ELSEVIER

Pattern Recognition Letters 16 (1995) 1137–1145

Pattern Recognition Letters

A random model for analyzing region quadtrees

Michael Vassilakopoulos [a], Yannis Manolopoulos [b,*,1]

[a] Department of Electrical & Computer Engineering, Aristotelian University of Thessaloniki, GR-540 06, Greece

[b] Department of Informatics, Aristotelian University of Thessaloniki, GR-540 06, Greece

Received 14 June 1994; revised 1 June 1995

Abstract

A new random tree model, suitable for analyzing the storage efficiency of region quadtrees, is introduced. For certain values of parameters this model is proved to be the formal equivalent of an older popular random tree model. A formula that expresses the average number of nodes in a region quadtree obeying this model is also given and proved in two different ways.

Keywords: Data structures; Average storage analysis; Probabilistic models of trees and images

1. Introduction

The region quadtree (Samet, 1990a; Samet, 1990b) is a very popular hierarchical data structure for the representation of binary images. We can view such an image as a $2^n \times 2^n$ binary array, for some natural number n, where an entry equal to 0 stands for a white pixel and an entry equal to 1 stands for a black pixel. If every pixel of this image is white (black), its quadtree is made up of a single white (black) node. If, however, the image is not unicolor, its quadtree is made up of a grey root that points to four children (subtrees). Each of these subtrees is a quadtree that represents a quadrant of the image. We assume here, that the first (leftmost) child of the root corresponds to the North-West quadrant, the second to the North-East quadrant, the third to the South-West quadrant and the fourth (rightmost) child of the root to the South-East quadrant of the image.

An example of an 8 by 8 binary image and its quadtree are shown in Figs. 1(a) and 1(c), respectively. Note that black (white) squares represent black (white) leaves, while circles represent grey nodes. The unicolor blocks, into which this image is partitioned by the quadtree external nodes, are depicted in Fig. 1(b). Recall that the quadtree for a 2^n by 2^n image is of height n, at most. Let us call such a tree a class-n quadtree. Note also, that there are $2^{(4^n)}$ different images of this size. A node corresponding to a single pixel is at level 0, while the root is at level n. A node at level i, where $0 \leqslant i \leqslant n$, represents a subarray of $2^i \times 2^i$ $(= 4^i)$ pixels; there are at most 4^{n-i} nodes at this level.

* Corresponding author. Email: manolopo@eng.auth.gr

[1] Research performed while the author was on sabbatical leave with the Computer Science Department, University of Maryland at College Park, MD 20742, USA.

SSDI 0167-8655(95)00072-0

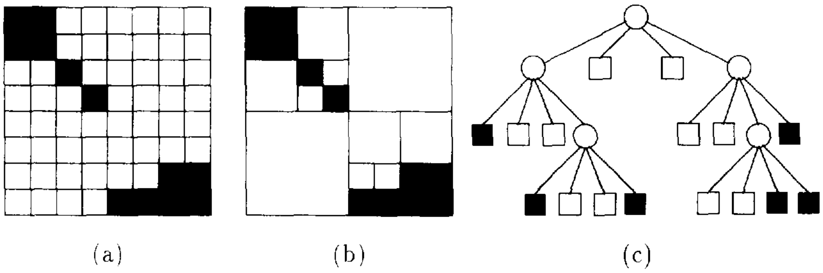

Fig. 1. (a) An 8 by 8 binary image. (b) The partitioning into unicolor quadrangular blocks. (c) Its quadtree.

The average storage efficiency of quadtrees has been studied in (Mathieu et al., 1987), where analysis is based on a model of random trees. It has also been studied in (Vassilakopoulos and Manolopoulos, 1993), where analysis is based on a model of random images and on the assumption that the sizes of internal and external nodes differ. However, random trees or images are not the only viewpoints for studying the quadtree storage-requirements. Analogous results for the representation of arbitrary iso-oriented rectangles by quadtrees appear in (Faloutsos, 1992; Shaffer, 1988) and for the representation of arbitrary curves or regions by quadtrees in (Burton et al., 1985). Worst-case results for the distribution of black nodes at various levels in a quadtree appear in (Unnikrishnan et al., 1987).

