YE.V. VODOLAZSKIY, Senior Research Associate,
International Research and Training Centre of Information
Technologies and Systems of the NAS and MES of Ukraine,
Glushkov ave., 40, Kyiv, 03187, Ukraine,
waterlaz@gmail.com

A MODIFICATION OF THE FRECHET DISTANCE FOR NONISOMORPHIC TREES

The paper presents a modification of the Frechet distance for nonisomorphic trees. While the classical Frechet distance between nonisomorphic trees is undefined, a new measure called similarity of a tree to a reference tree is given that is defined for a wider class of trees. A polynomial time algorithm is given for determining whether similarity of one tree to another is less than a given number.

Keywords: computational geometry, Frechet distance, trees.

Introduction

Many image processing methods involve selecting contours on the image and processing those contours with various algorithms such as computing the Frechet distance between the selected contour and the reference contour. Another common approach is to perform a so-called image skeletonization [6, 8] instead of contour selection. Computing the Frechet distance between curves, both open and closed, is a relatively well researched problem [1, 3, 5, 7]. Since an image contour is always some sort of a curve, comparing contours with the Frechet metric does not impose much difficulty. The skeleton of a simply connected binary image is in some sense an acyclic set of points (a tree). Computing the Frechet distance between trees [4] can be problematic since the metric is either undefined or equals infinity for non-isomorphic trees, depending on the point of view. An alternative for the Frechet metric is known [2], which is essentially a mixture of the Frechet distance between polygonal lines and Hausdorff distance. With this approach the distance between any pair of trees is defined and finite. We propose another variation on the Frechet distance that is also defined for nonisomorphic pairs of trees and is ideologically closer to the original Frechet metric, although the proposed distance is not symmetric and is therefore not a metric.

Problem formulation

Let $\Gamma = (V, E)$ be an acyclic connected graph (a tree) with a finite number of vertices $V \subset \mathbf{R}^k$, which are points in a k-dimensional linear space, and a set of edges $E \subset V \times V$. Each edge $(u,v) \in V$ defines a linear segment $\lambda(u, v) = \{\alpha u + (1 - \alpha) v \mid \alpha \in [0, 1]\}$, and the whole tree Γ defines a subset of points in $\mathbf{R}^k$, the union of all linear segments $X = \bigcup_{(u,v) \in E} \lambda(u,v)$, which are called points of the tree Γ.

Definition 1. The Frechet distance between trees Γ_1 and Γ_2 with their corresponding sets of points X_1 and X_2 is

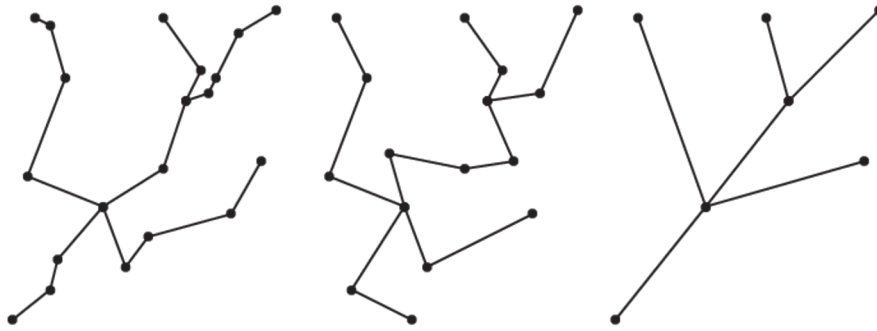$$f(X_1, X_2) = \inf_{\varphi \in \Phi} \max_{x \in X} d(x, \varphi(x))$$

Fig. 1. An example of trees with non-empty Φ. For any of given pair of trees the Frechet distance is defined and is finite
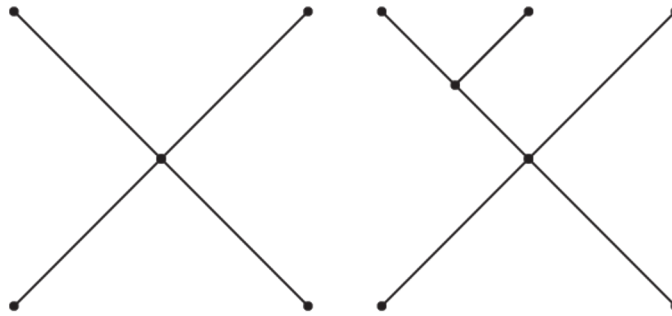

Fig. 2. A pair of trees with no continuous bijection and undefined Frechet distance

where Φ is the set of all continuous bijections between X_1 and X_2.

