Discrete Applied Mathematics 157 (2009) 3447–3456

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

A benchmark set for the reconstruction of hv-convex discrete sets

Péter Balázs *

Department of Image Processing and Computer Graphics, University of Szeged, Árpád tér 2., H-6720 Szeged, Hungary

ARTICLE INFO

Article history:
Received 1 September 2008
Received in revised form 26 January 2009
Accepted 20 February 2009
Available online 2 April 2009

Keywords:
Discrete tomography
hv-convex discrete set
Random generation
Reconstruction
Analysis of algorithms

ABSTRACT

In this paper we summarize the most important generation methods developed for the subclasses of hv-convex discrete sets. We also present some new generation techniques to complement the former ones thus making it possible to design a complete benchmark set for testing the performance of reconstruction algorithms on the class of hv-convex discrete sets and its subclasses. By using this benchmark set the paper also collects several statistics on hv-convex discrete sets, which are of great importance in the analysis of algorithms for reconstructing such kinds of discrete sets.

1. Introduction

The goal of Discrete Tomography (DT) [22,23] is to reconstruct discrete sets (finite subsets of the 2D integer lattice defined up to translation) from the number of its elements lying on parallel lattice lines along several (usually horizontal, vertical, diagonal, and antidiagonal) directions, called projections. It has several applications in pattern recognition, image processing, electron microscopy, angiography, non-destructive testing, and so on. The main challenge in DT is that practical limitations usually reduce the number of available projections to at most about four—which results in a possibly extremely large number of solutions of the same reconstruction task. This can cause the reconstructed discrete set to be quite different from the original one. In addition, the reconstruction problem can be NP-hard, depending on the number and directions of the projections. In certain cases one can facilitate the reconstruction task by supposing that the set to be reconstructed has some geometrical properties. Thus, the search space of the possible solutions can be reduced which can yield fast and less ambiguous reconstructions.

A common problem in Discrete Tomography arises in comparing reconstruction methods from the viewpoint of speed, accuracy, noise sensitivity, etc. In the past 15–20 years many reconstruction algorithms have been developed for solving the reconstruction problem by using different techniques. The average performance of those reconstruction algorithms were often tested on certain subclasses of hv-convex discrete sets. The reason of this is that the reconstruction in those classes has a well-developed theory including heuristics and exact reconstruction algorithms, as well as some important results regarding the complexity and ambiguity of the reconstruction. As an example, the reconstruction of hv-convex discrete sets from two projections is known to be NP-complete while it can be solved in polynomial time with the additional condition that the set is connected in the same time. The key to obtain an exact comparison of the average performance of different reconstruction algorithms is to develop uniform random generators for the studied classes. Unfortunately, for some subclasses of the hv-convex discrete sets no efficient method was known to generate elements of those classes by using uniform random distributions. In addition, even if there was a uniform generator for a certain class of discrete sets,

* Tel.: +36 62 546396; fax: +36 62 546397.
E-mail address: pbalazs@inf.u-szeged.hu.

0166-218X/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.dam.2009.02.019

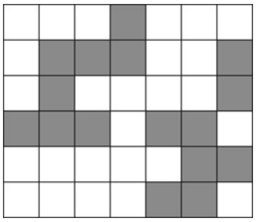

Fig. 1. A discrete set of size 6×7 represented by a binary picture.

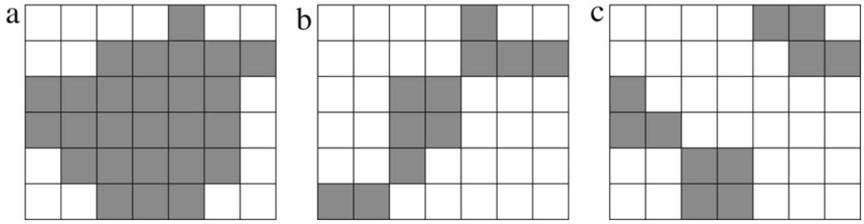

Fig. 2. (a) An hv-convex polyomino, (b) an hv-convex 8-connected but not 4-connected discrete set, and (c) a general hv-convex discrete set.

different authors used their own benchmark sets obtained by that uniform generator. Thus – until now – no exact overall comparison of the algorithms were possible.

In this paper we describe methods for generating elements of the most frequently studied subclasses of the hv-convex discrete sets—from uniform random distributions. This work is an extended version of [1] and it is strongly based also on the results of [2]. However, it is new in its concept that our aim is to design a complete benchmark set for the class of hv-convex discrete sets. Thus – as a main contribution – some new generation techniques are also proposed. The paper is structured as follows. First, the necessary definitions are introduced in Section 2. In Section 3 we describe methods for generating hv-convex discrete sets according to size and perimeter, possibly with certain additional properties. After that, in Section 4 we collect some statistics that can affect the complexity of several reconstruction algorithms developed for the hv-convex class. Section 5 concludes the paper.

2. Definitions

The finite subsets of the 2D integer lattice are called *discrete sets*. The *size* of a discrete set is defined by the size of its minimal bounding discrete rectangle (i.e. not the number of its elements). A discrete set F of size $m \times n$ is defined up to a translation and it is usually represented by a binary picture formed from unitary cells (see Fig. 1). We refer to the topmost row of the discrete set as the first row, and to the leftmost column of the set as the first column. Thus, the upper left corner of the minimal bounding rectangle of a discrete set is always the $(1, 1)$ position, and the remaining positions of the minimal bounding rectangle (and of the discrete set as well) are addressed consequently. Discrete sets with empty rows and/or columns are not of interest in this study.

