# Mapping polygons to the grid with small Hausdorff and Fréchet distance

# Mapping polygons to the grid with small Hausdorff and Fréchet distance*

Quirijn W. Bouts† Irina Kostitsyna† Marc van Kreveld§
Wouter Meulemans‡ Willem Sonke† Kevin Verbeek†

## 1 Introduction

Transforming the representation of objects from the real plane onto a grid has been studied for decades due to its applications in computer graphics, computer vision, and finite-precision computational geometry [8]. Two interpretations of the grid are possible: (i) the grid graph, consisting of vertices at all points with integer coordinates, and horizontal and vertical edges between vertices at unit distance; (ii) the pixel grid, where the only elements are pixels, which are unit squares. In the latter interpretation, one can choose between 4-neighbor or 8-neighbor grid topology.

The issues involved when moving from the real plane to a grid already start with the definition of a line segment on a (pixel) grid, also called a *digital straight segment* [10]. For example, it is already difficult to represent line segments such that the intersection between any pair is a connected set (or empty). More generally, the challenge is to represent objects on a grid in such a way that certain properties of those objects in the real plane transfer to related properties on the grid; connectedness of the intersection of two line segments is an example of this.

While most of the research related to *digital geometry* is done from the graphics or vision perspective, computational geometry has made a number of contributions as well. Besides finite-precision computational geometry [8] these include snap rounding [6, 7, 9], consistent digital rays with small Hausdorff distance [5], mosaic maps [4], and schematization by map matching [11].

We consider the problem of representing a simple polygon $P$ as a polygon in the grid with small distance between them. A *grid cycle* is a simple cycle of edges and vertices of the grid graph corresponding to the grid. A *grid polygon* is a set of pixels whose boundary is a grid cycle. Two of the standard ways of measuring the distance are the Hausdorff distance [1] and the Fréchet distance [2]; we will consider both.

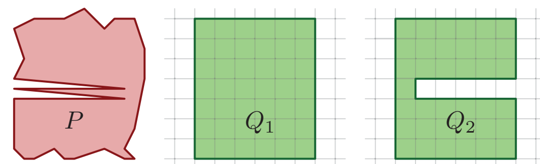



Figure 1: $d_H(P, Q_1)$ is small but $d_H(\partial P, \partial Q_1)$ is not. $d_H(P, Q_2)$ and $d_H(\partial P, \partial Q_2)$ are both small but the Fréchet distance $d_F(\partial P, \partial Q_2)$ is not.

Let $X$ and $Y$ be two subsets of a metric space. The (directed) *Hausdorff distance* $d_H(X, Y)$ from $X$ to $Y$ is defined as the maximum distance from any point in $X$ to its closest point in $Y$. In Section 2 we show that for any simple polygon $P$, a grid polygon $Q$ exists with $d_H(P, Q) \leq \frac{1}{2}\sqrt{2}$ and $d_H(Q, P) \leq \frac{3}{2}\sqrt{2}$ on the unit grid. Furthermore, the constructed polygon satisfies the same bounds for the distance between the boundaries $\partial P$ and $\partial Q$. This is not equivalent, since the point that realizes the maximum smallest distance to the other polygon may lie in the interior (Fig. 1).

Under the Hausdorff distance, the polygon boundary $\partial Q$ does not necessarily intuitively resemble $\partial P$ (Fig. 1, $P$ and $Q_2$). Therefore, the Fréchet distance $d_F$ [2] between the boundaries may be a better measure for similarity. Unlike the Hausdorff distance, however, not every polygon boundary $\partial P$ can be represented by a grid cycle with constant Fréchet distance. In Section 3 we present a condition on the input polygon boundary related to fatness (in fact, to $\kappa$-straightness [3]) and show that it allows a grid cycle representation with constant Fréchet distance.