Definition 2. The Hausdorff distance between sets X and Y is the number

$$h(X,Y) = \max[\max_{x \in X} \min_{y \in Y} d(x,y), \max_{y \in Y} \min_{x \in X} d(x,y)]. (1)$$

The problem of the Frechet distance between trees is immediately apparent from the definition, since the value $f(X_1, X_2)$ is only defined when the set of continuous bijections Φ is not empty. The set Φ is never empty when Γ_1 and Γ_2 are polygonal lines. The same, however, can not be said for non-isomorphic pairs of trees.

Examples of trees that have at least one continuous bijection is given on Fig. 2, while

Fig. 2 gives an example of a pair of trees with no continuous bijection.

The Frechet metric needs to be weakened in order to be used on non-isomorphic trees. We do so by not limiting φ to continuous bijections.

For any function φ: $X_1 \rightarrow X_2$ and a point $x_2 \in X_2$ denote $\varphi^{-1}(x_2) = \{x_1 \in X_1 \mid \varphi(x_1)\}$ the preimage of $x_2 \in X_2$. Note that $\varphi^{-1}(x_2)$ in no way implies that φ is an invertible function.

For any continuous function φ: $X_1 \rightarrow X_2$ an image $X_2^* = \varphi(X_1^*)$ of any connected subset $X_1^* \subset X_1$ is also connected.

Definition 3. A function φ: $X_1 \rightarrow X_2$ is called monotone if preimage $\varphi^{-1}(x_2)$ of any point $x_2 \in X_2$ is a connected set.

Definition 4. For a given number ε the set X_1 is called ε–similar to the reference set X_2, if a continuous monotone function φ: $X_1 \to X_2$ exists such that φ(X_1)=X_2 and $d(x_1, \varphi(x_1)) \le \varepsilon$ for all $x_1 \in X_1$.

Definition 5. Similarity δ(X_1,X_2) of the set X_1 to the reference set X_2 is the exact lower bound on such values ε that X_1 is ε–similar to X_2.

The problem is to determine for given trees Γ_1, Γ_2 and a number ε whether X_1 is ε-similar to X_2, where X_1 and X_2 are the sets of points of Γ_1 and Γ_2 respectively.

General Properties of Set Similarity

The relation of ε-similarity of X_1 to X_2 as well as the function δ(X_1,X_2) of similarity is not symmetric with respect to its arguments. Therefore, δ(X_1,X_2) is not a metric. Nevertheless, it is ideologically close to the Hausdorff and Frechet metrics and lies in the middle of them in the sense of the following theorem.

Theorem 1. Let $h(X_1,X_2)$ be the Hausdorff distance between X_1 and X_2, let $f(X_1,X_2)$ be the Frechet distance between X_1 and X_2.

For all X_1 and X_2 the following inequalities are valid

$$h(X_1,X_2) \le \delta(X_1,X_2) \le f(X_1,X_2). \qquad (2)$$

The equality δ(X_1,X_2) = 0 holds if and only if X_1=X_2.

Proof. We first prove that when $f(X_1,X_2) = f^*$, the inequality δ(X_1,X_2) $\le f^*$ holds. Indeed, the expression $f(X_1,X_2) = f^*$ means that for any $\varepsilon^* > f^*$ a continuous bijection φ*: $X_1 \leftrightarrow X_2$ exists, such that for any $x \in X_1$ the inequality $d(x_1, \varphi(x_1)) \le \varepsilon^*$ holds. Since φ* is continuous and invertible, then according to Definition 3 it is monotone and according to Definition 4 the set X_1 is ε*–similar to X_2 for all $\varepsilon^* > f^*$. Therefore, δ(X_1,X_2) $\le f^*$ ccording to Definition 5.