A discrete set F is 4-*connected* (8-*connected*), if for any two positions $P \in F$ and $Q \in F$ of the set there exist a sequence of distinct positions $(i_0, j_0) = P, \dots, (i_k, j_k) = Q$ such that $(i_l, j_l) \in F$ and $|i_l - i_{l+1}| + |j_l - j_{l+1}| = 1$ ($\max\{|i_l - i_{l+1}|, |j_l - j_{l+1}|\} = 1$) for each $l = 0, \dots, k-1$. The 4-connected sets are also called polyominoes [21]. If the discrete set is not 4-connected then it consists of several polyominoes. The maximal 4-connected subsets of a discrete set F are called the *components of F*. Those components are always uniquely determined. For example, the discrete set in Fig. 1 has three components. A discrete set is called *horizontally and vertically convex* (shortly, hv-convex) if all the rows and columns of the set are 4-connected. Let us introduce the notations $\mathcal{P}$, $\mathcal{Q}$, and $\mathcal{HV}$ for the class of hv-convex polyominoes, hv-convex 8-connected discrete sets, and general hv-convex discrete sets, respectively. Obviously, $\mathcal{P} \subset \mathcal{Q} \subset \mathcal{HV}$. Fig. 2 shows some examples of discrete sets belonging to those classes.

A polyomino F is *northeast directed* (NE-directed for short) if there is a particular point $P \in F$ such that for each point $Q \in F$ there is a sequence $P_0 = P, \dots, P_t = Q$ of distinct points of F such that each point P_l of the sequence is north or east of P_{l-1} for each $l = 1, \dots, t$ (see Fig. 3a). Similar definitions can be given for SW-, SE-, and NW-directedness. An hv-convex polyomino is called *NW-parallelogram polyomino* if it is both NW- and SE-directed. Similarly, an hv-convex polyomino is called *NE-parallelogram polyomino* if it is both NE- and SW-directed (see Fig. 3b).

3. A benchmark set of hv-convex discrete sets

Although the reconstruction from two projections in the class of general hv-convex discrete sets is NP-complete [29] several methods can effectively solve this problem by applying some heuristic [25], metaheuristic [11,18,28] or

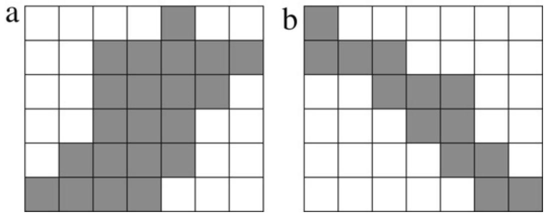

Fig. 3. (a) An hv-convex NE-directed polyomino, and (b) an hv-convex NW-parallelogram polyomino.

optimization [15] technique. Besides, for hv-convex polyominoes and hv-convex 8-connected sets different reconstruction algorithms have been developed to find one solution in polynomial time. The so-called *kernel-shell* method approximates the solution iteratively by a nondecreasing and a nonincreasing sequence at the same time [8,9,13]. This algorithm has a worst case time complexity of $O(mn \cdot \log mn \cdot \min\{m^2, n^2\})$. Another observation is that the reconstruction task can also be transformed into a 2SAT task [14,26] and then it is solvable in $O(mn \cdot \min\{m^2, n^2\})$ time in the worst case. The comparison of the average execution times of the two reconstruction approaches led to the design of a hybrid reconstruction algorithm [7] that has the same worst case time complexity of $O(mn \cdot \min\{m^2, n^2\})$ and remains fast in the average case as well. Recently, an algorithm has been also published that can perform the reconstruction in the class of hv-convex 8-connected but not 4-connected discrete sets in $O(mn \cdot \min\{m, n\})$ time [5] and which finds all the solutions of this class having a given pair of projections, yielding that the number of solutions is polynomial. On the contrary, it was proven that for certain pairs of projections there can be exponentially many hv-convex 4-connected sets having those projections [17].

These strong theoretical results regarding the number of solutions and the complexity of reconstruction caused the class of hv-convex discrete sets to become a class of main interest in Discrete Tomography. Elements of those classes are often used to test newly introduced reconstruction methods in order to gain information about the advances and drawbacks of the technique studied. For those tasks, there is a need for appropriately generated benchmark sets consisting of elements of a large variety from the hv-convex class.

3.1. hv-convex polyominoes

For the class $\mathcal{P}$ we already have closed formulas for enumerating hv-convex polyominoes according to several parameters. The semiperimeter (i.e. the half of the length of the boundary) of an hv-convex polyomino with size $m \times n$ is obviously $m + n$. In [16] it was proved that the number P_{n+4} of hv-convex polyominoes with a semiperimeter value of $n + 4$ is

$$P_{n+4} = (2n + 11)4^n - 4(2n + 1)\binom{2n}{n}. \tag{1}$$

Later, in [20] it was shown that the number $P_{m+1,n+1}$ of hv-convex polyominoes of size $(m + 1) \times (n + 1)$ is

$$P_{m+1,n+1} = \frac{m + n + mn}{m + n}\binom{2m + 2n}{2m} - \frac{2mn}{m + n}\binom{m + n}{m}^2. \tag{2}$$

In addition, in [24] the authors described a fast probabilistic method that generates hv-convex polyominoes having fixed perimeter with asymptotic probability 0.5. The method was extended in [7] to be able to generate hv-convex polyominoes with fixed size, as well.

