to Construct Rough Space

In this section, we suppose that the rough space is constructed by a very simple mapping $\varphi_{\delta} : R \to Z, x \to \lfloor x/\delta \rfloor$, and the equivalence relation induced from φ_{δ} can lead to the regular partition in n-dimension Euclidean space \Re^n . Although more complex mappings can be used to construct the rough space and the approximate configuration formed by the un-regular partition in \Re^n may be more appropriate to some practical problems, the analysis of the geometric properties in such spaces will become a very difficult work and most useful principles in transformation that mentioned above will be lost.

Proper Transformation

In addition, we must notice that some propositions and definitions introduced in this section are just tenable under the condition of proper transformation. In other words, some principles may not stand up in some extreme situations. For example, a rough line segment will turn to be a rough point when being transformed into the upper space that is rough enough to cover the segment with only one equivalence class. For some configurations, especially digital character, the improper transformation into rougher spaces can not guarantee some important topological features invariable, such as connectivity and curvature, which will lead to the recognition error. Furthermore, choosing the proper transformation space actually belongs to the issue of seeking the proper granular level for solution, and it should be considered depending on the specific problem. In the following sections, the methods of computing the proper roughness of upper space in transformation will be further introduced according to the specific applications.

Extension to n-Dimensional Space

The research work in this paper focuses on introducing rough geometry and its application in digital character recognition, thus the definitions and properties given above are mainly considered in 2D spaces. The generalization of this theory from 2D to nD will be our future work considering the specific applications, and the novel properties discovered will be further compared with the similar research work in [7].

3 Application of Rough Geometry

3.1 Application in Equichordal Point Problem

In Euclidean geometry, the Equichordal Point Problem can be formulated in simple geometric terms. If C is a Jordan curve on the plane and $P, Q \in C$ then the line segment PQ is called a chord of the curve C. A point inside the curve is called equichordal if every two chords through this point have the same length. For example, it is a well-known fact that there exits one equichordal point in a circle and the center of circle is the equichordal point. But can a convex shape have more than one equichordal point? This question was posed by Fujiwara in 1916 and independently by BlaschkeRothe and Weizenbock in 1917. Since then, the problem whether there exists a closed convex curve of two equichordal points had been a classic issue in traditional geometry until it was resolved by M.R.Rychlik in 1997 [20]. He proved that there exists no closed convex curve of more than one equichordal point. The Euclidean curve of two equichordal points and the analysis of its features are also introduced in the related research work (see, *e.g.*, [12,20]).

Although the research of M.R.Rychlik is of significant theoretical value, the closed convex curve of more than one equichordal point can exist in other spaces rougher than the Euclidean space. Because the representations of the shapes in the real world are always the approximations of the Euclidean geometric elements rather than absolutely accurate and abstract, the results from the analysis of the traditional geometric problem in rough space may be available in some specific applications. In the following paragraphs, we will introduce how to construct the proper rough space to represent the convex shape of two equichordal points.

Fig. 10.5 represents the closed convex curve in rough space SPACE(b/600), where a is the distance between two equichordal points, b is the length of the common chord and a = 0.3b in Euclidean space. The partition is so fine that we can denote a and b in the rough space instead of the distance approximation for simplicity.

The curve of two equichordal points in Euclidean space can be constructed as follows (see Fig. 7): the two equichordal points are laid on the horizontal axis symmetrically, the distance between the two points is a and the length of the common chord is b, given a point in the plane denoted by number 0 as the initial point, from the initial point 0 a line segment of length b through the left equichordal point should be drawn, and the other end point of this segment will be denoted as point 1, then from the point 1 the second line segment of length bcan be made through the right equichordal point and the new end point will be marked as number 2. In such process, the ordinal line segments passing through the left and right equichordal points respectively are created repeatedly, and the coordinates of the n+1th point can be computed from the nth point according to the iterative formula in the polar coordinate system and complex space [12]. The related research had proven that for any initial point in the plane, the iteration will converge to a pair of conjugated points, and all points denoted by the even numbers that converge to the right equichordal point can form a continuous curve, similarly, the all points denoted by the odd numbers that converge to the



Fig. 7. Curve of a = 0.5b

left equichordal point will create another continuous curve, these two curves will be named even curve and odd curve respectively in the next paragraphs.