We now prove that if δ(X_1,X_2) $\le \delta^*$ then $h(X_1,X_2) \le \delta^*$. The equality δ(X_1,X_2) $= \delta^*$ means that for any $\varepsilon^* < \delta^*$ there exists a monotone continuous functions φ: $X_1 \to X_2$ such that for any $x \in X_1$ the

inequality $d(x_1, \varphi^*(x_1)) \le \varepsilon^*$ holds. For a function φ* there exists a function ψ*: $X_1 \to X_2$ such that ψ*(x_2) ∈ φ*(x_2) and

$$d(x_1, \varphi^*(x_1)) \le \varepsilon^* \text{ for all } x_1 \in X_1,$$

$$d(\psi^*(x_2), x_2) \le \varepsilon^* \text{ for all } x_2 \in X_2.$$

From the inequalities it obviously follows that

$$\min_{x_2 \in X_2} d(x_1, x_2) \le \varepsilon^* \text{ for all } x_1 \in X_1,$$

$$\min_{x_1 \in X_1} d(x_1, x_2) \le \varepsilon^* \text{ for all } x_2 \in X_2,$$

$$\max_{x_1 \in X_1} \min_{x_2 \in X_2} d(x_1, x_2) \le \varepsilon^*,$$

$$\max_{x_2 \in X_2} \min_{x_1 \in X_1} d(x_1, x_2) \le \varepsilon^*,$$

and finally

$$\max \{\max_{x_1 \in X_1} \min_{x_2 \in X_2} d(x_1, x_2), \max_{x_2 \in X_2} \min_{x_1 \in X_1} d(x_1, x_2)\} \le \varepsilon^*.$$

According to Definition 2 the left-hand side of the last inequality is the Hausdorff distance $h(X_1,X_2)$. Since $h(X_1,X_2) \le \varepsilon^*$ for all $\varepsilon^* > \delta^*$, then $h(X_1,X_2) \le \delta^* = \delta(X_1,X_2)$.

The last statement of the theorem

$$[\delta(X_1,X_2) = 0] \Leftrightarrow [X_1 = X_2]$$

directly follows from the proved inequalities (2) and the fact that both h and f are metrics.

Properties of Tree Similarity

Let $\Gamma = \langle V, E \rangle$ be a tree that defines a set of points X.

Definition 6. Denote $P(u,v) \subseteq X$ the minimal connected subset of X that contains u and v.

Definition 7. A point $u \in V \subset \mathbf{R}^k$ is called a branching point of X if u has strictly more than two adjacent vertices. A point $u \in V \subset \mathbf{R}^k$ is called a terminal point of X if u has exactly one adjacent vertex. All other points from X are called ordinary points.
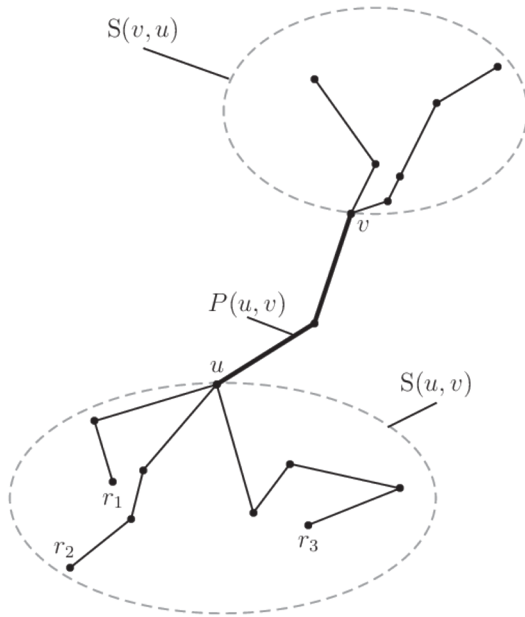
Fig. 3. An example of an unbranched section $P(u,v)$ and two sectors $S(u,v)$ and $S(v,u)$ that together form the whole set X
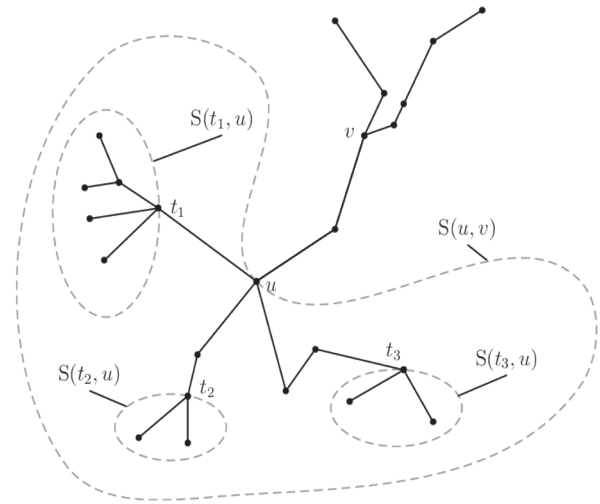