3.2. hv-convex 8-connected discrete sets

Now, let us study the class $\mathcal{Q}$ of hv-convex 8-connected discrete sets. We first give a recursive formula for the number of hv-convex 8-connected discrete sets having a fixed semiperimeter. In order to do this we have to generalize the concept of semiperimeter for hv-convex 8-connected but not 4-connected discrete sets (the class defined by $\mathcal{Q} \setminus \mathcal{P}$), as well. Since the components of such a set are hv-convex polyominoes, this can be done in a straightforward way.

Definition 1. Let $F \in \mathcal{Q} \setminus \mathcal{P}$ having components $F_1, \ldots, F_k$. The semiperimeter of F is defined as the sum of the semiperimeters of all F_i $(i = 1, \ldots, k)$.

Moreover, we recall an already known fact about elements of the class $\mathcal{Q} \setminus \mathcal{P}$ from [5].

Proposition 2. *The smallest containing discrete rectangles of the components of an hv-convex 8- but not 4-connected discrete set are either connected to each other with their upper left and bottom right corners, or with their upper right and bottom left corners.*

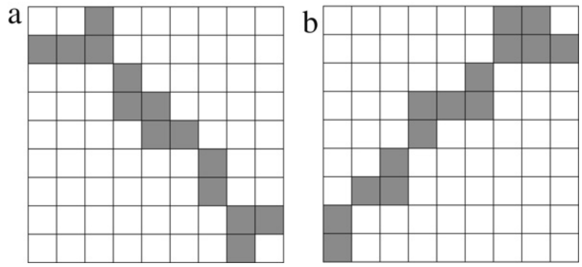

Fig. 4. An hv-convex 8- but not 4-connected set of (a) NW-type and (b) one of NE-type.

In the former case we say that the discrete set is of NW-type. Analogously, in the latter case we say that the set is of NE-type. Fig. 4 shows two examples of this.

Let us introduce the notations D_n, L_n, and Q_n for the number of NW-directed polyominoes, NW-parallelogram polyominoes, and hv-convex 8-connected discrete sets having semiperimeter n, respectively. Moreover let $\mathcal{T}$ denote the class of 8-connected discrete sets of NW-type whose components are all NW-parallelogram polyominoes (this time including sets consisting of a single NW-parallelogram polyomino, as well), and let T_n denote the number of sets in class $\mathcal{T}$ with semiperimeter n. With these notations we obtain $T_2 = L_2 = D_2 = 1$ and (for technical reasons by setting $T_0 = 1$) the following recursive formulas

Theorem 3. *For each* $n > 2$

$$T_n = L_n + \sum_{m=2}^{n-2} L_m T_{n-m} \tag{3}$$

and

$$Q_n = P_n + 2\sum_{m=2}^{n-2} \sum_{k=2}^{n-m} D_m D_k T_{n-m-k}. \tag{4}$$

Proof. If a discrete set $F \in \mathcal{T}$ with semiperimeter n has just one component then it is a NW-parallelogram polyomino with semiperimeter n. Otherwise, it contains a NW-parallelogram polyomino with semiperimeter m (where $1 < m < n-1$) as a subset in the upper left-hand corner and – following from Proposition 2 – the rest of F is a discrete set with semiperimeter $n-m$ which also belongs to the $\mathcal{T}$ class. This observation can be concisely expressed by the recursive formula (3).

An hv-convex 8-connected set of semiperimeter n is possibly an hv-convex polyomino which gives the first term on the right-hand side of (4) where P_n is defined by (1). Due to symmetry the number of discrete sets of $\mathcal{Q} \setminus \mathcal{P}$ of NW-type and of NE-type are the same. Therefore, it is sufficient to calculate the number of sets of NW-type having semiperimeter n and multiply the result by 2. Recalling from [5], for such a set of $\mathcal{Q}$ it is always true that $F_1, \ldots, F_{k-1}$ are NW-directed and $F_2, \ldots, F_k$ are SE-directed (that is, $F_2, \ldots, F_{k-1}$ are NW-parallelogram polyominoes). In particular, we also get that there are hv-convex 8-connected sets which have just two components and with no parallelogram polyominoes between them. Additionally, the structure of a set of $\mathcal{Q}$ of NW-type is the following. It contains an NW-directed polyomino with semiperimeter m in the upper left corner (where $1 < m < n-1$), an SE-directed polyomino with semiperimeter k in the bottom right corner (where $1 < k \leq n-m$) and the remaining part (if not empty) is a discrete set with semiperimeter $n-m-k$ belonging to the class $\mathcal{T}$ (see Fig. 4a). Note that the number of NW-directed polyominoes with semiperimeter n are exactly the same as the number of NE-directed polyominoes with semiperimeter n. Thus we get the formula (4). □

Considerations similar to Theorem 3 yielded recursive formulas for counting hv-convex 8- but not 4-connected discrete sets according to their size [1,2]. Based on those formulas in [2] an algorithm was also supplied to generate discrete sets of the class $\mathcal{Q} \setminus \mathcal{P}$ with given sizes by using uniform random distributions. Unfortunately, this algorithm generates sets of size $m \times n$ in $O(mn \cdot \min\{m, n\})$ time with an $O(m^2n^2)$ preprocessing time which makes the method inappropriate to generate sets of larger sizes.