In the present article we introduce a new random tree model, which is based on a set-theoretic recursive description of region quadtrees. Among other approaches which use recursive techniques, we note the work reported in (Flajolet et al., 1993) which gives stochastic divide-and-conquer recurrences analyzing the serching efficiency of point quadtrees. In this article, also, we prove that, for certain values of parameters, our model expresses formally the popular random tree model of Shaffer and Samet (1982; 1985). Besides, we prove in two different ways a formula that gives the average number of nodes for a quadtree obeying this model.

2. Models of randomness

Shaffer and Samet have given a descriptive definition of a model of random quadtrees (Samet, 1982; Samet and Shaffer, 1985). According to this model "*each leaf node is assumed to be equally likely to appear at any position and level in the tree*". We characterize this model as a random tree model, since the probabilistic assumptions concern tree nodes and not image blocks. It is generally considered a very realistic model for images usually found in practice. It has also been used for the analysis of neighbor finding algorithms (Samet, 1982; Samet and Shaffer, 1985), producing results close to statistics of real tests. However, there has never been presented a formal equivalent of the above definition.

The random model used in (Mathieu et al., 1987) also assigns probabilities to the nodes of the quadtree. According to this model a random tree is built by using a branching process that starts at the root (level n). In general, a level-i node is colored black or white with probability b_i in both cases; otherwise, it is colored grey with probability $1-2b_i$. In case a node is grey, the branching process continues for each of the four children of this node. Since single pixels can not be grey, we must have $b_0 = 1/2$. For nodes at higher levels we must have $0 \leqslant b_i \leqslant 1/2$. Note that according to the definition of this model a grey node may have four children that are all black or white (the branching process is not restricted so as not to produce four sibling leaves of the same color). This quadtree variation is of no practical use, since the memory savings introduced by creating maximal white and black blocks are lost.

Another model is based on pixels (Vassilakopoulos and Manolopoulos, 1993) (this is a random image model). A pixel is black with probability p and white with probability $1-p$, independently of any other pixel. This means that a block which corresponds to a node at level i is black with probability $p^{(4^i)}$, white

Fig. 2. The four different sets of sub-quadtrees represented by the notation $\langle \blacksquare\square\square\square \rangle$.

with probability $(1-p)^{(4^i)}$ and grey with probability $1-p^{(4^i)}-(1-p)^{(4^i)}$. When $p=0.5$ an image mainly consists of very small unicolor blocks and a quadtree is not an efficient representation, since it is almost a full structure (leaf nodes appear almost exclusively at level 0). However, when p differs significantly from 0.5, the probability of unicolor blocks that are larger than 1×1 pixels increases (for one of the colors). This means that the random image model can express coherence to some extend and represent images found in practice (for example biological images). We also note reference (Cross and Jain, 1983) which reports an image model, based on Random Markov Fields, expressing coherence well.

3. A new model of random quadtrees

First, we examine the set-theoretic representation of quadtrees. Note that we can view a region quadtree as a finite ordered tuple that consists of grey, black and white nodes and corresponds to the preorder traversal of the tree. With this in mind, we can devise a symbolic way to represent the set of all class-n quadtrees, Q_n (a finite set having as elements tuples), as has been done for other trees and combinatorial structures (Vitter and Flajolet, 1990).

For the sake of simplicity, in this section the symbols $\bigcirc$, $\square$ and $\blacksquare$ will denote the sets $\{\bigcirc\}$, $\{\square\}$ and $\{\blacksquare\}$, respectively. As *class-i sub-quadtree* we characterize a class-i quadtree that has at least one grey node. We will use the symbol S_i to denote the set of all class-i sub-quadtrees ($n \geqslant i \geqslant 1$). We are interested in a certain kind of sets of class-i sub-quadtrees: such a set is a cartesian product of $\bigcirc$ and four other sets, $T_{i,1}$, $T_{i,2}$, $T_{i,3}$ and $T_{i,4}$, where each of $T_{i,1}$, $T_{i,2}$, $T_{i,3}$ and $T_{i,4}$ is either a subset of S_{i-1} or a leaf. We will refer to such a set as a *Cartesian i-Set*. A class-i sub-quadtree that belongs in such a cartesian i-set consists of a grey root and its NW, NE, SW and SE subtrees which belong in $T_{i,1}$, $T_{i,2}$, $T_{i,3}$ and $T_{i,4}$, respectively. We shall use the notation $\langle T_{i,1}, T_{i,2}, T_{i,3}, T_{i,4}\rangle$ to denote the union of all the cartesian i-sets that are defined as a cartesian product of $\bigcirc$ and each one of the distinct permutations of $T_{i,1}$, $T_{i,2}$, $T_{i,3}$ and $T_{i,4}$. An example showing the four different sets of sub-quadtrees whose union is represented by the notation $\langle \blacksquare\square\square\square \rangle$ appears in Fig. 2. Note that each of these sets includes only one sub-quadtree. In addition, in the symbolic equations that follow the $+$ symbol represents set union. Thus, the set of all class-n quadtrees is:

$$Q_n = \square + \blacksquare + S_n,$$

$$\begin{aligned} S_i = {} & \langle \blacksquare\square\square\square \rangle + \langle \blacksquare\blacksquare\square\square \rangle + \langle \blacksquare\blacksquare\blacksquare\square \rangle \\ & + \langle S_{i-1}\square\square\square \rangle + \langle S_{i-1}\blacksquare\square\square \rangle + \langle S_{i-1}\blacksquare\blacksquare\square \rangle + \langle S_{i-1}\blacksquare\blacksquare\blacksquare \rangle \\ & + \langle S_{i-1}S_{i-1}\square\square \rangle + \langle S_{i-1}S_{i-1}\blacksquare\square \rangle + \langle S_{i-1}S_{i-1}\blacksquare\blacksquare \rangle \\ & + \langle S_{i-1}S_{i-1}S_{i-1}\square \rangle + \langle S_{i-1}S_{i-1}S_{i-1}\blacksquare \rangle + \langle S_{i-1}S_{i-1}S_{i-1}S_{i-1} \rangle \quad \forall i > 1, \end{aligned}$$

$$S_1 = \langle \blacksquare\square\square\square \rangle + \langle \blacksquare\blacksquare\square\square \rangle + \langle \blacksquare\blacksquare\blacksquare\square \rangle.$$

In the tables of Fig. 3 you can see the number of different cartesian i-sets whose union is represented by each of the $\langle\rangle$ notations above. These are all the possible $\langle\rangle$ notations of unions of cartesian i-sets in the sense that any other $\langle\rangle$ notation is equivalent to one of them. It is not difficult to see that Q_n includes all the possible class-n quadtrees since the definition of Q_n is a top-down constructive definition where all the possible

< ■□□□ >	4
< ■■□□ >	6
< ■■■□ >	4

< S_{i-1}□□□ >	4
< S_{i-1}■□□ >	12
< S_{i-1}■■□ >	12
< S_{i-1}■■■ >	4
< $S_{i-1}S_{i-1}S_{i-1}S_{i-1}$ >	1

< $S_{i-1}S_{i-1}$□□ >	6
< $S_{i-1}S_{i-1}$■□ >	12
< $S_{i-1}S_{i-1}$■■ >	6
< $S_{i-1}S_{i-1}S_{i-1}$□ >	4
< $S_{i-1}S_{i-1}S_{i-1}$■ >	4

Fig. 3. All possible $\langle\rangle$ notations and the number of different subtree sets each one represents.

configurations for the children of a grey node appear (excluding the illegal configuration where all four children are unicolor leaves).

Note that the above symbolic equations (when fully expanded) define the set of all class-n quadtrees as a large expression made of union and cartesian product operators and a number of only three initial operands, the sets O, ■ and □. In addition, these symbolic equations describe a branching process by which we can construct any legal class-n region quadtree. More specifically,

- At the start, we perform the *initial branching*: we choose between the root being a black node, a white node or a grey node (the tree being a member of the set of class-n sub-quadtrees).
- At any level i, $n \geqslant i > 1$, for any grey node at this level, we perform a *level-i branching*: we choose between 79 different sub-quadtree sets (those of all the three tables of Fig. 3) so that the subtree rooted at this node belongs to the chosen set.
- At level 1, for any grey node at this level, we perform a *level-1 branching*: we choose between 14 different sub-quadtree sets (those of the leftmost table of Fig. 3) so that the subtree rooted at this node belongs to the chosen set.

Note that except for the initial branching, where we choose between different colors, at all other branchings we choose between different sub-quadtree configurations (we choose the cartesian i-set which the subtree rooted at the grey node of each branching belongs to). This approach is different to the branching process described in (Mathieu et al., 1987), where at each node (starting at the root), we always choose between white, black and grey colors; if we color a node grey, we continue the process recursively for each one of its children choosing always between all these three colors.