The shape formed by the continuous curves that converge to the conjugated points on horizontal axis is shown in Fig. 8.1, the segment of the even curve in the first quadrant and the part of the odd curve in the third quadrant are convex, but the corresponding proportions of the curves in the second and fourth quadrants are not convex and have pulsation. The even curve and odd curve are not connected, but when the proper initialization makes the two continuous curves close enough on the horizontal axis, we can consider the shape as a closed curve approximately. In this closed curve, the chord is the line segment from the even number point to the odd number point. The approximate closed curve constructed by passing the left equichordal point first as mentioned above is called the right curve see Fig. 8.1, and the similar approximate construction through the right equichordal point first is called the left curve, see Fig. 8.2. We can also infer that for any pair of equichordal points on horizontal axis, the left and right curves that are symmetric with respect to the vertical axis can be constructed by adopting the symmetric initial points.

Based on the above introduction, the closed curve of two equichordal points in different rough spaces can be represented. Let the distance between two equichordal points a is 0.5 and the length of the common chord b is 1. Given a rough space $SPACE(\delta)$, when $\delta = 1/20$, the equivalence class in the space is too small to shield the pulsation, thus the approximation of the closed curve is not convex in SPACE(1/20), see Fig. 9.1. While $\delta = 1/9$, the left and right curves will have the common upper approximation that is convex in SPACE(1/9), see Fig. 9.2.

Next we will describe the impacts of the ratio between distance a and the chord length b to the shape of closed curve in the same space. Given a rough



Fig. 8. Right and Left Curves



Fig. 9. Closed Curve in Different Spaces



Fig. 10. Right Curves in SPACE(b/600)

space $SPACE(\delta)$, where $\delta = b/600$ and the common chord length b is a fixed value, the different shapes of the right curve according to the variant distance a are shown in Fig. 10.

It is apparent that the pulsation of the closed curve is gradually becoming weaker as the distance a reduces. As illustrated in Fig. 10.1-10.4, the approximation of the right curve is not convex and not symmetric about the vertical axis. We can define the convexity of the approximation of the right curve in the rough space according to its symmetric left curve as follows. If the upper approximation of the right curve can cover the left, this approximation is the common representation of both curves of two equichordal points in the rough space. Because the two curves are symmetric respect to the vertical axis, the common approximation is symmetric about the vertical axis according to Theorem 6. As introduced above, the closed curve is convex in the first and third quadrants, according to Theorem 4, the approximation is still convex in these quadrants, since its symmetry, the approximate shape is completely convex in all quadrants. Furthermore, we can obtain the result that the length approximations of all chords in the rough space are equal from Theorem 7. In this way, a convex closed curve of two equichordal points can be obtained in rough space. As shown in Fig. 10.5, when the ratio of a to b is 0.3, the pulsation is completely covered by the convex and symmetric approximation of closed curves in SPACE(b/600).

From the paragraphs above, we have learnt the important factors that influence the shape of the closed curve of two equichordal points. It can be inferred that for any common chord length b and the distance between two equichordal points a, given any initial point in the plane, there must exists a rough space in which left and right curves have the common upper approximation, and this approximation is a convex shape of two equichordal points. Especially when the corresponding partition in the space is fine enough, the rough configuration may turn to be a real closed convex curve in our vision. Thus the closed convex curve of two equichordal points can be constructed in the rough space.

As introduced in this section, the application of rough geometry in the Equichordal Point Problem indicates that the rough configurations in the approximate space have their own properties different from those of the shapes in Euclidean space, and the novel results may be obtained from observing the traditional geometric problems in the rough space.

3.2 Application in Principal Curves

The term "Principal Curves" was first proposed by Hastie and Stuetzle in 1984, and principal curves are usually defined as "self-consistent" smooth curves which pass through the "middle" of a n-dimensional probability distribution (see, *e.g.*, [3,4]). Principal curves can provide a nonlinear summary of the data through reflecting the data distribution in low dimensional space, and the curves' shape is suggested by the data. In another view, the principal curves are the skeleton of data set and the data set is the "cloud" around the curves. They construct the one-dimensional manifold of the data in high dimensional space and can be viewed as the nonlinear generation of the principal component analysis (PCA).

Because principal curves can preserve the most information of the data distribution, they usually serve as an efficient feature extraction tool. The field



Fig. 11. Principal Curves of '0' in Different Spaces

has been very active since Hastie and Stuetzle's groundbreaking work, numerous alternative methods for estimating principal curves have been proposed and analyzed (see, *e.g.*, [5,6,23,24]). The applications of this theory in various fields such as image analysis, feature extraction, and speech processing have demonstrated that principal curves are not only of theoretical interest, but also have a legitimate place in the family of practical unsupervised learning techniques (see, *e.g.*, [13,14,29]).