Remark 4. The author of [2] proposed to generate the parallelogram components of given sizes of the hv-convex 8- but not 4-connected sets by generating simply directed hv-convex polyominoes with given sizes and then omit them if they are not parallelogram polyominoes. This method is quite time-consuming and therefore, when preparing our benchmark set, we used a more effective linear-time method that was described in [12].

Remark 5. In [5] the authors suggested to generate hv-convex 8- but not 4-connected discrete sets of a given size by using the method of rejection. They used the algorithm described in [6] to generate hv-convex 8-connected sets, and if the set was 4-connected as well then they omitted it. As we will see in Section 4, this method is also not suitable to generate discrete sets of larger sizes.

However, on the basis of Theorem 3 now we can outline a more effective random generator according to the semiperimeter of the hv-convex 8- but not 4-connected set.

Algorithm GENQ for generating hv-convex 8-connected but not 4-connected discrete sets from a uniform random distribution

Input: The integer n.
Output: An hv-convex 8- but not 4-connected set with semiperimeter n.

Step 1 calculate T_i for each $i = 1, \dots, n$;
Step 2 for $m = 2$ **to** $n - 2$
  for $k = 2$ **to** $n - m$
    calculate $D_m D_k T_{n-m-k}$;
Step 3 identify the semiperimeters u and b of the upper left and bottom right directed components, respectively, by choosing a number $r \in [1, Q_n]$ from a uniform random distribution;
Step 4 $t := n - u - r$;
Step 5 while $(t > 0)$
  { identify the semiperimeter p of the upper left NW-parallelogram component by choosing a number $r \in [1, T_t]$ from a uniform random distribution;
  $t := t - p$; }
Step 6 generate the components knowing their semiperimeters by using uniform random distributions;
Step 7 flip vertically the generated discrete set with 1/2 probability;

Theorem 6. *Algorithm GENQ generates a random* 8*- but not* 4*-connected hv-convex discrete set of semiperimeter n in $O(n \log n)$ time with $O(n^2)$ preprocessing time and $O(n^2)$ memory requirements.*

Proof. For the number of NW-directed (parallelogram) polyominoes we obtain from [10] the direct formulas

$$D_n = \binom{2(n-2)}{n-2}, \quad \text{and} \quad L_{n-1} = \frac{1}{n-1}\binom{2(n-2)}{n-2}. \tag{5}$$

The first two steps of Algorithm GENQ are for the preprocessing and they trivially can be performed in $O(n^2)$ time. We store all the increasing partial sums of the values calculated in these two steps (i.e. the values $L_1, L_1 + L_2 \cdot T_{n-2}, L_1 + L_2 \cdot T_{n-2} + L_3 \cdot T_{n-3}$ and so on, and similarly for the $D_m D_k T_{n-m-k}$'s where $m = 2, \dots, n-2$ and $k = 2, \dots, n-2-m$). For this we need $O(n^2)$ memory. The stored values give a unique partitioning of the interval $[1, Q_n]$ and $[1, T_t]$ Due to the storing of the increasing partial sums it can be decided in $O(\log n)$ time which of the intervals the randomly generated numbers of steps 3 and 5 fall into. Since Step 5 is iterated at most $n/2$ times, we get that the total execution time of Steps 3–5 is $O(n \log n)$. In Step 6 we generate an hv-convex NW-directed polyomino with semiperimeter u, an hv-convex SE-directed polyomino with semiperimeter b, and NW-parallelogram polyominoes with semiperimeters determined in Step 5. All these components can be generated in $O(n)$ time [16,10] and so the complexity part of the theorem follows.

Due to the construction, the generated sets are hv-convex 8- but not 4-connected sets of NW-type. Finally, in Step 7 we also take into account that the set to be generated can be either of NW-type or of NE-type with the same probability. □

For further details on the analysis of random generation algorithms of the above type the reader is referred to [19].

3.3. General hv-convex discrete sets

Regarding the generation of general hv-convex sets we first recall a concept of [2]. Let F be a discrete set with $k \geq 2$ components such that $I_l \times J_l = \{i_l, \dots, i'_l\} \times \{j_l, \dots j'_l\}$ is the minimal bounding rectangle of the lth component of F. We say that the components of F are *disjoint* if for any $1 \leq l, l' \leq k I_l \cap I_{l'} \neq \emptyset$ or $J_l \cap J_{l'} \neq \emptyset$ only if $l = l'$. Obviously, if an hv-convex discrete set has more than one components then they are disjoint. Now, without loss of generality we can assume that $i_l < i_{l+1}$ for each $l = 1, \dots, k-1$. F is called *canonical* if $j_l < j_{l+1}$ for each $l = 1, \dots, k-1$. That is, the discrete set is canonical if the minimal bounding rectangles of the components are connected to each other with their bottom right and upper left corners (see Fig. 5). The following proposition shows the connection of the general and canonical hv-convex discrete sets.