Fig. 4. An example of sector $S(u,v)$, which is a union of sectors $S(t_1,u)$, $S(t_2,u)$, $S(t_3,u)$ and unbranched sections $P(t_1,u)$, $P(t_2,u)$, $P(t_3,u)$. Also $F(t_1,u) = S(t_1,u) \cup P(t_1,u)$, $F(t_2,u) = S(t_2,u) \cup P(t_2,u)$ and $F(t_3,u) = S(t_3,u) \cup P(t_3,u)$

A path $P(u,v)$ can either go through branching points of X or not. If all points $r \in P(u,v)\backslash\{u,v\}$ are not branching points the path is called an unbranched section (see Fig. 4).

Definition 8. For any unbranched section $P(u,v)$ denote $S(u,v)$ the set of such points t that $P(t,u) \cap P(u,v) = \{u\}$ and call it a sector (see Fig. 4).

Any terminal vertex is a special case of a sector.

Definition 9. For any pair of $r \in V$ and $s \in V$ such that each r and s are either branching points or terminals denote the set $Z(r,s)=(X \backslash S(r,u)) \backslash S(s,v) \cup \{r,s\}$, where u and v are branching points on $P(r,s)$ (see Figure 4) and call it a subtree of X.

Obviously $P(r,s) \subseteq Z(r,s)$ is part of the subtree $Z(r,s)$ without all the branches.

Definition 10. A flower $F(u,v)$ *is* the union of the corresponding unbranched path and the sector.

$$F(u,v)=S(u,v) \cup P(u,v).$$

The provided definitions result in a number of properties.

Let r be a terminal point and let s be a branching point such that $P(r,s)$ is an unbranched section. Then $S(r,s) = \{r\}$ and $S(s,r) \cup P(r,s) = X$. Or even more general.

Property 1. If $P(u,v)$ is an unbranched section (see Fig. 4), then

$$X=S(u,v)\cup P(u,v)\cup S(v,u).$$

Any sector that is not a terminal vertex is a union of flowers as the following property states.

Property 2. Let $S(u,v)$ be a sector and let $\{t_1, t_2, ..., t_{n-1}, t_n\}$ be the set of all branching $t_i \in S(u,v)$ such that $P(t_i,u)$ is an unbranched section. In this case (Fig. 4) sector $S(u,v)$ can be expressed as

$$S(u,v) = \bigcup_{i=1}^{n}[S(t,u) \bigcup P(t,u)] =$$
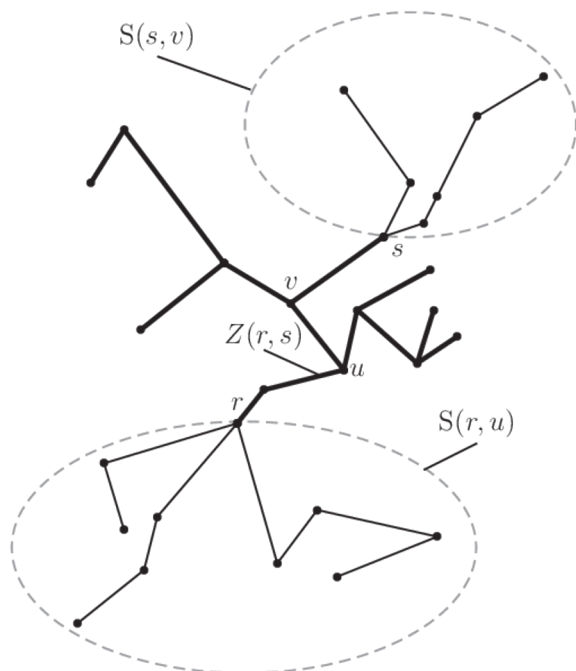$$= \bigcup_{i=1}^{n}[F(t,u)]$$

Fig. 5. A subtree $Z(r,s)$ (shown in bold) is a part of the whole tree X that is enclosed between two sectors $S(r,u)$ and $S(s,v)$ such that $X = S(r, u) \cup Z(r, s) \cup S(r, v)$

So to sum up, a sector is a union of flowers while each flower is a union of a sector and an unbranched path.

For any sector S a single set of flowers F_1, F_2,......, F_n exist that satisfy Property 2. We may say that secto S consists of flowers F_1, F_2,..., F_n.