## 2 Hausdorff distance

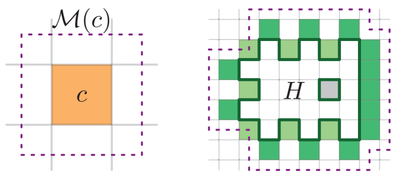
In this section, we present an algorithm that achieves a low Hausdorff distance between both the boundaries and the interiors of the input polygon $P$ and the resulting grid polygon $Q$. We say that two cells are *adjacent* if they share a segment. If two cells share only a point, then they are *point-adjacent*.

*Research on the topic of this paper was initiated at the 1st Workshop on Applied Geometric Algorithms (AGA 2015) in Langbroek, The Netherlands, supported by the Netherlands Organisation for Scientific Research (NWO) under project no. 639.023.208.

†Dept. of Mathematics and Computer Science, TU Eindhoven, The Netherlands, [q.w.bouts|i.kostitsyna|w.m.sonke|k.a.b.verbeek]@tue.nl. Supported by the Netherlands Organisation for Scientific Research (NWO) under project no. 639.023.208 and 639.021.541.

‡giCentre, City University London, United Kingdom, wouter.meulemans@city.ac.uk. Supported by Marie Skłodowska-Curie Action MSCA-H2020-IF-2014 656741.

§Dept. of Information and Computing Sciences, Utrecht University, m.j.vankreveld@uu.nl.




(a) (b)

Figure 2: (a) Module $\mathcal{M}(c)$ (dashed) of cell $c$. (b) Illustration to Lemma 1. $Q_1 \cap B$ in green; $Q_2 \cap B$ in gray; curve $C$ dashed.

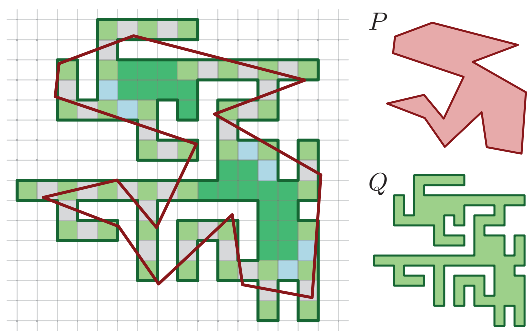



Figure 3: Example of the Hausdorff algorithm; the input and output are shown on the right. Colors: $Q_1$, $Q_2$, $Q_3$, $Q_4$.

**Algorithm.** We represent the grid polygon $Q$ as a set of cells (or pixels). If two cells $c_1 \in Q$ and $c_2 \in Q$ are point-adjacent, and there is no cell $c \in Q$ that is adjacent to both $c_1$ and $c_2$, then $c_1$ and $c_2$ share a *point-contact*. We construct $Q$ as the union of four sets $Q_1$, $Q_2$, $Q_3$, $Q_4$ (not necessarily disjoint). To define these sets, we define the *module* $\mathcal{M}(c)$ of a cell $c$ as the two-by-two square region centered at the center of $c$ (see Fig. 2(a)). Furthermore, since we can number the rows and columns, we can speak of even-even cells, odd-odd cells, odd-even cells, and even-odd cells. The four sets are defined as follows; see also Fig. 3.

$\boldsymbol{Q_1}$**:** All cells $c$ for which $\mathcal{M}(c) \subseteq P$.

$\boldsymbol{Q_2}$**:** All even-even cells $c$ for which $\mathcal{M}(c) \cap P \neq \emptyset$.

$\boldsymbol{Q_3}$**:** For all cells $c_1, c_2 \in Q_1 \cup Q_2$ that share a point-contact, the two cells that are adjacent to both $c_1$ and $c_2$ are in $Q_3$.

$\boldsymbol{Q_4}$**:** A maximal set of cells that does not introduce holes, and where each cell $c \in Q_4$ is adjacent to two cells in $Q_2$ and $\mathcal{M}(c) \cap P \neq \emptyset$.

We note that the set $Q_1 \cup Q_2$ is sufficient to achieve the desired Hausdorff distance. We add the set $Q_3$ to resolve point-contacts, and the set $Q_4$ to make the set $Q$ connected.