Proposition 7 ([1]). *Each canonical hv-convex discrete set with $k \geq 2$ components can be transformed into k! different general hv-convex discrete sets by using a suitable permutation of order k on the column sets of the components. Conversely, for a general hv-convex discrete set F with $k \geq 2$ components there exist exactly one hv-convex canonical discrete set F' and a uniquely determined permutation π of order k such that F' can be transformed into F by applying π on the column sets of F'.*

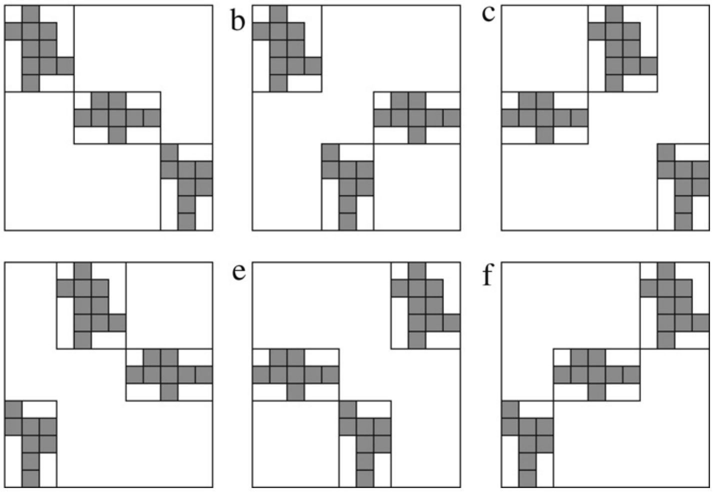

Fig. 5. (a) A canonical hv-convex discrete set, and (b)–(f) all the derived hv-convex discrete sets by applying the permutation (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1), respectively.

Fig. 5 represents the relation described in Proposition 7.

Now we can study the generation problem in the general class of hv-convex discrete sets. Let $C_n^{(t)}$ denote the number of canonical hv-convex discrete sets with semiperimeter n and exactly t components. Moreover, let HV_n denote the number of general hv-convex discrete sets having semiperimeter n. Then the following relations hold

Theorem 8. *For each $t > 1$ and $n > 2$*

$$C_n^{(t)} = \sum_{m=2}^{n-2} P_m C_{n-m}^{(t-1)} \tag{6}$$

and

$$HV_n = \sum_{t=1}^{\lfloor n/2 \rfloor} t! \cdot C_n^{(t)}. \tag{7}$$

Proof. Formula (6) can be proven similarly to Theorem 4 of [2], while Eq. (7) follows from Proposition 7 and the fact that the number of components can be at most $\lfloor n/2 \rfloor$. □

By setting the initial values $C_i^{(1)} = P_i$ $(i = 2, \ldots, n)$ and $C_i^{(t)} = 0$ when $i < \lfloor t/2 \rfloor$ on the basis of Theorem 8 an algorithm similar to Algorithm GENQ can be outlined to generate hv-convex discrete sets from uniform random distributions according to their semiperimeter. For more details we also suggest [2]. As a consequence we get

Theorem 9. *A random hv-convex discrete set of semiperimeter n can be generated from a uniform distribution in $O(n \log n)$ time with $O(n^2)$ preprocessing time and $O(n^2)$ memory requirements.*

With similar observations in [2] an algorithm was supplied to generate general hv-convex discrete sets of size $m \times n$ in $O(mn \cdot \min\{m, n\})$ time with $O(m^2 n^2 \min\{m, n\})$ preprocessing time. Unfortunately, due to the huge computational complexity this algorithm is not suitable to generate sets of larger sizes.

4. Statistics on hv-convex discrete sets

By applying the enumeration and generation methods presented in this paper we constructed a benchmark collection.[1] It consists of hv-convex discrete sets of various sizes and perimeters as given below

- 100-100 polyominoes of size $10 \times 10, 20 \times 20, \ldots, 100 \times 100$,
- 100-100 polyominoes of size $150 \times 150, 200 \times 200, \ldots, 500 \times 500$,
- 100-100 8- but not 4-connected sets of size $10 \times 10, 20 \times 20, \ldots, 100 \times 100$,

[1] The benchmark collection is available from the author upon request.

Table 1
The values of P_n, Q_n, and HV_n.

n	P_n	Q_n	HV_n
2	1	1	1
3	2	2	2
4	7	9	9
5	28	36	36
6	120	154	162
7	528	668	732
8	2 344	2 916	3 368
9	10 416	12 740	15 520
10	46 160	55 570	71 618
11	203 680	241 692	329 988
12	894 312	1 047 604	1 518 090
13	3 907 056	4 524 464	6 971 112
14	16 986 352	19 470 660	31 963 904
15	73 512 288	83 500 968	146 390 016

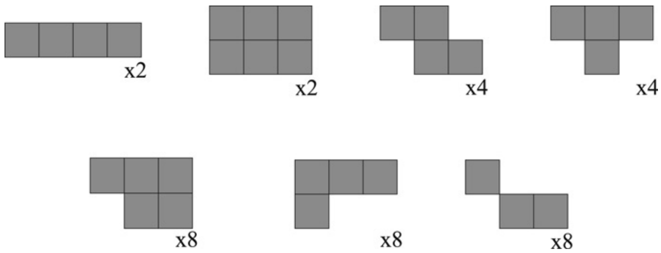

Fig. 6. Some hv-convex binary pictures with a semiperimeter value of 5. The numbers tell us that there are other solutions that can be obtained by mirroring or/and rotating the given discrete set.