According to the definition of principal curves given by Hastie and Stuetzle, the self-consistency means each point of the curves is the average of all points that project there. Thus, the complexities of the algorithms producing principal curves are always closely relative to the scale of the data set. But in some practical problems, it may be not necessary to traverse all initial data points to produce the skeleton of the data distribution. In fact, the approximate representations of the curves that can catch the most important topological features of the distribution are sufficient for some recognition works, and these approximations can be obtained through only the rough points that can preserve the object's primary structure. For example, existing recognition methods of off-line handwritten characters usually generate the features from all pixels contained in the configurations, but the objects can be viewed at rougher granular level to obtain the same results (see Fig. 11).

Depending on the invariants of transformation among the rough spaces introduced in Section 2, such as the invariants of convexity, we can use principal curves to extract the geometric features of the characters in the spaces rougher than original images. This process can bring several benefits for the recognition work as follows: first, the efficiency of the algorithms for generating the skeletons will be improved as the scale of the original data set is greatly reduced; second, the detrimental effects of the trivial details produced from the redundant data in the character figures can be weakened at rougher level, and this result can also lead to the simplification of the classification rules as the third advantage.

As mentioned above, our exploratory work tries to apply the rough geometry to the character recognition. In our experiments, the polygonal line algorithm of the principal curves methods (see, *e.g.*, [5,6]) is adopted to produce the skeletons



Fig. 12. Process of Off-line Characters Recognition

of the off-line handwritten digits. Furthermore, the proper rough spaces for upper transformation can be obtained according to the thickness of the character and the classifier is constructed based on rough sets methods. The flow diagram of the system is shown in Fig. 12.

As illustrated in Table 1, the skeletons and the recognition results of the sample digit can be obtained in different rough spaces. One pixel of the original digital image is considered as the smallest equivalence class in the rough spaces and the corresponding $\delta = 1$, so the finest rough space constructed from the pixels is denoted by SPACE(1). The polygonal curves algorithm is used to extract the skeletons for generating the geometric features of characters at five different granular levels. The sample figures' visions, extracted skeletons and recognition results of two persons' handwritings in different rough spaces are displayed in the following table, in which N is the scale of the digital image, P is the number of points in the skeleton, and K is the components number of the principal curves extracted from the character.

In the experiment, we discovered that the transformation of upper approximation just causes little damage of the characters' geometric features that we are interested, and the skeletons got from the rough spaces are well enough for the recognition work. This observation is coincident with the invariants of the transformation introduced in the Section 2. From the analysis of the experimental data, we can see that the scales of the original data and the skeletons and the iteration times in the process for producing the principal curves are greatly reduced as the rough space transforms. It also should be noticed that the false recognition result caused by the trivial details can be rectified through the transformation (see Table 1). Accordingly, training with the features generated in the proper rough spaces, the classifier will be further simplified. As mentioned above, the efficiency of the recognition algorithm based on principal curves and rough sets can be effectively improved through the application of rough geometry as the preprocessing step in feature generation.

SPACE(1)	SPACE(4)	SPACE(8)	SPACE(12)	SPACE(20)
9	9	9	9	9
9	9	9	9	9
$N:500\times500$	$N:125\times 125$	$N:62\times 62$	$N:41 \times 41$	$N:25\times 25$
P: 159	P:60	P:46	P: 32	P: 24
K:3	K:2	K:2	K:2	K:2
Result:5	Result:9	Result:9	Result:9	Result:9
9	9	9	9	9
9	9	9	9	9
$N:500 \times 500$	$N:125\times 125$	N:62 imes 62	$N:41 \times 41$	$N:25 \times 25$
P:150	P:56	P:45	P:32	P:22
K:3	K:3	K:2	K:2	K:2

Table 1. Recognition Results of Figure '9' in Rough Spaces

4 Conclusion and Prospect

In traditional geometry, the geometric elements are defined absolutely abstract and accurate, but the configurations we see in the real world always have lengths or sizes. Rough geometry attempts to combine the rough set theory with the geometric methods, generate and analyze the graphics at the rougher granular levels through the approximation transformation. The aim of the investigation is to construct the proper geometric spaces more available for problem solving. The motivation of the research and some principles of rough geometry have been introduced, and we also presented the applications of this theory in the traditional geometry problem and characters recognition. Although the new geometry is expected to be an effective tool for measuring the configurations in approximate spaces, at present it is just based on the personal views and ideas immature, perhaps controversial. There is still a long way to go before it turns to be an integrated system. In our future work, the improvement and the enrichment in theory will be continued, while the applications of the theory will be studied further as well. In the next paragraphs, we will describe the questions induced from our exploratory research according to the basic issues of the granular computing (see, *e.g.*, [25,26]), and expect the possible solutions for these problems in the future work.