**Lemma 1** *The set $Q_1 \cup Q_2$ is hole-free, even when including point-adjacencies.*

**Proof.** For the sake of contradiction, let $H$ be a maximal set of cells comprising a hole. Consider the set $B$ of all cells in $Q_1 \cup Q_2$ that surround $H$ and are adjacent to a cell of $H$. Since $Q_2$ contains only even-even cells, every cell in $Q_2 \cap B$ must be (point-)adjacent to two cells in $Q_1 \cap B$ (see Fig. 2(b)). Hence, the boundary of the union of all modules of cells in $Q_1 \cap B$ is a single closed curve $C$; if this union contains a hole, $P$ would contain a hole as well. Since $C \subset P$ due to the definition of $Q_1$, the interior of $C$ must also be in $P$. Finally note that $C$ is a rectilinear curve through the centers of cells, but not through the center of a cell in $H$. Hence, the module of a cell in $H$ is completely inside $C$, implying $H \subset Q_1$; this contradicts our assumption. $\square$

**Lemma 2** *The set $Q$ is simply connected and does not contain point-contacts.*

**Proof.** Consider a point-contact between two cells $c_1, c_2 \in Q_1 \cup Q_2$ and a cell $c \notin Q_1 \cup Q_2$ that is adjacent to both $c_1$ and $c_2$ ($c \in Q_3$). Since $Q_2$ contains only even-even cells, we may assume that $c_1 \in Q_1$. Recall that $\mathcal{M}(c_1) \subseteq P$ by definition. We may further assume that $c_1$ is an odd-odd cell, for otherwise a cell in $Q_2$ would eliminate the point-contact. Hence, all cells point-adjacent to $c_1$ are in $Q_1 \cup Q_2$, and thus $c$ has three adjacent cells in $Q_1 \cup Q_2$. This implies that adding $c \in Q_3$ to $Q_1 \cup Q_2$ cannot introduce point-contacts or holes. Similarly, cells in $Q_4$ connect two oppositely adjacent cells in $Q_2$, and thus cannot introduce point-contacts (or holes, by definition). Combining this with Lemma 1 implies that $Q$ is hole-free and does not contain point-contacts.

It remains to show that $Q$ is connected. For the sake of contradiction, assume that $Q$ is not connected, so take two cells $c_1$ and $c_2$ in different connected components. We may further assume that $c_1, c_2 \in Q_2$, as cells in $Q_1 \cup Q_3 \cup Q_4$ must be adjacent or point-adjacent to a cell in $Q_2$. Let $p \in \mathcal{M}(c_1) \cap P$, $q \in \mathcal{M}(c_2) \cap P$ and consider a path $\pi$ between $p$ and $q$ inside $P$. Every even-even cell $c$ with $\mathcal{M}(c) \cap \pi \neq \emptyset$ must be in $Q_2$. Furthermore, the modules of even-even cells cover the plane. Thus, there must be two cells $c, c' \in Q_2$ in different components such that the module of the cell adjacent to both $c$ and $c'$ intersects $\pi$. This contradicts the maximality of $Q_4$. $\square$

**Upper bounds.** To prove our bounds, note that $\mathcal{M}(c) \cap P \neq \emptyset$ holds for every cell $c \in Q$. This is explicit for cells in $Q_1$, $Q_2$, and $Q_4$. For cells in $Q_3$, note that these cells must be adjacent to a cell in $Q_1$, and thus contain a point in $P$.

**Lemma 3** $d_H(P,Q), d_H(\partial P, \partial Q) \leq \frac{1}{2}\sqrt{2}$.

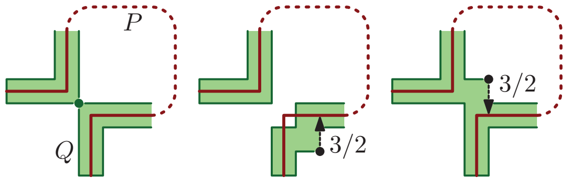



Figure 4: A polygon that does not admit a grid polygon with Hausdorff distance smaller than 3/2. The brown line signifies a very thin polygon.