- 100-100 8- but not 4-connected sets of semiperimeter 100, 200, ..., 1000,
- 100-100 general hv-convex sets of size 10×10, 20×20, ..., 100×100,
- 100-100 general hv-convex sets of semiperimeter 100, 200, ..., 1000.

The recursive formulas of Section 3 allow us to examine some important properties of hv-convex discrete sets that can affect the reconstruction complexity. In order to get such statistics we first calculated the number of hv-convex discrete sets in the classes studied. Table 1 shows the number of elements in the classes $\mathcal{P}$, $\mathcal{Q}$, and $\mathcal{HV}$ with semiperimeter n for the first 15 values of n – represented by P_n, Q_n, and HV_n, respectively (the first column can also be calculated via formula (1) and it enumerates the first 15 elements of Sequence A005436 in [27]). For $n = 5$ the corresponding hv-convex binary images are shown in Fig. 6.

Knowing that $\mathcal{P} \subset \mathcal{Q} \subset \mathcal{HV}$ and with the aid of the statistics presented in Table 1, we can describe the relative cardinality of the classes examined. With this information we can, for example, address questions concerning the relative occurrence of certain hv-convex discrete sets and calculate the probability that an hv-convex discrete set chosen from a uniform random distribution has some special properties which can facilitate the reconstruction task.

Example 10. Using the entries of Table 1 we can calculate the probability that an hv-convex discrete set with semiperimeter value of 6 chosen from a uniform random distribution is an hv-convex polyomino (i.e. it consists of one component), which turns out to be $120/162 \approx 0.74$. If we increase the semiperimeter value to 10, say, then this probability decreases to $46160/71618 \approx 0.64$. Such information is especially useful in the reconstruction task as hv-convex polyominoes can be reconstructed from their horizontal and vertical projections in polynomial time. In contrast, if the hv-convex set has at least two components then the reconstruction is NP-hard (see the introduction here). Hence with this method we can calculate the probability that the reconstruction of the randomly chosen hv-convex set can be performed using a polynomial-time algorithm to reconstruct an hv-convex polyomino.

Example 11. In [5] the authors presented a very fast algorithm for the reconstruction of hv-convex 8-connected but not 4-connected discrete sets. From the first few entries of Table 1 we have the suggestion that the number of such kinds of sets rapidly decreases as the semiperimeter value increases. To verify this, we calculated the first 100 values of P_n/Q_n (see Fig. 7). From this figure it is evident, that – unfortunately – even for sets of relatively small sizes there is almost no chance to apply this fast reconstruction algorithm in practice (assuming that the sets to be reconstructed are from a uniform random distribution), and things get worse if we want to reconstruct sets of bigger sizes. This observation gives an explanation why the method of rejection is not an effective way to generate hv-convex 8- but not 4-connected discrete sets (recall Remark 5).

It is also possible to describe the true distribution of the number of components of the generated hv-convex discrete set of the $\mathcal{HV}$ class since, in this case, we can enumerate the discrete sets of a given class that have exactly k components

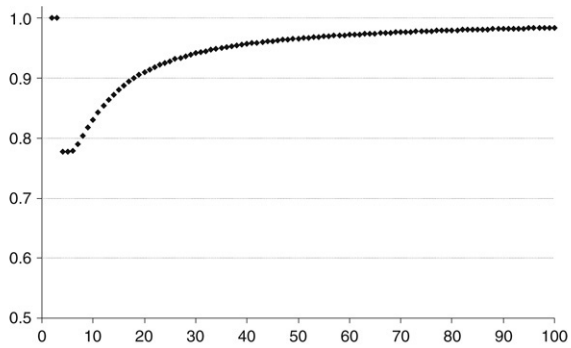

Fig. 7. The ratio P_n/Q_n (*vertical axis*) depending on the semiperimeter value n (*horizontal axis*).

