Available online at www.sciencedirect.com
SCIENCE DIRECT®
Theoretical Computer Science 340 (2005) 364–380
Theoretical Computer Science
www.elsevier.com/locate/tcs

Collage of two-dimensional words

Christian Choffrut*, Berke Durak
Université Paris VII, L.I.A.F.A., 2 Place Jussieu, 75221 Paris, France

Abstract

We consider a new operation on one-dimensional (resp. two-dimensional) word languages, obtained by piling up, one on top of the other, words of a given recognizable language (resp. two-dimensional recognizable language) on a previously empty one-dimensional (resp. two-dimensional) array. The resulting language is the set of words "seen from above": a position in the array is labeled by the topmost letter. We show that in the one-dimensional case, the language is always recognizable. This is no longer true in the two-dimensional case which is shown by a counter-example, and we investigate in which particular cases the result may still hold.

Keywords: Regular languages; Picture languages

1. Introduction

The present paper deals with the notion of recognizable collection of pictures, a picture being a matrix whose entries (pixels) are taken in a finite alphabet (colors). The reader unfamiliar with the formal definition might find it suggestive to think of the set of chessboards of arbitrary dimension or of the set of squares with, say, their north-west to south-east diagonal marked with some particular color, as typical examples.

Assume we are given a collection of strips of wallpapers of different textures in such a way that it forms a recognizable collection. Assume further that starting from an empty frame we can paste these strips one at a time, in any arbitrary way, with possible overlapping but without rotation. At each position, the visible pixel is that belonging to the last pasted

* Corresponding author.
E-mail addresses: cc@liafa.jussieu.fr (C. Choffrut), durak@liafa.jussieu.fr (B. Durak)
URLs: http://www.liafa.jussieu.fr/∼cc (C. Choffrut), http://www.liafa.jussieu.fr/∼durak.

0304-3975/$ - see front matter © 2005 Published by Elsevier B.V.
doi:10.1016/j.tcs.2005.03.034

strip. This is reminiscent of the so-called painter's algorithm achieving face elimination in computer graphics where the objects nearest to the observer are painted last. Our result says that if we start with a recognizable collection of strips reduced to one column (resp. to one row), then all possible collages form again a recognizable collection. This property is obtained by studying the particular case of one-dimensional pictures, i.e., words, and by extending to two-dimensional pictures via row- (or column-) Kleene concatenation. Furthermore, we show that this closure property no longer holds when this hypothesis fails; using a counting argument, we show that there exists a finite language consisting of two strips whose collage is not recognizable. There exist general simple conditions guaranteeing recognizability of the collage in terms of the parameters of the collage, such as the maximum number of levels of strips. In the case where the alphabet is unary, yielding thus binary pictures with a color for the background and a color for the foreground, the collage is recognizable whatever the collection of strips (it may even be non-recursive).

As far as we know, the operation of collage as we mean it here is new. In [7, Proposition 5.1.], the author considers the operation consisting of tiling a picture with non-overlapping strips and shows a closure property for recognizable pictures. Concerning one-dimensional pictures, the notion of quasiperiodicity, which is remotely connected to our notion of collage, was introduced in [1]. In our terminology it is a collage of a unidimensional picture with a unique strip as explained above and where the overlapping occurrences of the strip are obliged to match. A final word of caution though: the term collage was coined in [5] as a means of defining pictures via recursive geometric functions in the spirit of fractals [2]. We use it here in a different meaning which we think appropriate for its kinship with the art movement in painting of the first decades of the 20th century.

2. The unidimensional case

Given a finite alphabet Σ, we denote by Σ^* the free monoid of *words* or *strings* over Σ, and by ε the empty string. The *product* or *concatenate* of two words u and v is simply denoted by uv. For a string $w \in \Sigma^*$, we denote by $|w|$ its length and by $w[i]$ the ith symbol of w, $i = 1, \ldots, |w|$. A string $z \in \Sigma^*$ is a *subword* or *factor* of w if there exist two strings $u, v \in \Sigma^*$ such that $w = uzv$ and we write $z = w[i \ldots j]$ where $|u| = i - 1$ and $|uz| = j$. If $t \in \Sigma^*$ has the same length as z, the substitution of t for z in w results in the word utv which we write $w \to utv$. We say that u is *placed at position i*. The notations $\overset{r}{\to}$ for the rth iterate and $\overset{*}{\to}$ for the reflexive and transitive closure of $\to$ are used with their standard meaning. Given a subset $W \subseteq \Sigma^*$ of *patches*, the operation of collage consists of producing words in $(\Sigma \cup \{\square\})^*$ ($\square$ is a new symbol not in Σ) by starting with a word of the form $\square^n$ and then repeatedly replacing random factors of the current word with elements of W. A word thus obtained is called a *collage* of W. Formally $\mathtt{C}^0(W) = \square^*$ and for all $k \geqslant 0$

$$\mathtt{C}^{k+1}(W) = \{w' \mid \exists w \in \mathtt{C}^k(W), w \to w'\}.$$

The set of collages of W is the union $\mathtt{Collage}(W) = \bigcup_{k \geqslant 0} \mathtt{C}^k(W)$. We say position $0 < j \leqslant n$ of $w \in \mathtt{C}^k(W)$ is *covered by an occurrence* $u \in W$ whenever there exists an