The main idea of the algorithm presented further is that for certain subsets $X_1^* \subseteq X_1$ and $X_2^* \subseteq X_2$ it is determined whether a continuous monotone mapping $\varphi : X_1^* \to X_2^*$ exists, such that

$$\max_{x \in X_1^*} d(x_1, \varphi(x_l)) \le \varepsilon.$$

Whenever such mapping exists we would say that X_1^* maps onto X_2^* and denote this as $G(X_1^*, X_2^*) = 1$. Otherwise, $G(X_1^*, X_2^*) = 0$ denotes that such mapping does not exist.

Algorithm Description

When X_1 and X_2 are points of trees Γ_1 and Γ_2 the concept of monotone and continuous function $\varphi: X_1 \to X_2$ has a very demonstrative interpretation expressed in the following three lemmas.

Lemma 1. For a monotone continuous function $\varphi : X_1 \to X_2$ and any three points $u,v,w \in X_1$ such that $w \in P(u,v)$ the relation $\varphi(w) \in P(\varphi(u), \varphi(v))$ holds.

Proof. Let us show that when the function $\varphi: X_1 \to X_2$ is continuous then either $\varphi(w) \in P(\varphi(u), \varphi(v))$ or φ is not monotone. Denote $r = med(\varphi(w), \varphi(u), \varphi(v))$ a median of points $\varphi(w)$, $\varphi(u)$, $\varphi(v)$, that is a point of intersection of paths $P(\varphi(u), \varphi(v))$, $P(\varphi(w), \varphi(v))$ and $P(\varphi(w),\varphi(u))$. Either the equality $r = \varphi(w)$ holds or inequality $r \ne \varphi(w)$ holds. If $r = \varphi(w)$ then $\varphi(w) \in P(\varphi(u), \varphi(v))$.

Suppose that $r \ne \varphi(w)$. Since the restriction of function φ to $P(u,w)$ is a continuous function, the point $w_1 \in P(w,v)$, $w_1 \ne w$, exists such that $\varphi(w_1) = r$. The restriction of φ to $P(w,v)$ is also a continuous function, and therefore there exists a point $w_2 \in P(w,v)$, $w_2 \ne w$, such that $\varphi(w_2) = r$. Therefore, $w_1^{-1} \in \varphi(r)$, $w_2^{-1} \in \varphi(r)$, and the point w, that belongs to $P(w_1,w_2)$ does not belong to $\varphi^{-1}(r)$, since $\varphi(w) \ne r$. This means that the preimage $\varphi^{-1}(r)$ is not connected and φ is not monotone.

Lemma 2. If $\varphi: X_1 \to X_2$ is a monotone continuous function then for any three points $u,v,w \in X_1$ the equality $\varphi(med(u,v,w)) = med(\varphi(u), \varphi(v), \varphi(w))$ holds.

Proof. Median r of the triple $u,v,w \in X_1$ belongs to paths $P(u,v)$, $P(u,w)$ and $P(v,w)$. Therefore, according to Lemma 1, the point $\varphi(r)$ belongs to paths $P((u), (v))$, $P(\varphi(u), \varphi(w))$ and $P(\varphi(v), \varphi(w))$, or in other words is a median of three points $\varphi(u)$, $\varphi(v)$, $\varphi(w) \in X_2$.

The algorithm works by determining whether certain subsets $X_1^* \subseteq X_1$ of tree X_1 can be mapped onto certain subsets $X_2^* \subseteq X_2$ of tree X_2. Recall that when such mapping exists it is denoted by $G(X_1^*, X_2^*)$. Computing $G(X_1^*, X_2^*)$ can be divided into these important special cases.

1. Can a sector $S(u,v)$ be mapped onto a sector $S(r,s)$ such that $\varphi(u) = r$.

2. Can a flower $F(u,v)$ be mapped onto a flower $F(r,s)$ such that $\varphi(v) = s$.

3. Can a flower $F(u,v)$ be mapped onto a single point $\{r\}$.

4. Can a subtree $Z(u,v)$ be mapped onto an unbranched path $P(r,s)$ such that $\varphi(u) = r$ and $\varphi(v) = s$.

Mapping a flower on a flower

A flower $F(u,v)$ can be mapped onto a flower $F(r,s)$ if there exists a sector $S(p,t) \subset F(u,v)$ that can be mapped onto a sector $S(r,s)$ such that the subtree $Z(p,v)$ can be mapped onto the unbranched section $P(r,s)$ (see Fig. 6). This can be written as

$$G(F(u,v), F(r,s)) =$$
$$= \bigvee_{\substack{S(p,t) \\ S(p,t) \subset F(u,v) \\ t \in P(p,v)}} G(Z(p,v), P(r,s)) \,\&\, G(S(p,t), S(r,s)).$$