Granulation

How can we construct the optimal approximate space and representation of the configuration according to the specific application?

This question refers to the constructions of the basic components of the granular computing: granules, granulated views and hierarchies, and these terms may correspond to the equivalence classes, rough configurations and rough spaces respectively in rough geometry. Depending on the existing definitions in rough geometry, although the upper approximations can preserve some geometric features of the original graphics, it also may lose some important information. For example, the upper approximation can damage the property of connectivity and increase the number of loops in the graphics, the changes of the topological features will bring us the undesirable effects. As the exiting method for constructing the approximations of the objects needs to be further improved, the following ideas may be helpful solutions in the future.

The other approximation forms can be adopted in the same rough space defined in Section 2, such as lower approximation, then through combining the information obtained from different approximations, we can maintain the most features of the original objects in the approximate representations. In other words, we can observe the graphics from multiple profiles to get the sufficient information about the features. The second suggestion for improving the approximation is that we can define the different approximate spaces to catch the most geometric features in transformation. But it always requires the more complex mapping to induce the un-regular partitions in the space rather than the simple construction of the rough space. Although it may be a difficult work, we must notice that the semantics of the objects are usually ignored in the construction of the approximate space, it is possible to form the optimal un-regular rough spaces based on the objects' contents to preserve the most features in transformation.

Computing with Granules

How can we construct the proper mappings between multi-level approximate spaces to preserve the most properties of the objects? How can we decide the optimal granular level for the problem solving?

The key points in the issue of computing with granules are mappings between different level of granulations, granularity conversion and property preservation, and they are also the essential targets of the transformation in rough geometry. In this paper, we define the mappings between different spaces from fine to rough like the upper approximation in rough sets, and it can lead to a simple transformation. This transformation will make it easy to seek the rules of properties preservation, but it also can cause the damage of the important geometric features as introduced above. So a more proper transformation should be defined according to the specific problems, such as the un-regular transformation may be defined based on the characters' structure to catch more geometric properties in changing spaces. Furthermore, the upper approximation transformation is the bottom-up way to construct the hierarchy, while the inverse transformation, i.e. the top-down approaches, may be also useful for features preservation, for example, the analysis of the important details in local areas of the whole rough configuration can rectify the properties' loss in transformation.

In the practical problems, people usually tend to choose a proper granular level for solution, and the ability of cruising among the different levels of granularity with freedom is factually an embodiment of the human intelligence. The issue also exists in granular computing as a research hotspot. In pattern recognition process, the optimal granular level is always expected to improve the recognition results, and at this level, the trivial details will be neglected while the import features should be preserved even more distinct. Choosing the proper granular level for problem solving in granular computing corresponds to the computation about the roughness of the space in rough geometry. We suppose that constructing a proper rough space in rough geometry may be through the following two ways. The first is to formulate the topological changing of the rough configurations according to the regular transformation, but it may be a difficult work. The other way is to construct the proper space according to the given parameters of roughness based on the specific application, these parameters can be obtained from data training or the empirical knowledge. For the off-line handwriting recognition, the proper rough spaces can be constructed according to the average thickness of the characters as the prior knowledge. Moreover the optimal roughness can also be obtained from the data training, and this method often requires the evaluation criterion of the topological variation.

By the way, it is usually believed that the geometric property values will become more and more accurate as the approximate space transforming from rough to fine, such as the area value of the closed configuration. But the properties do not always behave like this, for example, the relative deviation of perimeter for the digitized polygon will converge to a fixed value when the image resolution turning big, in which the relative deviation is computed from the absolute difference between the property values for the approximation of the graphics and that for the same graphics in Euclidean space [7]. Thus choosing the proper rough level to catch the geometric features in the approximate space is one of the most important issues in the rough geometry research.

In this section, we have discussed the existing problems and the future work on rough geometry depending on the basic issues of granular computing. The new theory is so immature that it needs further development in many aspects, but it will provide a new perspective to observe the geometric elements in reality and encourage us to analyze the objects in multiple levels and views. Furthermore, the applications of rough geometry in practical problems are also attached importance in the related work, so the research subject will be valuable in both theory and application.

Acknowledgements

This work was supported by National Natural Science Foundation of China (Serial No. 60475019, 60775036) and The Research Fund for the Doctoral Program of Higher Education (Serial No. 20060247039).