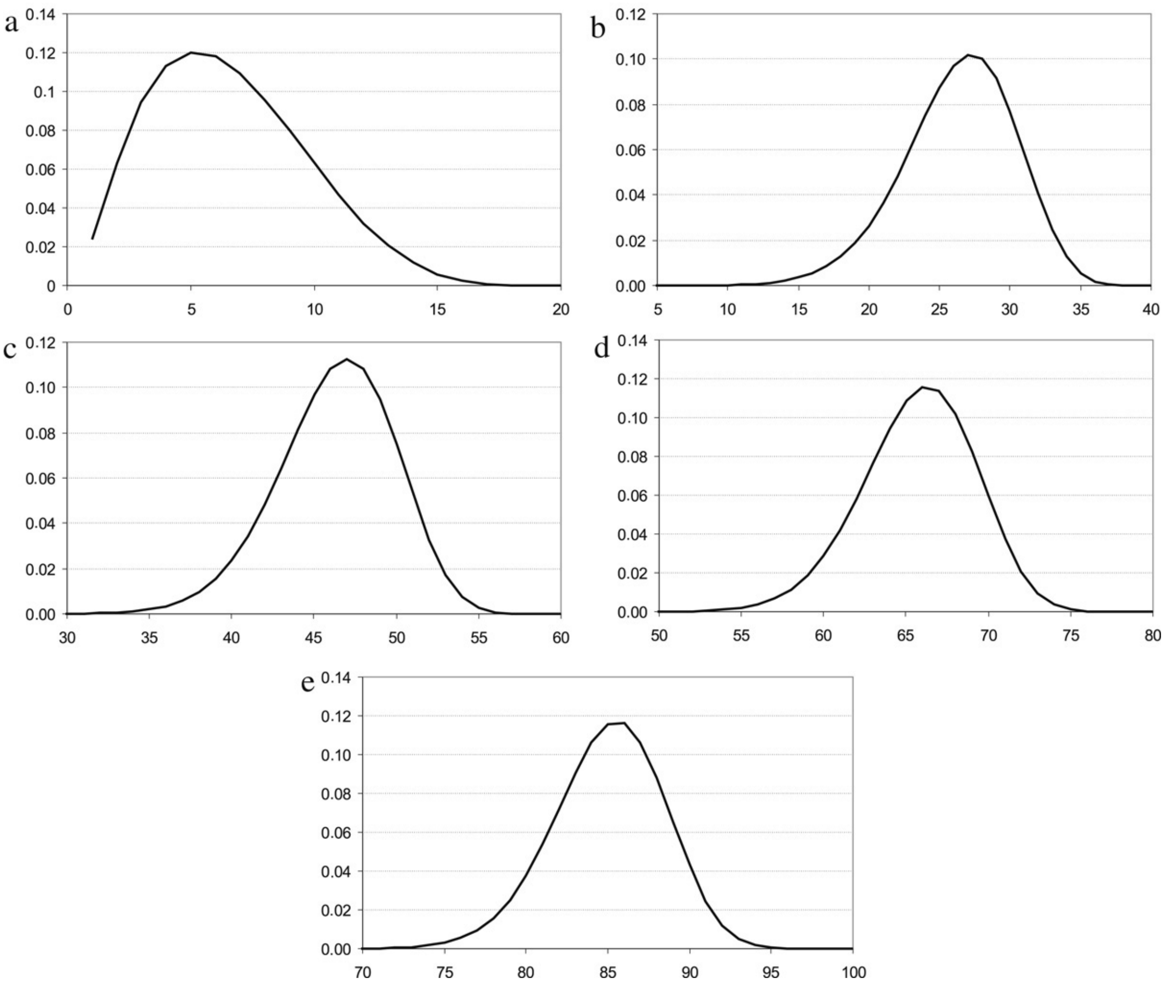

Fig. 8. The distributions of the number of components in the $\mathcal{HV}$ class for sets of size (a) 20×20, (b) 40×40, (c) 60×60, (d) 80×80, and (e) 100×100.

(see, e.g., formula (6)). Table 2 lists the expectation values and the variances of the variables which represent the number of components of a discrete set generated by using a uniform random distribution from the $\mathcal{HV}$ class when the size of the minimal bounding rectangle is $n \times n$ for some fixed positive integer n. In addition, the corresponding distributions are depicted in Fig. 8.

This piece of information can be very useful when reconstructing images like these. For example, the number of components of an hv-convex set also affects the accuracy of the reconstruction heuristic that was presented in [3]. Namely, more the components the hv-convex discrete set has, it is more likely that ambiguity will occur in the reconstruction.

Table 2
The expectation value $E_{HV}(n)$ and the variance $D^2_{HV}(n)$ of the number of components of a set with a minimal bounding rectangle of size $n \times n$ in the $\mathcal{HV}$ class. The values have been rounded to 5 digits.

n	$E_{HV}(n)$	$D^2_{HV}(n)$
20	6.53981	9.84446
40	26.33821	16.00766
60	46.30283	12.92260
80	65.70631	12.05665
100	84.99456	11.80716

Statistics about the expected number of components also opens the way to design reconstruction algorithms that exploit information known beforehand about the expected number of components [4].

Finally, with the aid of the benchmark set it is also possible to make some conjectures on certain properties of reconstruction tasks and algorithms. For example, the algorithm presented in [5] can find all the hv-convex 8-connected but not 4-connected discrete sets in polynomial time from two projections. This result implies that the number of discrete sets of the class $\mathcal{Q} \setminus \mathcal{P}$ with the same horizontal and vertical projections is bounded by a polynomial. However, when applied the reconstruction algorithm on the 8-connected but not 4-connected discrete sets of the benchmark collection we found that in each cases the number of different solutions having the same projections was at most 4. Even if we know from [5] that for the pair of vectors (1, 2, 3, 2, 1), (1, 2, 3, 2, 1) the number of solutions is 6, we have the following

Conjecture 12. *The number of different hv-convex 8- but not 4-connected discrete sets having given horizontal and vertical projections is at most 6.*

5. Conclusions

In this paper we have collected methods to generate hv-convex discrete sets (which possibly have certain connectedness properties as well). Besides, we also presented some new generation algorithms to complement the former ones in order to design a complete benchmark set for analysing the average performance of reconstruction algorithms developed for the class of hv-convex discrete sets and its frequently studied subclasses. The new generation methods are designed according to the semiperimeter of the set to be generated but they can be extended to generate discrete sets with fixed area (or with fixed area and semiperimeter) without any difficulty. In addition, they can be generalized to higher dimensions as well. By using the benchmark set supplied we have also collected some statistics on several subclasses of hv-convex discrete sets which can be used to analyse the efficacy (speed, accuracy, noise sensitivity, etc.) of certain reconstruction algorithms developed for the classes studied or to make some conjectures on other parameters, as well.

