Doctopic: Algorithms, automata, complexity and games ARTICLE IN PRESS TCS:10819

Theoretical Computer Science ••• (••••) •••–•••

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Two-dimensional comma-free and cylindric codes ☆

Marcella Anselmo [a], Maria Madonia [b]

[a] Dip. di Informatica, V. G. Paolo II, 132, Università di Salerno, I-84084 Fisciano (SA), Italy
[b] Dip. Matematica e Informatica, Università di Catania, Viale Andrea Doria 6/a, 95125 Catania, Italy

A R T I C L E I N F O

Article history:
Received 7 November 2015
Received in revised form 19 May 2016
Accepted 10 June 2016
Available online xxxx

Keywords:
Two-dimensional languages
Codes
Comma-free codes
Circular codes

A B S T R A C T

A two-dimensional code of pictures is defined as a set $X \subseteq \Sigma^{**}$ such that any picture over Σ is tilable in at most one way with pictures in X. It has been proved that it is undecidable whether a finite set of pictures is a code. Here we introduce two classes of picture codes: the comma-free codes and the cylindric codes, with the aim of generalizing the definitions of comma-free (or self-synchronizing) code and circular code of strings. The properties of these classes are studied and compared, in particular in relation to maximality and completeness. As a byproduct, we introduce self-covering pictures and study their periodicity issues.

1. Introduction

The theory of variable-length codes is an important part of computer science both in its theoretical and practical issues. The theory is strongly related to combinatorics on words, formal languages, automata and semigroup theory. The aim is to find structural properties of codes to be exploited for their construction. We refer to [10] for complete references.

During the last fifty years, some significant work has been done to transfer the formalisms and the results from the string language theory into the two-dimensional (2D) world (see for example [8,11,17,18,29]). The extension of the classical notion of string (or word) to two dimensions leads to the definition of polyomino, in its different declinations – labeled polyominoes, directed polyominoes, as well as rectangular labeled polyominoes, that we will refer to as *pictures*. In the literature, one can find different attempts to generalize the notion of code to 2D objects. A set C of polyominoes is a *code* if every polyomino is tilable in at most one way with (copies of) elements of C. The results show that in the 2D context most important properties are lost. A major result, due to D. Beauquier and M. Nivat, states that the problem whether a finite set of polyominoes is a code is undecidable, and the same result holds for dominoes, too ([9]). Related particular cases were studied in [1]. In [21], the codes of directed polyominoes equipped with catenation operations are considered, and a few special decidable cases are detected. The codes of labeled polyominoes, also called bricks, are studied in [26], and some further undecidability results are proved. In relation to the operations of row and column concatenations and in connection with doubly-ranked monoids, a definition of code of pictures (rectangular arrays of symbols) is given in [13] with the main goal of extending syntactic properties to two dimensions. Unfortunately most of the results are negative.

In this paper we consider the definition of code of pictures introduced in [4], which refers to the operation of tiling star. The *tiling star* of a language X, as defined in [29], is the set X^{**} of all pictures that are tilable (in the polyominoes style)

☆ Partially supported by MIUR Project 2010LYA9RH *"Automata and Formal Languages: Mathematical and Applicative Aspects"*, MIUR Project *"PRISMA PON04a2 A/F"*, University of Salerno FARB Project ORSA138754, and University of Catania PRA Project 2012.

E-mail addresses: anselmo@dia.unisa.it (M. Anselmo), madonia@dmi.unict.it (M. Madonia).

http://dx.doi.org/10.1016/j.tcs.2016.06.016
0304-3975/© 2016 Elsevier B.V. All rights reserved.

Please cite this article in press as: M. Anselmo, M. Madonia, Two-dimensional comma-free and cylindric codes, Theoret. Comput. Sci. (2016), http://dx.doi.org/10.1016/j.tcs.2016.06.016

by elements of X. Then, X is a code if any picture in X^{**} is tilable in a unique way. In [4], it has been proved that it is undecidable whether a given set of pictures is a code. This is actually not surprising, because it is coherent with the result that it is undecidable whether a picture language can be (tiling system) recognized in an unambiguous way ([7]).

The search for decidable classes of picture codes has followed the footprints of the well established theory of string codes. Two generalizations of the notion of prefix code have been proposed and investigated in [4,6], the prefix codes and the strong prefix codes. Subsequently, the codes with finite deciphering delay have been introduced ([3]). All these classes have good decidability properties; it is decidable whether a finite language is a prefix code, or a strong prefix code, or a code with a given finite deciphering delay. It is meaningful to note that in these definitions the pictures are considered with a preferred scanning direction from a corner to the opposite one. For example, if X is a (top-left to bottom-right) prefix code, the decoding process of a picture p starts on its top-left corner, and proceeds towards the bottom right corner, in such a way that one can decide without ambiguity which is the next element in X.