Mapping a sector on a sector

If a sector $S(u,v) = \bigcup_{i=1}^{m} F(t_i, u)$ is mapped onto a sector $S(u,v) = \bigcup_{i}^{n} F(p_i, r)$ via a function φ then $\varphi(u) = r$ and each flower $F(t_i, u)$, $i = 1,..., m$ must either map onto some flower $F(p_i, r)$ or map onto the vertex r. Moreover, no two flowers can map on the same flower, but multiple flowers can map onto the vertex r. Whether such a correspondence exists can be determined by solving a maximum matching problem on a bipartite graph (see Fig. 7).

Mapping a flower on a point

Testing whether $G(F(t, u), r) = 1$ is straightforward. Indeed,

$G(F(t, u), r) = 1 \Leftrightarrow \forall\, x \in F(t, u) \cap V_1 : d(x,r) \le \varepsilon$, so the algorithm needs to check the proximity of finite number of points to r.

Mapping a subtree on an unbranched path

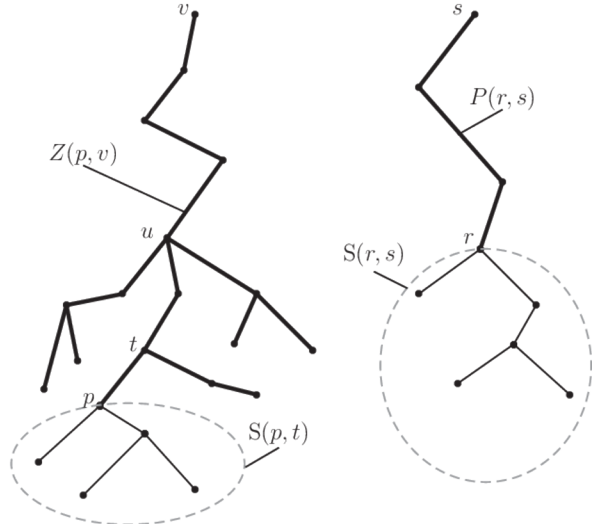Testing $G(Z(p, v), P(r, s)) = 1$ may seem tricky. However, it is not much different from testing

Fig. 6. A flower $F(u,v)=S(p,t) \cup Z(p,v)$ (on the left) being mapped onto a flower $F(r,s)= S(r,s) \cup P(r,s)$ (on the right). The subtree $Z(p,v)$ and the unbranched section $P(r,s)$ are shown in bold
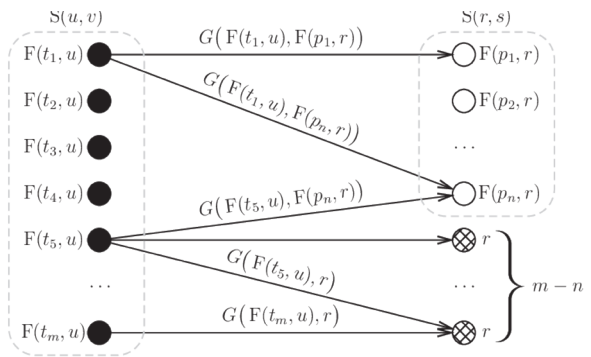

Fig. 7. A maximum bipartite matching problem for computing $G(S(u,v),S(r,s))$, when $S(u,v) = \bigcup_{i=1}^{m} F(t_i, u)$ and $S(u,v) = \bigcup_{i=1}^{n} F(p_i, r)$

$G(P(p, v), P(r, s)) = 1$, which is the problem solved by Alt and Godau [1] for testing whether the Frechet distance between polygonal curves does not exceed.

Conclusions

We have presented a measure of similarity of one tree to another reference tree. This measure of similarity is not a metric, however, it is weaker than the Frechet metric, which allows it to be finite on nonisomorphic trees, but stronger than the Hausdorff metric.

The presented algorithm for determining if the similarity between to trees is less than a given number ε takes polynomial time and works by reducing the problem of comparing two trees into smaller problems of comparing subsets of these trees (called flowers and sectors), which in turn are reduced to smaller and smaller problems and so on.