Acknowledgements

This work was supported by grant OTKA T048476. Preliminary version of the paper was presented at the 12th International Workshop on Combinatorial Image Analysis, Buffalo, NY, USA, April 2008 [1].

References

[1] P. Balázs, On the number of hv-convex discrete sets, in: V.E. Brimkov, R.P. Barneva, H.A. Hauptman (Eds.), Combinatorial Image Analysis, in: Lecture Notes in Computer Science, vol. 4958, Springer, Berlin, 2008, pp. 112–123.
[2] P. Balázs, A framework for generating some discrete sets with disjoint components by using uniform distributions, Theoret. Comput. Sci. 406 (2008) 15–23.
[3] P. Balázs, On the ambiguity of reconstructing hv-convex binary matrices with decomposable configurations, Acta Cybernet. 18 (3) (2008) 367–377.
[4] P. Balázs, Reconstruction of binary images with few disjoint components from two projections, Lecture Notes in Comput Sci. 5359 (2008) 1147–1156.
[5] P. Balázs, E. Balogh, A. Kuba, Reconstruction of 8-connected but not 4-connected hv-convex discrete sets, Discrete Appl. Math. 147 (2005) 149–168.
[6] E. Balogh, Generation and reconstruction of hv-convex 8-connected discrete sets, Acta Cybernet. 15 (2001) 185–200.
[7] E. Balogh, A. Kuba, Cs. Dévényi, A. Del Lungo, Comparison of algorithms for reconstructing hv-convex discrete sets, Linear Algebra Appl. 339 (2001) 23–35.
[8] E. Barcucci, A. Del Lungo, M. Nivat, R. Pinzani, Reconstructing convex polyominoes from horizontal and vertical projections, Theoret. Comput. Sci. 155 (1996) 321–347.
[9] E. Barcucci, A. Del Lungo, M. Nivat, R. Pinzani, Medians of polyominoes: A property for the reconstruction, Int. J. Imaging Syst. Technol. 9 (1998) 69–77.
[10] E. Barcucci, A. Frosini, S. Rinaldi, On directed-convex polyominoes in a rectangle, Discrete Math. 298 (2005) 62–78.
[11] K.J. Batenburg, An evolutionary algorithm for discrete tomography, Discrete Appl. Math. 151 (2005) 36–54.
[12] N. Bonichon, M. Mosbah, Watermelon uniform random generation with applications, Theoret. Comput. Sci. 307 (2003) 241–256.
[13] S. Brunetti, A. Del Lungo, F. Del Ristoro, A. Kuba, M. Nivat, Reconstruction of 4- and 8-connected convex discrete sets from row and column projections, Linear Algebra Appl. 339 (2001) 37–57.
[14] M. Chrobak, Ch. Dürr, Reconstructing hv-convex polyominoes from orthogonal projections, Inform. Process. Lett. 69 (6) (1999) 283–289.
[15] G. Dahl, T. Flatberg, Optimization and reconstruction of hv-convex (0, 1)-matrices, Discrete Appl. Math. 151 (2005) 93–105.
[16] M.P. Delest, G. Viennot, Algebraic languages and polyominoes enumeration, Theoret. Comput. Sci. 34 (1984) 169–206.
[17] A. Del Lungo, Polyominoes defined by two vectors, Theoret. Comput. Sci. 127 (1994) 187–198.
[18] V. Di Gesù, G. Lo Bosco, F. Millonzi, C. Valenti, A memetic algorithm for binary image reconstruction, Lecture Notes in Comput. Sci. 4958 (2008) 384–395.

[19] P. Flajolet, P. Zimmermann, B. Van Cutsem, A calculus for the random generation of labelled combinatorial structures, Theoret. Comput. Sci. 132 (1994) 1–35.
[20] I. Gessel, On the number of convex polyominoes, Ann. Sci. Math. Québec 24 (2000) 63–66.
[21] S.W. Golomb, Polyominoes, Charles Scriber's Sons, New York, 1965.
[22] G.T. Herman, A. Kuba (Eds.), Discrete Tomography: Foundations, Algorithms and Applications, Birkhäuser, Boston, 1999.
[23] G.T. Herman, A. Kuba (Eds.), Advances in Discrete Tomography and its Applications, Birkhäuser, Boston, 2007.
[24] W. Hochstättler, M. Loebl, C. Moll, Generating convex polyominoes at random, Discrete Math. 153 (1996) 165–176.
[25] A. Kuba, The reconstruction of two-directionally connected binary patterns from their two orthogonal projections, Comput. Vis. Graph. Image Process. 27 (1984) 249–265.
[26] A. Kuba, Reconstruction in different classes of 2D discrete sets, Lecture Notes in Comput. Sci. 1568 (1999) 153–163.
[27] N.J.A. Sloane, The on-line encyclopedia of integer sequences, http://www.research.att.com/~njas/sequences/.
[28] C. Valenti, A genetic algorithm for discrete tomography reconstruction, Genet. Program. Evolvable Mach. 9 (2008) 85–96.
[29] G.W. Woeginger, The reconstruction of polyominoes from their orthogonal projections, Inform. Process. Lett. 77 (2001) 225–229.