We can create a probabilistic model for this branching process by assigning probabilities to the different choices for every branching. This process must be legal under the fundamental probability axioms. Thus, for every specific branching the probabilities of all the different choices must sum to 1. We would like to find a simple assignment of probabilities to the different choices of our branching model that would lead to a formal definition of Samet and Shaffer's model. In the rest of this paragraph we present an arbitrary parametric assignment of probabilities. Remember that at the initial branching we have three choices: the root to be black, white or grey. We call x the probability each of the first two choices. Then, the probability of the third choice will be $1-2x$. Remember also that at all levels above level 1 we have 79 choices. We would like all of them except for the choice of the cartesian i-set represented by $\langle S_{i-1}S_{i-1}S_{i-1}S_{i-1}\rangle$ to be equiprobable (for the sake of simplicity). We call y_i the probability of each of these 78 choices. Since the sum of the probabilities of all the 79 choices must be 1, the probability of choosing the cartesian i-set represented by $\langle S_{i-1}S_{i-1}S_{i-1}S_{i-1}\rangle$ will be $1-78y_i$. Accordingly, at level 1 we would like all 14 choices to be equiprobable.

We can express formally this assignment of probabilities by defining some probability measures. The measure $\pi : Q_n \to [0,1]$ which expresses the probability of a class-n quadtree equals:

$$\pi(t) = \begin{cases} x & \text{if } t \in ■, \\ x & \text{if } t \in □, \\ (1-2x)v_n(t) & \text{if } t \in S_n. \end{cases}$$

The measure $v_i : S_i \to [0,1]$ ($n \geqslant i > 1$) which expresses the probability of a class-i sub-quadtree equals:

$$
v_i(t) = \begin{cases} y_i & \text{if } t \in \langle \blacksquare\square\square\square \rangle + \langle \blacksquare\blacksquare\square\square \rangle + \langle \blacksquare\blacksquare\blacksquare\square \rangle, \\ y_i v_{i-1}(t_1) & \text{if } t \in \langle t_1 \square\square\square \rangle + \langle t_1 \blacksquare\square\square \rangle + \langle t_1 \blacksquare\blacksquare\square \rangle + \langle t_1 \blacksquare\blacksquare\blacksquare \rangle,\ t_1 \in S_{i-1}, \\ y_i v_{i-1}(t_1) v_{i-1}(t_2) & \text{if } t \in \langle t_1 t_2 \square\square \rangle + \langle t_1 t_2 \blacksquare\square \rangle + \langle t_1 t_2 \blacksquare\blacksquare \rangle,\ t_1, t_2 \in S_{i-1}, \\ y_i v_{i-1}(t_1) v_{i-1}(t_2) v_{i-1}(t_3) & \text{if } t \in \langle t_1 t_2 t_3 \square \rangle + \langle t_1 t_2 t_3 \blacksquare \rangle,\ t_1, t_2, t_3 \in S_{i-1}, \\ y_i v_{i-1}(t_1) v_{i-1}(t_2) v_{i-1}(t_3) v_{i-1}(t_4) & \text{if } t \in \langle t_1 t_2 t_3 t_4 \rangle,\ t_1, t_2, t_3, t_4 \in S_{i-1}. \end{cases}
$$

Finally, the measure $v_1 : S_1 \to [0, 1]$ which expresses the probability of a class-1 sub-quadtree equals:

$$
v_1(t) = \frac{1}{14}, \quad t \in S_1.
$$

Let us call this branching process along with this parametric assignment of branching probabilities the "*new random quadtree model*". If we can find values for x and y_i in the interval $[0, 1]$ such that the probability of a black or white node existing anywhere in the tree is constant, then we shall have proved that the new random quadtree model is equivalent to Samet and Shaffer's descriptive model.

