Appeared in: Nordic Journal of Computing 2(1995), 70-84.

Efficiency Analysis of Overlapped Quadtrees

Michael VASSILAKOPOULOS,
Department of Electrical & Computer Eng.,
Aristotelian University of Thessaloniki,
54006 Thessaloniki, Greece.
e-mail: vasilako@eng.auth.gr

Yannis MANOLOPOULOS*
Department of Informatics,
Aristotelian University of Thessaloniki,
54006 Thessaloniki, Greece.
e-mail: manolopo@eng.auth.gr

and Brigitte KRÖLL
Institut für Theoretische Informatik,
ETH Zentrum,
8092 Zürich, Switzerland.
e-mail: kroell@inf.ethz.ch

CR categories and subject descriptors: E.1, E.2, F.2.2, G.2.1, G.2.2, H.2.2
General Terms: Algorithms, Performance
Keywords: Quadtrees, Random Images, Overlapping Data Stuctures

Abstract

The technique of overlapping region quadtrees results in significant space reduction. This new form of data structure is suitable for representing sequences of similar binary raster images emerging in many areas. Apart from a short introduction to the method of overlapping, an analysis of the space reduction achieved is presented in this article. First, a general theorem about the average number of common nodes of two images obeying a general model is given. Next, two models of image similarity are introduced and the respective formulae are derived from the general one. Evaluating these formulae for various parameters shows the efficiency of overlapping.

*Research performed while at the Department of Computer Science, University of Maryland, College Park, MD 20742, USA.

1 Introduction

The region quadtree is the most popular member in the family of quadtree data structures. It is used for the representation of binary images, that is $2^n \times 2^n$ binary arrays (for a positive integer n), where a 1 (0) entry stands for a black (white) picture element. In particular, it is a degree four tree with height n, at most. Each node corresponds to a square array of pixels (the root corresponds to the whole image). If all of them have the same color (black or white) the node is a leaf of that color. Otherwise, the node is colored gray and has four children. Each of these children corresponds to one of the four square subarrays to which the array of that node is partitioned. We assume here, that the first (leftmost) child corresponds to the left, upper subarray, the second to the right, upper subarray, the third to the left, lower subarray and the fourth (rightmost) child to the right, lower subarray. For more details regarding quadtrees see [8] and [11]. Figure 1.c shows a quadtree for the 8×8 pixel array of Figure 1.a.

Now, we want to save memory space by overlapping quadtrees, that is to use the same memory area for storing common subtrees of different trees. Consider two quadtrees that represent an image and a change of that image. If the two images have common subimages covering the same coordinates, the corresponding subtrees of the two trees that represent these images are equal and may be kept only once in memory, while they belong to paths of both trees. In case there are not any common subimages between the two images we have two isolated trees. In case the two instances are two identical images so are the trees (the pointers to the roots of the trees remain distinct although they point to the same node). If there are not only two images, but a long sequence of images representing for example a gradually changing image, common substructures may be shared by more than two consecutive trees. This sharing is transparent to the algorithms accessing the trees. In other words, there is not any access time overhead because of overlapping. Figures 1.a and 1.b demonstrate two similar images and Figures 1.c and 1.d demonstrate the respective quadtrees. The result of applying overlapping to these trees is shown in Figure 1.e. Note, that when there is a different substructure (which might even be a different pixel for a certain image position) the total path to this substructure is stored. This guarantees that the access time of any node remains the same.

The similarity between consecutive image data in many applications is extensive.