References

- 1. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 2nd edn. Publishing House of Electronics Industry, Beijing (2006)
- Hassanien, A.: Fuzzy rough sets hybrid scheme for breast cancer detection. Image and Vision Computing 25(2), 172–183 (2007)
- 3. Hastie, T.: Principal Curves and Surfaces. Unpublished doctoral dissertation, Stanford University, USA (1984)
- Hastie, T., Stuetzle, W.: Principal curves. Journal of the American Statistical Association 84(406), 502–516 (1988)
- 5. Kégl, B.: Principal curves: learning, design, and applications. Unpublished doctoral dissertation, Concordia University, Canada (1999)
- Kégl, B., Krzyzak, A.: Learning and design of principal curves. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(3), 281–297 (2000)
- Klette, R., Rosenfeld, A.: Digital Geometry: Geometric Methods for Digital Image Analysis. Beijing World Publishing Corporation, Beijing (2006)
- Lin, T.Y.: Granular Computing on Binary Relations I: Data Mining and Neighborhood Systems. In: [19], pp. 107–121 (1998)
- Lin, T.Y.: Granular Computing on Binary Relations II: rough set representations and belief functions. In: [19], pp. 121–140 (1998)
- Lin, T.Y.: Granular Computing: Fuzzy Logic and Rough Sets. In: Skowron, A., Polkowski, L. (eds.) Computing with words in information/intelligent systems, pp. 183–200. Physica-Verlag, Heidelberg (1999)
- 11. Lin, T.Y.: Granular computing rough set perspective. The Newsletter of the IEEE Computational Intelligence Society 2(4), 1543–4281 (2005)
- 12. Ma, Y.: Rough Geometry. Computer Science 33(11A), 8 (2006) (in Chinese)
- Miao, D.Q., Tang, Q.S., Fu, W.J.: Fingerprint Minutiae Extraction Based on Principal Curves. Pattern Recognition Letters 28, 2184–2189 (2007)
- Miao, D.Q., Zhang, H.Y.: Off-Line Handwritten Digit Recognition Based on Principal Curves. Acta Electronica Sinica 33(9), 1639–1644 (2005) (in Chinese)
- Mushrif, M.M., Ray, A.K.: Color image segmentation: Rough-set theoretic approach. Pattern Recognition Letters 29, 483 (2008)
- Pal, S.K., Mitra, P.: Multispectral image segmentation using the rough-setinitialized EM algorithm. IEEE Transactions on Geoscience and Remote Sensing 40(11), 2495–2501 (2002)

- Pawlak, Z.: Rough sets. International Journal of Computer and Information Sciences 11, 341–356 (1982)
- Pawlak, Z.: Rough sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht (1991)
- Polkowski, L., Skowron, A. (eds.): Rough sets in knowledge discovery. Physica-Verlag, Heidelberg (1998)
- Rychlik, M.R.: A complete solution to the equichordal point problem of Fujiwara, Blaschke, Rothe and Weizenbock. Inventiones Mathematicae 129, 141–212 (1997)
- Skowron, A.: Toward intelligent systems: calculi of information granules. Bulletin of International Rough Set Society 5, 9–30 (2001)
- Skowron, A., Stepaniuk, J.: Information Granules: Towards Foundations of Granular Computing. International Journal of Intelligent Systems 16, 57–85 (2001)
- Tibshirani, R.: Principal curves revisited. Statistics and Computation 2, 183–190 (1992)
- Verbeek, J.J., Vlassis, N., Kröse, B.: A k-segments algorithm for finding principal curves. Pattern Recognition Letters 23, 1009–1017 (2002)
- Yao, Y.Y.: Information granulation and rough set approximation. International Journal of Intelligent Systems 16(1), 87–104 (2001)
- Yao, Y.Y.: A partition model of granular computing. In: Peters, J.F., Skowron, A., Grzymała-Busse, J.W., Kostek, B.z., Świniarski, R.W., Szczuka, M.S. (eds.) Transactions on Rough Sets I. LNCS, vol. 3100, pp. 232–253. Springer, Heidelberg (2004)
- 27. Zadeh, L.A.: Fuzzy sets and information granulation.advances in fuzzy set theory and applications. North-Holland Publishing, Amsterdam (1979)
- Zadeh, L.A.: Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Set s and Systems 19, 111–127 (1997)
- Zhang, H.Y., Miao, D.Q.: Analysis and Extraction of Structural Features of Off-Line Handwritten Digits Based on Principal Curves. Journal of Computer Research and Development 42(8), 1344–1349 (2005) (in Chinese)