Note that

- The set of all class-n quadtrees is a probability space. This means that $P(Q_n) = 1$ and that each class-n quadtree is a distinct outcome in this space.
- The set of all class-i sub-quadtrees ($n \geqslant i \geqslant 1$) is a probability space. This means that $P(S_i) = 1$ and that each class-i sub-quadtree is a distinct outcome in this space.
- The probability of any subset of quadtrees (sub-quadtrees) equals the sum of their probabilities.
- If a grey node has two, three or four children which are sub-quadtrees (not leaves), then these sub-quadtrees are independent to each other (the branching process for any of these sub-quadtrees is independent to the branching process for any of its sibling sub-quadtrees)

We can now prove the following proposition:

Proposition 1. *If $x = 1/(2n-2)$ and $y_i = 1/52i$, the new random quadtree model is equivalent to Samet and Shaffer's random tree model.*

Proof. The case of black nodes is examined. Due to symmetry the same arguments hold for white nodes, too. We denote by A_n^z the set of all class-n quadtrees that include a certain black node z (that is, a black node in a certain position at level). Let j denote the level of z. When $n > j$, we denote by B_i^z the set of all class-i sub-quadtrees ($i > j$) that include z. Note that A_n^z is an event of the probability space Q_n and B_i^z is an event of the probability space S_i ($n \geqslant i > j$). The probability of A_n^z must be constant, equal to a value c ($1 \geqslant c > 0$) for any level and position of z. We shall use the notation

$$
\langle \overbrace{T_{i,1}}, T_{i,2}, T_{i,3}, T_{i,4} \rangle
$$

to denote the union of all the cartesian i-sets that are defined as a cartesian product of $\bigcirc$ and each of the distinct permutations of $T_{i,1}$, $T_{i,2}$, $T_{i,3}$ and $T_{i,4}$ where $T_{i,1}$ has a certain constant position in these permutations.

In Fig. 4 you can see an abstract example of the branchings we perform at each level when we construct quadtrees that have a certain node black. The plain dotted rectangles stand either for a black or a white node; the dotted rectangles that include a circle stand either for a black, a white, or a grey node. Note that at the last branching a configuration with four black nodes can not exist.

We distinguish the following two cases:

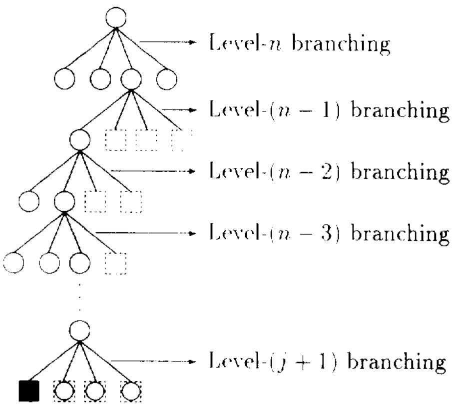

Fig. 4. An abstract example of the branchings we perform when constructing quadtrees with a certain black node.

1. $j = n$. We have $P(A_n^z) = x$; we must then have

$$x = c. \tag{1}$$

2. $j < n$. We have $A_n^z = B_n^z$. Since, $\forall t \in A_n^z$, $\pi(t) = (1-2x)v_n(t)$ we have $P(A_n^z) = (1-2x)P(B_n^z)$. For all levels $i > j+1$ the set B_i^z may be partitioned into the following four distinct subsets:

   - $C_i^z = \langle \overbrace{B_{i-1}^z \square\square\square} \rangle + \langle \overbrace{B_{i-1}^z \blacksquare\square\square} \rangle + \langle \overbrace{B_{i-1}^z \blacksquare\blacksquare\square} \rangle + \langle \overbrace{B_{i-1}^z \blacksquare\blacksquare\blacksquare} \rangle$
   - $D_i^z = \langle \overbrace{B_{i-1}^z S_{i-1} \square\square} \rangle + \langle \overbrace{B_{i-1}^z S_{i-1} \blacksquare\square} \rangle + \langle \overbrace{B_{i-1}^z S_{i-1} \blacksquare\blacksquare} \rangle$
   - $E_i^z = \langle \overbrace{B_{i-1}^z S_{i-1} S_{i-1} \square} \rangle + \langle \overbrace{B_{i-1}^z S_{i-1} S_{i-1} \blacksquare} \rangle$
   - $F_i^z = \langle \overbrace{B_{i-1}^z S_{i-1} S_{i-1} S_{i-1}} \rangle$

   Obviously, $P(B_i^z) = P(C_i^z) + P(D_i^z) + P(E_i^z) + P(F_i^z)$. Note that in C_i^z, D_i^z, E_i^z and F_i^z the root of the child-subtree that belongs in B_{i-1}^z is part of the path that leads to the black node in question. We have that