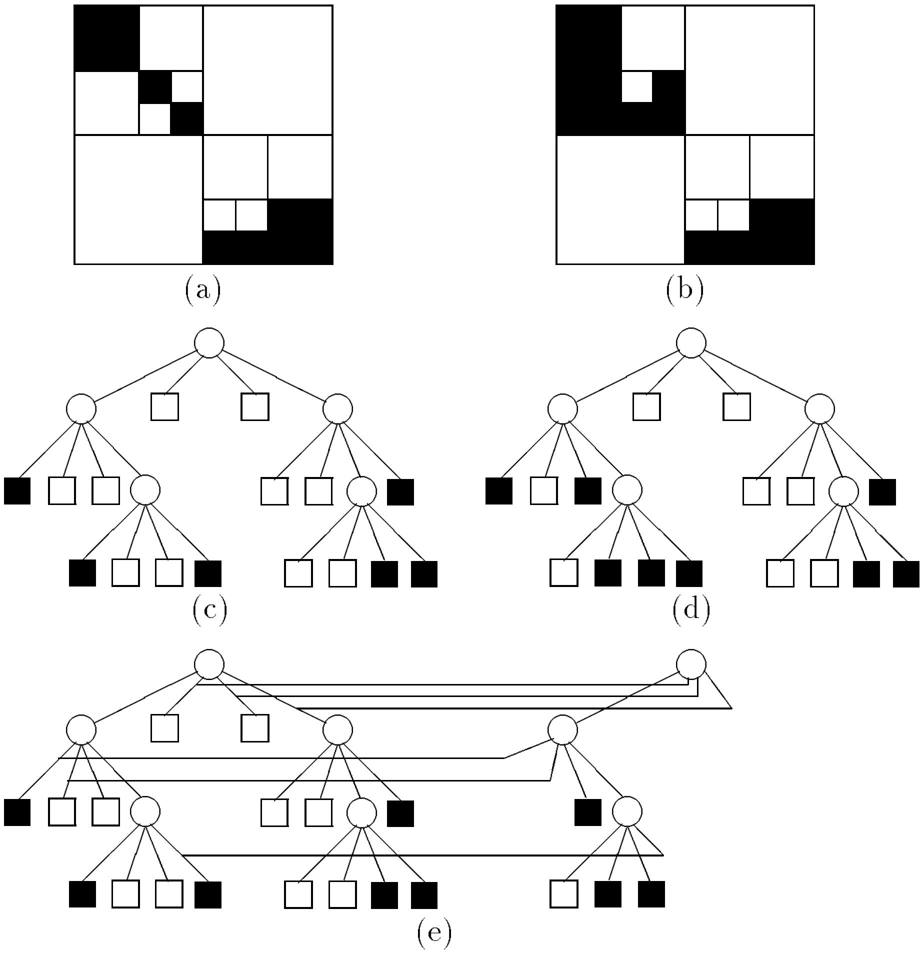

Figure 1: (a),(b): two similar 8x8 images, (c),(d): the corresponding quadtrees and (e): the overlapped structure.

These include Animation, Image Processing, Geographic Information Systems or even Meteorology Weather Pictures. While the amount of space needed to store each of these images is vast, overlapping quadtrees stand for an excellent storing structure that does not influence the access ability.

Overlapping quadtrees constitute a main memory structure. However, there are also structures which are suitable for storing sets of similar images in secondary memory, like the Multivalued Quadtree [1]. The use of common substructures in quadtrees has been originally proposed in [4] as one of the advantages that functional languages (and especially Miranda) provide over procedural languages for graphics programmers, while an algorithm that performs map overlaying and is similar to the one used in overlapping is described in [3]. More details about overlapping quadtrees, along with the necessary algorithms and early simulation results, can be found in [13]. This article focuses on analyzing the space reduction achieved. The analysis presented extends the few previous theoretical studies on the space efficiency of the quadtree. More specifically, there have been given formulae for the average and/or worst-case number of nodes of a quadtree representing rectangles [6, 7, 12], polygons [2, 9] and random images [14]. This work is about the average number of common nodes when quadtrees represent similar images.

2 Analysis of overlapping quadtrees

In the following sections we will make use of probability theory, thus we will give some fundamental properties here.

Let Ω be a finite sample space and let $\mathcal{P}(\Omega)$ denote the set of all subsets of Ω. A map $P : \mathcal{P}(\Omega) \rightarrow [0, 1]$ with

$$\begin{aligned} P(\Omega) &= 1 \qquad \text{and} \\ P(A \cup B) &= P(A) + P(B) \qquad \text{for disjoint sets } A, B \subset \Omega \end{aligned}$$

is called a *probability measure* on Ω.