**Proof.** Let $p \in P$ and consider the even-even cell $c$ such that $p \in \mathcal{M}(c)$. Since $c \in Q_2$, the distance $d_H(p,Q) \le d_H(p,c) \le \frac{1}{2}\sqrt{2}$. Now consider a point $p \in \partial P$. There is a $2 \times 2$-set of cells whose modules contain $p$. This set contains an even-even cell $c \in Q$ and an odd-odd cell $c' \notin Q$. The latter is true, because odd-odd cells in $Q$ must be in $Q_1$. Therefore, the point $q$ shared by $c$ and $c'$ must be in $\partial Q$. Thus, $d_H(p, \partial Q) \le d_H(p,q) \le \frac{1}{2}\sqrt{2}$. ◻

**Lemma 4** $d_H(Q,P), d_H(\partial Q, \partial P) \le \frac{3}{2}\sqrt{2}$.

**Proof.** Let $q \in Q$ and let $c \in Q$ be the cell that contains $q$. Since $\mathcal{M}(c) \cap P \neq \emptyset$, we can choose a point $p \in \mathcal{M}(c) \cap P$. It directly follows that $d_H(q,P) \le d_H(q,p) \le \frac{3}{2}\sqrt{2}$. Now consider a point $q \in \partial Q$, and let $c \in Q$ and $c' \notin Q$ be two adjacent cells such that $q \in \partial c \cap \partial c'$. We claim that $(\mathcal{M}(c) \cup \mathcal{M}(c')) \cap \partial P \neq \emptyset$. If $c \notin Q_1$, then the claim directly follows. Otherwise, $\mathcal{M}(c) \subseteq P$ implies that $\mathcal{M}(c') \cap P \neq \emptyset$ and clearly $\mathcal{M}(c') \nsubseteq P$. This in turn implies that $\mathcal{M}(c') \cap \partial P \neq \emptyset$. Let $p \in (\mathcal{M}(c) \cup \mathcal{M}(c')) \cap \partial P$. Then $d_H(q, \partial P) \le d_H(q,p) \le \frac{3}{2}\sqrt{2}$. ◻

**Theorem 5** *For every simple polygon $P$ there exists a simply connected grid polygon $Q$ without point-contacts such that $d_H(P,Q), d_H(\partial P, \partial Q) \le \frac{1}{2}\sqrt{2}$ and $d_H(Q,P), d_H(\partial Q, \partial P) \le \frac{3}{2}\sqrt{2}$.*

**Lower bounds.** In Fig. 4 a polygon is shown for which no grid polygon has Hausdorff distance below 3/2 between the boundaries or interiors. A naive construction of a grid polygon results in the left drawing of Fig. 4 which is not a simple polygon. To make it simple, we can either remove a cell (center) or add a cell (right). Both methods result in $d_H(\partial Q, \partial P) \ge 3/2 - \epsilon$, for any $\epsilon > 0$. Alternatively, we can fill the entire upper-right part of the grid polygon (not shown), resulting in a high $d_H(Q,P)$.

In the $L_\infty$ distance, the lower bound given in Fig. 4 also holds. Interestingly, in this measure, our algorithm achieves a Hausdorff distance of 3/2 (the upper-bound proofs can be straightforwardly modified to show this).

## 3 Fréchet distance

The Fréchet distance $d_F$ between two curves is generally a better measure for similarity than the Hausdorff distance; see [2] for a definition of the measure. We consider computing a grid polygon $Q$ whose boundary has constant Fréchet distance to the boundary of the input polygon $P$. We study under what conditions on $\partial P$ this is possible and prove a bound.

**Obesity.** Some input polygons $P$ do not admit a grid polygon $Q$ such that their boundaries have low Fréchet distance; see for example the polygon in Fig. 7(a). Intuitively, any grid polygon boundary $\partial Q$ approximating $\partial P$ must significantly deviate from it, because the grid is too coarse to follow $\partial P$ closely.