   - C_i^z is the union of 8 cartesian i-sets each one of which is a cartesian product involving B_{i-1}^z (e.g. level-$(n-1)$ branching in Fig. 4: each of the 3 leaves that are not on the path can be black or white). Since, $\forall t \in C_i^z$, we have $v_i(t) = y_i v_{i-1}(t1)$, where $t1 \in B_{i-1}^z$, summing up probabilities for every element of C_i^z we get $P(C_i^z) = 8y_i P(B_{i-1}^z)$.
   - D_i^z is the union of 12 different cartesian i-sets each one of which is a cartesian product involving B_{i-1}^z and S_{i-1} (e.g. level-$(n-2)$ branching in Fig. 4: the grey node not on the path can take 3 positions, while each of the 2 leaves that are not on the path can be black or white). Since, $\forall t \in D_i^z$, we have $v_i(t) = y_i v_{i-1}(t1) v_{i-1}(t2)$, where $t1 \in B_{i-1}^z \wedge t2 \in S_{i-1}$, summing up probabilities for every element of D_i^z we get $P(D_i^z) = 12y_i P(B_{i-1}^z) P(S_{i-1})$.
   - E_i^z is the union of 6 different cartesian i-sets each one of which is a cartesian product involving B_{i-1}^z and twice the set S_{i-1} (e.g. level-$(n-3)$ branching in Fig. 4: the leaf that is not on the path can take 3 positions, while at each position it can be black or white). Since, $\forall t \in E_i^z$, we have $v_i(t) = y_i v_{i-1}(t1) v_{i-1}(t2) v_{i-1}(t3)$, where $t1 \in B_{i-1}^z \wedge t2, t3 \in S_{i-1}$, summing up probabilities for every element of E_i^z we get $P(E_i^z) = 6y_i P(B_{i-1}^z) P(S_{i-1})^2$.
   - F_i^z is a cartesian i-set which is a cartesian product involving B_{i-1}^z and three times the set S_{i-1} (e.g. level-n branching in Fig. 4: the 3 nodes that are not on the path are all grey). Since, $\forall t \in F_i^z$, we have $v_i(t) = y_i v_{i-1}(t1) v_{i-1}(t2) v_{i-1}(t3) v_{i-1}(t4)$, where $t1 \in B_{i-1}^z \wedge t2, t3, t4 \in S_{i-1}$, summing up probabilities for every element of F_i^z we get $P(F_i^z) = (1 - 78y_i) P(B_{i-1}^z) P(S_{i-1})^3$.

Since $P(S_{i-1}) = 1$ we get $P(B_i^-) = (1 - 52y_i)P(B_{i-1}^-)$.
For level $i = j + 1$ we distinguish two further sub-cases.

(a) $j > 0$. As we did above for B_i^- we can partition B_{j+1}^- into four distinct subsets that, apart from the black node in question, have none, one, two or three children which belong in S_j. Each of these subsets is the union of a number cartesian i-sets. In total, at level-$(j+1)$ branching there are 26 such sets, where one specific child is black (e.g. level-$(j+1)$ branching in Fig. 4: each of the 3 nodes that are not on the path can be grey, white, or black excluding the illegal configuration where all of them are black). Following analogous arguments as above and making use of the fact that $P(S_j) = 1$ we get $P(B_{j+1}^-) = 26y_{j+1}$.
We want $P(A_n^-) = c$. Substituting the values of B_i^- $(n \geqslant i > j)$ we have

$$(1-2x)(1-52y_n)(1-52y_{n-1})\cdots(1-52y_{j+2})26y_{j+1} = c, \quad n-1 \geqslant j \geqslant 1. \tag{2}$$

(b) $j = 0$. B_1^- contains 7 sub-quadtrees (in any of these sub-quadtrees there is one specific child of the root which is black.) Since, $\forall t \in B_1^-$, we have $v_1(t) = 1/14$, we get $P(B_1^-) = 1/2$.
We want $P(A_n^-) = c$. Substituting the values of B_i^- $(n \geqslant i > j)$ we have

$$(1-2x)(1-52y_n)(1-52y_{n-1})\cdots(1-52y_2)1/2 = c. \tag{3}$$

Eqs. (1), (2) and (3) form a system of equations that expresses the conditions under which the *new random quadtree model* is equivalent to Samet and Shaffer's model. This system may be solved as follows: we substitute the value of c given by Eq. (1) into Eqs. (2) and (3). By induction on y_i and Eq. (2) we can show that

$$y_i = \frac{1}{26}\frac{x}{1-2(n-i-1)x}, \quad n \geqslant i > 1. \tag{4}$$

For all $i > 1$, substituting the value of y_i from the above equation into Eq. (3) we find that $x = 1/(2n+2)$. Substituting this value into Eq. (4) we find that $y_i = 1/52i$. □

4. Average efficiency of random quadtrees

We can now prove the following proposition.