For any event $A \subset \Omega$ we denote the complement $\Omega \setminus A$ of A by $\overline{A}$. Two events $A, B \subset \Omega$ are called *stochastically independent*, if $P(A \cap B) = P(A) \cdot P(B)$ holds. More than two events $A_1, \ldots, A_m \subset \Omega$ ($m \in \mathbf{N}$, where $\mathbf{N}$ denotes the set of all positive integers) are said to be *stochastically independent*, if $P(A_{i_1} \cap \ldots \cap A_{i_r}) = P(A_{i_1}) \cdot \ldots \cdot P(A_{i_r})$ for all subsets $\{i_1, \ldots, i_r\} \subset \{1, \ldots, m\} (1 \leq r \leq m)$.

For explanations about *random variable*, *expected value* and *independence of random variables* see [5]. For a random variable X defined on the probability space (Ω, P) we denote the expected value of X with $E_P[X]$.

2.1 Probabilistic assumptions

In the following the positive integer n that determines the size of the quadtrees under consideration is fixed. A quadtree corresponding to a $2^n \times 2^n$ image is called a *class-n quadtree*. Let Ω denote the set of all class-n quadtrees. One class-n quadtree represents an image of $2^n * 2^n = 4^n$ pixels. That implies that Ω is a finite set with $|\Omega| = 2^{4^n}$. In our model we assume that a single pixel is black (resp. white) with probability p (resp. $1-p$), where $0 < p < 1$, and that the events for all pixels to be black are stochastically independent. Let $P_p : \Omega \to [0,1]$ denote the probability measure which reflects this idea, that is $P_p(\omega) = p^{b(\omega)} * (1-p)^{w(\omega)}$, where $b(\omega)$ (resp. $w(\omega)$) is the number of black (resp. white) pixels in the $2^n \times 2^n$ image corresponding to ω. Note, that this is clearly a random-image model in contrast to the random-tree model introduced by Samet, where probabilities are assigned directly to the nodes of a quadtree.

The level of the root of a class-n quadtree is denoted as level n, the level of the sons of the root as level $n-1$ and so forth, that is, leaves representing single pixels are at level 0. A node at level i ($0 \leq i \leq n$) represents a $2^i \times 2^i$ subimage of the corresponding image (there are at most 4^{n-i} such nodes). That implies, that a node at level i, resp. the $2^i \times 2^i$ subimage, is black with probability $b_i(p) := p^{4^i}$, white with probability $w_i(p) := (1-p)^{4^i}$ or gray with probability $g_i(p) := 1 - b_i(p) - w_i(p)$. Obviously, $b_{i+1}(p) = b_i(p)^4$ and $w_{i+1}(p) = w_i(p)^4$.

We use this random pixel model to get a first impression of how well overlapping works, that is how much storage space we save in this case. The same model of randomness which assumes independency of coloring for each pixel has also been used in [10]. Further this model is connected with real life to some extend for a particular area of values of p. When p equals 0.5 the respective images mainly consist of small black and white areas. In this case a quadtree is a representation of limited efficiency, since it is almost a full structure (leaf nodes appear almost exclusively at the lowest level). Such images are rare in real applications. However, when p is quite different to 0.5 the probability of having large unicolor blocks is significant. In other words, when p is not close to 0.5 our model expresses coherence to a significant extend and can represent images found in practice (for example medical, or meteorological images).

2.2 Average number of nodes for a single image

In this section we prove a formula for the average number of nodes of a quadtree $\omega \in \Omega$ obeying P_p. We need this as denominator for computing the average space reduction induced by overlapping quadtrees. Here, we give a short proof for that formula presented also in [14]. First, let us introduce some terminology to simplify the following argumentation. Let ω^* denote the *completion of* ω, that is, the complete class-n quadtree with 4^n leaves at level 0, which we obtain from ω by substituting each black (resp. white) leaf at level i in ω by a complete degree four tree of height i with 4^i black (resp. white) leaves at level 0, and doing this for all $i \in \{1, \ldots, n\}$. The jth node (counted from the left side) at level i of ω^* is denoted by $v_{ij}(\omega^*)$ for $i \in \{0, \ldots, n\}$ and $j \in \{1, \ldots, 4^{n-i}\}$. Further let $v_{ij}(\omega)$ denote that node (if it exists) in ω that corresponds to $v_{ij}(\omega^*)$, and let $x_{ij}(\omega)$ be that $2^i \times 2^i$ subimage of the image represented by ω that corresponds to $v_{ij}(\omega^*)$. Note, that $x_{ij}(\omega)$ is defined for all $i \in \{0, \ldots, n\}$ and $j \in \{1, \ldots, 4^{n-i}\}$, but $v_{ij}(\omega)$ is only defined, if the corresponding node exists.