However, this problem is caused only by the thin spikes: if we assume that $P$ does not have those, we can do better. Let $|ab|_{\partial P}$ be the distance from $a$ to $b$ along $\partial P$. As defined in [3], a curve $C$ is *$\kappa$-straight* if for any two points $a, b \in C$, $|ab|_C \le \kappa \cdot |ab|$. In fact, we need this property on $\partial P$ only when $|ab| \le \sqrt{2}$, as we must deal with several parts of $\partial P$ being in the same grid cell. We therefore define a weaker fatness measure called *$\beta$-obesity*: a polygon $P$ is *$\beta$-obese* if for any two points $a, b \in \partial P$ with $|ab| \le \sqrt{2}$, $|ab|_{\partial P} \le \beta$.

**Algorithm.** The algorithm constructs $Q$ via a grid cycle $C$ representing $\partial Q$. Consider for all grid graph vertices a $1 \times 1$-square centered on the vertex, and let $C'$ be the cyclic chain of vertices whose square is intersected by $\partial P$, in the order in which $\partial P$ visits them (see Fig. 5). Note that $C'$ may contain duplicates. Now $C$ is obtained by iteratively finding a duplicate with minimal distance along the curve between the two occurrences, and removing the corresponding subchain. After all the duplicates are removed, the grid polygon $Q$ with boundary $C$ is returned, unless $C$ encloses no cells. In that case $C$ is a single vertex

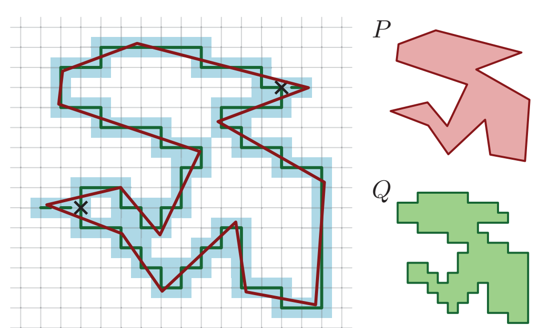



Figure 5: Example of the Fréchet algorithm; the input and output are shown on the right. The two crosses mark points appearing in $C$ twice, hence their subpaths (shown dashed) are removed.

or two vertices connected by one edge, and we let the grid polygon $Q$ consist of a single cell intersecting $P$.

**Upper bounds.** $Q$ cannot contain duplicate vertices, so it must be a grid polygon. Therefore we need to prove only the bound on the Fréchet distance.

**Theorem 6** *Given a $\beta$-obese polygon $P$ with $\beta \geq \sqrt{2}$, there exists a grid polygon $Q$ such that $d_F(\partial P, \partial Q) \leq (\beta + \sqrt{2})/2$.*

**Proof.** Consider $C$ (representing $\partial Q$) as obtained by the algorithm described above (ignoring the case where $C$ did not enclose cells). We will show that $d_F(\partial P, C) \leq (\beta + \sqrt{2})/2$. We will define a mapping between $\partial P$ and $C$ that gives rise to the parameterizations of $\partial P$ and $C$, needed to bound the Fréchet distance, and show that the distance between mapped points is at most $(\beta + \sqrt{2})/2$.

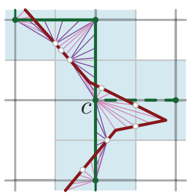


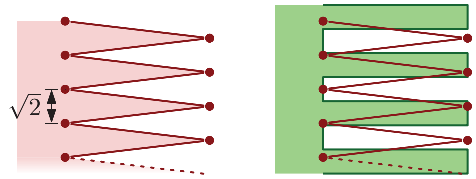

Figure 6: Mapping between $C$ and $\partial P$.