Proposition 2. *Consider quadtrees obeying the new random quadtree model with* $x = 1/(2n+2)$ *and* $y_i = 1/52i$. *The average number of nodes,* $\overline{N_n}$, *of a class-n quadtree obeys the equation*

$$\overline{N_n} = \frac{4^{n+2} - 3n - 7}{9(n+1)}. \tag{5}$$

First Proof. This proof is based on methods appearing in (Vitter and Flajolet, 1990). These methods are general since they can be applied to a number of combinatorial structures that are defined as sets through symbolic equations. Let $|t|$ represent the number of nodes of t, where $t \in Q_n + \sum_{i=1}^{n} S_i$. We will define two generating functions. Let

$$\pi(z) = \sum_{t \in Q_n} \pi(t) z^{|t|}$$

and

$$v_i(z) = \sum_{t \in S_i} v_i(t) z^{|t|}, \quad n \geqslant i \geqslant 1.$$

The coefficient of z^l in $\pi(z)$ (in $v_i(z)$) is the probability of the set of quadtrees (sub-quatrees) that have l nodes. Note that $\pi'(1) = \overline{N_n}$. Note also that $\pi(1) = 1$ and $v_i(1) = 1$, $n \geqslant i \geqslant 1$. Using Fig. 3, in which we can see the number of different cartesian i-sets whose union is represented by each of the $\langle\rangle$ notations, and the definitions of our probability measures we reach the following equations

$$\pi(z) = 2\frac{1}{2n+2}z + \frac{n}{n+1}v_n(z),$$
$$v_i(z) = \frac{14}{52i}z^5 + \frac{32}{52i}z^4 v_{i-1}(z) + \frac{24}{52i}z^3 v_{i-1}(z)^2 + \frac{8}{52i}z^2 v_{i-1}(z)^3 + (1 - \frac{78}{52i})z v_{i-1}^4, \quad n \geqslant i > 1,$$
$$v_1(z) = z^5.$$

Differentiating and setting $z = 1$ we get

$$\pi'(1) = \frac{1}{n+1} + \frac{n}{n+1}v_n'(1),$$
$$v_i'(1) = 1 + \frac{4}{i} + 4(\frac{i-1}{i})v_{i-1}'(1), \quad n \geqslant i > 1,$$
$$v_1'(1) = 5.$$

Setting $u_i = i(v_i'(1) - 1), n \geqslant i \geqslant 1$, we get

$$\pi'(1) = 1 + \frac{1}{n+1}u_n,$$
$$u_i = 4i + 4u_{i-1}, \quad n \geqslant i > 1,$$
$$u_1 = 4.$$

Making use of induction on u_n it is easily proved that

$$\pi'(1) = 1 + \frac{4^n}{n+1}\sum_{i=1}^{n}(1/4)^{i-1}i$$

Note that $\sum_{i=1}^{n}(1/4)^{i-1}i = f'(1/4)$, where $f(z) = \sum_{i=1}^{n} z^i = (z^{n+1} - z)/(z-1)$. This remark leads us to

$$\pi'(1) = 1 + \frac{4^{n+2} - 12n - 16}{9(n+1)} = \frac{4^{n+2} - 3n - 7}{9(n+1)}. \qquad \square$$

Second Proof. Consider a tree $q \in Q_n$. Let $L_n(q)$ denote the number of leaf nodes and $N_n(q)$ the number of both leaf and internal nodes in q. At level i of such a tree ($n \geqslant i \geqslant 0$) there may exist at most 4^{n-i} nodes. For all levels, we associate a random variable to each of these potential nodes. This variable equals 1 when the related node exists and is a leaf. Otherwise, it equals 0. The sum of all these variables for every level gives us $L_n(q)$. Thus, the expected number of leaf nodes, $\overline{L_n}$, in a class-n quadtree equals the sum of the average values of all the random variables associated with potential nodes. The definition of our random tree model implies that the average value of such a variable equals $1/(n+1)$. This means that $\overline{L_n} = \sum_{i=0}^{n} 1/(n+1) \cdot 4^{n-i} = (4^{n+1} - 1)/(3n+3)$. By induction on the number of internal nodes, we can prove that in any class-n quadtree q we have $N_n(q) = (4L_n(q) - 1)/3$. Taking expectations we have

$$\overline{N_n} = \frac{4\overline{L_n} - 1}{3} = \frac{4^{n+2} - 3n - 7}{9(n+1)}. \qquad \square$$