REFERENCES

1. Alt H., Godau M., 1995.“Computing the Frechet distance between two polygonal curves”. Int. J. Comput. Geometry Appl.Vol. 5,pp. 75−91.

2. Berezsky O., Zarichnyi M., 2016. “Frechet distance between weighted rooted trees”.MatematychniStudii. Dec., Vol. 48.

3. Berezsky O., Zarichnyi M., 2018. “Gromov-Frechet distance between metric curves”.MatematychniStudii. 2018. Aug., Vol. 50.

4. Buchin M., Krivosija A., Neuhaus A., 2020.“Computing the Frechet distance of trees and graphs of bounded tree width”.arXiv:2001. 10502.

5. Buchin K., 2017.“Four Soviets Walk the Dog: Improved Bounds for Computing the Fr chet Distance”. Discrete & Computational Geometry.

6. Lee T., Kashyap R., Chu, C., 1994.“Building Skeleton Models via 3-D Medial Surface Axis Thinning Algorithms”. CVGIP: Graphical Models and Image Processing. Vol. 56, no. 6, pp. 462−478.

7. Schlesinger M. I., Vodolazskiy E. V., Yakovenko V.M., 2016.“Frechet Similarity of Closed Polygonal Curves”. International Journal of Computational Geometry & Applications. Vol. 26, no. 01.

8. Zhang T. Y., Suen C. Y., 1984.“A Fast Parallel Algorithm for Thinning Digital Patterns.Commun. ACM. New York, NY, USA, Vol. 27, no. 3, pp. 236−239.

ЛІТЕРАТУРА

1. Alt H., Godau M. Computing the Frechet distance between two polygonal curves. Int. J. Comput. Geometry Appl. 1995. Vol. 5. P. 75−91.

2. Berezsky O., ZarichnyiM.Frechet distance between weighted rooted trees.MatematychniStudii. 2016. Dec. Vol. 48.

3. Berezsky O., ZarichnyiM.Gromov-Frechet distance between metric curves.MatematychniStudii. 2018. Aug. Vol. 50.

4. Buchin M., Krivosija A., Neuhaus A. Computing the Frechet distance of trees and graphs of bounded tree width. arXiv:2001.10502. 2020.

5. Buchin K. Four Soviets Walk the Dog: Improved Bounds for Computing the Fr chet Distance. Discrete & Computational Geometry. 2017.

6. Lee T., Kashyap R., Chu C. Building Skeleton Models via 3-D Medial Surface Axis Thinning Algorithms. CVGIP: Graphical Models and Image Processing. 1994. Vol. 56, no. 6. P. 462−478.

7. Schlesinger M. I., Vodolazskiy E. V., Yakovenko V. M.Frechet Similarity of Closed Polygonal Curves. International Journal of Computational Geometry & Applications. 2016. Vol. 26, no. 01.

8. Zhang T. Y., Suen C. Y. A Fast Parallel Algorithm for Thinning Digital Patterns.Commun. ACM. New York, NY, USA, 1984. Vol. 27, no. 3. P. 236−239.

Є.В. Водолазський, старший науковий співробітник,
Міжнародний науково-навчальний центр
інформаційних технологій та систем НАН та МОН України,
03187, м. Київ, просп. Академіка Глушкова, 40, Україна,
waterlaz@gmail.com

МОДИФІКАЦІЯ МЕТРИКИ ФРЕШЕ ДЛЯ НЕІЗОМОРФНИХ ДЕРЕВ

Вступ. Скелетизація зображень для подальшого порівняння пар скелетів — поширена практика в розпізнаванні зображень. У більшості випадків скелет зображення — ациклічна підмножина поля зору. Поширеною метрикою для порівняння підмножин є метрика Фреше. Однак, відстань Фреше визначено лише для пар ізоморфних дерев, що зводить нанівець можливість практично застосувати таку метрику на деревах.

Ціль статті. Необхідно розробити метод порівняння дерев, ідеологічно близький до метрики Фреше, але визначений для пар неізоморфних дерев.

Результати. У статті запропоновано модифікацію метрики Фреше для неізоморфних дерев. Нова числова характеристика названа близькістю дерева до еталону і визначена в тому числі для деяких класів пар неізоморфних дерев. Запропоновано поліноміальний алгоритм розпізнавання того, що одне дерево є близьким до іншого з точністю до заданого числа.

Ключові слова: обчислювальна геометрія, метрика Фреше, дерева.