First we define a mapping between $\partial P$ and the vertices of $C'$ in the natural way: by proximity. We map the edges of $C'$ to points of $\partial P$, namely to the points where $\partial P$ intersects a boundary of a $1 \times 1$-square centered on a vertex of $C'$. This mapping is also simply by proximity. We convert this mapping into one between $\partial P$ and $C$: Whenever we remove a subchain from $C'$, that whole subchain is mapped to the vertex that is the start and end of that subchain (refer to Fig. 6). Once $C$ is obtained, only a single connected component of $\partial P$ is mapped to any vertex of $C$ and only one edge of $C$ is mapped to any point of $\partial P$. The resulting mapping is monotone by construction.

Consider any vertex $c$ of $C$. If $\partial P$ visits the $1 \times 1$-square $s$ of $c$ only once, then exactly the part of $\partial P$ inside $s$ maps to $c$, and the distance between $c$ and the part of $\partial P$ mapped to it is at most $\sqrt{2}/2$. If $\partial P$ visits $s$ twice, then the part of $\partial P$ outside $s$ between these visits is also mapped to $c$. The length of this boundary external to $s$ is at most $\beta$, so its furthest point is at most $\beta/2$ away from $s$ and hence at most $\beta/2 + \sqrt{2}/2$ from $c$, leading to the desired bound. When $\partial P$ visits $s$ more than twice, the same argument can be used.

Finally, the distance between edges of $C$ and points of $\partial P$ is at most $\sqrt{2}/2$, which is easy to see. □

**Lower bound.** Though we omit a full proof of our lower bound, its essence lies with constructing a polygon as sketched in Fig. 7(a). The border $\partial Q$ of a grid polygon with low Fréchet distance to $\partial P$ needs to follow the spikes in $\partial P$. However, as the grid is too coarse, there is not enough vertical space to do so (Fig. 7(b)). By using spikes of length linear in $\beta$, we get the bound claimed below in Theorem 7.




(a) (b)

Figure 7: A polygon (left) for which any grid polygon will have high Fréchet distance (right).

**Theorem 7** *For any $\beta > \sqrt{2}$, there exists a $\beta$-obese polygon $P$ for which for any grid polygon $Q$, $d_F(\partial P, \partial Q) \geq \frac{1}{4}\sqrt{\beta^2 - 2}$.*

## References


[1] H. Alt, B. Behrends, and J. Blömer. Approximate matching of polygonal shapes. *Ann. Math. Artif. Intell.*, 13(3-4):251–265, 1995.

[2] H. Alt and M. Godau. Computing the Fréchet distance between two polygonal curves. *Int. J. Comput. Geom. & Appl.*, 5:75–91, 1995.

[3] H. Alt, C. Knauer, and C. Wenk. Comparison of distance measures for planar curves. *Algorithmica*, 38(1):45–58, 2003.

[4] R. G. Cano, K. Buchin, T. Castermans, A. Pieterse, W. Sonke, and B. Speckmann. Mosaic drawings and cartograms. *Comp. Graph. Forum*, 34(3):361–370, 2015.

[5] J. Chun, M. Korman, M. Nöllenburg, and T. Tokuyama. Consistent digital rays. *Discr. & Comput. Geom.*, 42(3):359–378, 2009.

[6] M. de Berg, D. Halperin, and M. H. Overmars. An intersection-sensitive algorithm for snap rounding. *Comput. Geom.*, 36(3):159–165, 2007.

[7] M. T. Goodrich, L. J. Guibas, J. Hershberger, and P. J. Tanenbaum. Snap rounding line segments efficiently in two and three dimensions. In *Proc. 13th SoCG*, pages 284–293, 1997.

[8] D. H. Greene and F. F. Yao. Finite-resolution computational geometry. In *Proc. 27th FOCS*, pages 143–152, 1986.

[9] J. Hershberger. Stable snap rounding. *Comput. Geom.*, 46(4):403–416, 2013.

[10] R. Klette and A. Rosenfeld. Digital straightness – a review. *Discr. Appl. Math.*, 139(1-3):197–230, 2004.

[11] W. Meulemans. *Similarity Measures and Algorithms for Cartographic Schematization*. PhD thesis, Technische Universiteit Eindhoven, 2014.