This paper continues this line of investigation and considers two meaningful classes of picture codes, namely comma-free and cylindric codes, which extend comma-free and circular codes of strings. The novelty lies in the fact that the definitions are "non-oriented", in the sense that they do not follow a specific decoding direction.

Let us recall the origin of comma-free codes of strings. Taking inspiration from some biological mechanisms in the information transmission via DNA, in 1958, Golomb, Gordon, and Welch [19] introduced the general concept of comma-free code and studied the quantitative aspects of comma-free codes. As a matter of fact, the authors of the early period of investigation, have considered largely the comma-free codes of constant length and under the need and influence of biology. But, it appeared several years later that the biological code is not a comma-free code, not even a code in the general sense. Nevertheless, the comma-free code has deservedly entered into coding theory. In the first years, the main line of study was concerned with the maximal size of comma-free codes of strings of a fixed length and on alphabets with a fixed number of letters. The research on variable-length comma-free codes was initiated later in 1969 [28]. The problems of completion and of finite completion of comma-free codes have been solved by N.H. Lam [22–24]. The comma-free codes of strings are also called self-synchronizing codes, due to the following property. They have an easy deciphering: if X is comma-free and in a word $w \in X^*$, some factor can be identified to be in X, then this factor is one term of the unique factorization of w on X.

The generalization of comma-free codes of strings to higher dimensions has been considered in [14]. A q-dimensional comma-free code is a set D of q-dimensional words or arrays of letters of fixed size, with the property that for any arrangement of the arrays in D on a plane, no "proper" factor is in D. The author shows a (not tight) bound on the cardinality of q-dimensional comma-free codes (note that comma-free codes are necessarily finite with this definition).

In this paper we introduce variable-length two-dimensional *comma-free sets*. Hence, a comma-free set can be infinite. A set of pictures is comma-free if no picture in X can be covered by the pictures in X. Any comma-free set is a code, that we call a *comma-free code*. The definition preserves the main property of self-synchronizability. If $X \subseteq \Sigma^{**}$ is a comma-free code and in a picture $p \in X^{**}$, some factor can be identified to be in X, then this factor is one term of the unique tiling decomposition of p on X. In a simple way, it is decidable whether a finite set is a comma-free set. Therefore finite comma-free codes are a first example of a decidable class of codes of pictures with a "non-oriented" definition.

In the theory of string codes, comma-free codes are studied inside the class of circular codes. The translation of the definition of circular code of strings into the two-dimensional world leads to some new situations to deal with. The role of a circle can be played in 2D by a cylinder, either horizontally or vertically placed. Now, a set X of pictures is a vertical *cylindric code* if the pictures of X cannot tile the lateral surface of any cylinder (for any height and radius) in two different ways. A main result is that it is undecidable whether a (finite) set of pictures is a cylindric code. This result shows a crucial difference with the comma-free codes. This is not really surprising, since the definition of comma-free code is based on a "local" property (that can be tested on pictures), while the definition of cylindric code is founded on a more "global" and broader property. Subsequently, the maximality of such notions of code is investigated and related to a new notion of completeness. Finally, the relationships among all the introduced classes are shown.

The investigation on comma-free sets has led us to consider *(non-) self-covering pictures*. A picture p is self-covering if it can be completely covered with some juxtaposed copies of itself. This notion deserves a separate consideration, since it represents a sort of periodicity on pictures. Note that matrix periodicity plays a fundamental role in two-dimensional pattern matching. The notion of periodicity in matrices and pictures has been investigated in [2,15,27], while two-dimensional quasi-periodicity was very recently studied in [16]. The definition starts from the consideration that in 1D a string is periodic if it self-overlaps across the middle point. Hence, in 2D a picture has been defined as periodic if it self-overlaps and this overlapping includes the center. The repetition of the self-overlapping produces some periodicity in the picture along one or more directions. Note that such periodicity may leave out some border positions (what is called a "free zone" in [27]) which may present no relation with the content in the rest of the picture. In [25], the well-known Fine and Wilf's Theorem on strings is generalized to two-dimensional words on a large class of convex domains through a geometric approach.

The definition of self-covering pictures, as considered here, gives rise to a different notion of periodicity. A picture p is self-covering if it can be completely covered with some juxtaposed copies of itself. Then, the repetition of such overlapping gives raise to a periodicity in all positions of p, without "free zones". Differently said, a self-covering picture can be cut out from an infinite periodic configuration with a "small" period. Note that in 1D, a self-covering string is the power of some of its prefixes (without final extra symbols); hence the self-covering strings correspond to a stronger kind of periodic strings. The notion of non-self-coverability corresponds to a 2D notion of aperiodicity. Non-self-covering pictures are special

Please cite this article in press as: M. Anselmo, M. Madonia, Two-dimensional comma-free and cylindric codes, Theoret. Comput. Sci. (2016), http://dx.doi.org/10.1016/j.tcs.2016.06.016