Using the latter proof method it is very easy to develop and prove formulae for the average numbers of white, black and grey nodes at a specific level of a class-n quadtree.

5. Conclusion

In this paper, we present a new model of random region quadtrees. We prove that for certain values of parameters this model is the formal equivalent of Samet and Shaffer's random tree model. We also prove in two ways a formula that expresses the average number of nodes in a quadtree. The method followed in the second proof can be easily used for producing further analytic results. Note that the presented model of random trees can capture coherence of unicolor regions that appears in images of real applications in a very good way, although, its definition as a branching process is rather complicated.

Future research could provide analogous models for other space dimensions, e.g. in R^1 a model for segment trees and in R^3 for oct-trees (Samet, 1990a; Samet, 1990b). In addition, the connection between the presented random image model and fractals remains to be investigated along the lines of reference (Faloutsos and Kamel, 1994), which examines the performance of spatial access methods using this technique. Moreover, future research could examine more complicated probability assignments for the presented random tree model.

Acknowledgment

The first author, who is a postgraduate scholar of the State Scholarship-Foundation of Greece, wishes to thank this foundation for its financial assistance to this research.

References

Burton, F.W., V.J. Kollias and J.G. Kollias (1985). Expected and worst-case storage requirements for quadtrees. *Pattern Recognition Lett.* 3 (2) 131–135.

Cross, G.R. and A.K. Jain (1983). Markov random field texture models. *IEEE Trans. Pattern Anal. Mach. Intell.* 5 (1) 25–39.

Faloutsos, C. (1992). Analytical results on the quadtree decomposition of arbitrary rectangles. *Pattern Recognition Lett.* 13 (1) 31–40.

Faloutsos, C. and I. Kamel (1994). Beyond uniformity and Independence: Analysis of R-trees using the concept of fractal dimension. *Proc. ACM Principles of Database Systems Conf.* (PODS), Minneapolis, 4–13.

Flajolet, P., G. Gonnet, C. Puech and J.M. Robson (1993). Analytic variations on quadtrees. *Algorithmica* 10 (7) 473–500.

Mathieu, C., C. Puech and H. Yahia (1987). Average efficiency of data structures for binary image processing. *Inform. Process. Lett.* 26 (2) 89–93.

Samet, H. (1982). Neighbor finding techniques for images represented by quadtrees. *Computer Graphics and Image Processing* 18 35–57.

Samet, H. (1990a). *The Design and Analysis of Spatial Data Structures.* Addison-Wesley, Reading, MA.

Samet, H. (1990b). *Applications of Spatial Data Structures.* Addison-Wesley, Reading, MA.

Samet, H. and C. A. Shaffer (1985). A model for the analysis of neighbor finding in pointer-based quadtrees. *IEEE Trans. Pattern Anal. Mach. Intell.* 7 (6) 717–720.

Shaffer, C.A. (1988). A formula for computing the number of quadtree nodes created by a shift. *Pattern Recognition Lett.* 7 (1) 45–49.

Vassilakopoulos, M. and Y. Manolopoulos (1993). Analytical results on the quadtree storage-requirements. *Proc. 5th Internat. Conf. on Computer Analysis of Images and Patterns* (CAIP), Budapest, 41–48.

Vitter, J.S. and Ph. Flajolet (1990). Average-case analysis of algorithms and data structures. In: J. van Leeuwen, Ed., *Handbook of Theoretical Computer Science. Vol. A. Algorithms and Complexity.* MIT Press/Elsevier, Cambridge, MA/Amsterdam, Ch. 9, 431–524.

Unnikrishnan, A., Y.V. Venkatesh and P. Shankar (1987). Distribution of black nodes at various levels in a linear quadtree. *Pattern Recognition Lett.* 6 (5) 341–342