Now we are able to prove

Theorem 2.1 *The expected number $E_p[N]$ of nodes of a class-n quadtree obeying the probability measure P_p is given by*

$$E_p[N] = 1 + \sum_{i=0}^{n-1} 4^{n-i} g_{i+1}(p).$$

Proof: For $i \in \{0, \ldots, n\}$ and $j \in \{1, \ldots, 4^{n-i}\}$ let $X_{ij} : \Omega \to \{0, 1\}$ denote the random variable with $X_{ij}(\omega) = 1$ if $v_{ij}(\omega)$ exists and $X_{ij}(\omega) = 0$ if $v_{ij}(\omega)$ doesn't exist. Obviously $X_{n1}(\omega) = 1$ holds for all $\omega \in \Omega$. For $i < n$ the equality $X_{ij}(\omega) = 1$ holds if and only if the father of $v_{ij}(\omega^*)$ is gray. The number of nodes of ω is given by $N(\omega) = \sum_{i=0}^{n} \sum_{j=1}^{4^{n-i}} X_{ij}(\omega)$. This leads to the following formula for the expected value $E_p[N]$ of $N : \Omega \to \mathbf{N}$.

$$\begin{aligned} E_p[N] &= E_p[\sum_{i=0}^{n} \sum_{j=1}^{4^{n-i}} X_{ij}] \\ &= \sum_{i=0}^{n} \sum_{j=1}^{4^{n-i}} E_p[X_{ij}] \end{aligned}$$

$$
\begin{aligned}
&= E_p[X_{n1}] + \sum_{i=0}^{n-1} \sum_{j=1}^{4^{n-i}} E_p[X_{ij}] \\
&= 1 + \sum_{i=0}^{n-1} \sum_{j=1}^{4^{n-i}} g_{i+1}(p) \\
&= 1 + \sum_{i=0}^{n-1} 4^{n-i} g_{i+1}(p).
\end{aligned}
$$

($E_p[X_{ij}]$ equals the probability that the father of $v_{ij}(\omega^*)$ is gray, which equals $g_{i+1}(p)$.)
$\square$

2.3 The general theorem

In order to estimate the space reduction induced by overlapping we will define a probability space that reflects the similarity of two quadtrees, each of them obeying a probability measure P_p, but possibly with different values for the parameter p. Formally that is a probability measure R on the product space $\Omega \times \Omega$, that is, two quadtrees ω_1 and ω_2 for which the space reduction should be measured occur with probability $R(\omega_1, \omega_2)$. To formalize the fact that each of the two quadtrees obeys a probability measure P_p, we define the projections $\pi_k : \Omega \times \Omega \to \Omega, (\omega_1, \omega_2) \mapsto \omega_k$ $(k = 1, 2)$ and require that the distribution R_{π_k} of π_k follows $R_{\pi_k} = P_{p_k}$ for $k = 1, 2$ with $0 < p_1, p_2 < 1$. Note, that in this generality $R(\omega_1, \omega_2)$ gives no information about the dependence of π_1 and π_2. In the following section we will specify R in two different ways, one of them is to require the independence of π_1 and π_2.

Up to the end of this section we will work on the sample space $\Omega \times \Omega$ with a probability measure R as described above. Further, we will assume an additional property of R which will be fulfilled by both measures discussed in the following. For $i \in \{0, \ldots, n\}$ and $j \in \{1, \ldots, 4^{n-i}\}$ let r_{ij} denote the probability that for a pair $(\omega_1, \omega_2) \in \Omega \times \Omega$ of quadtrees chosen according to R the image $x_{ij}(\omega_1)$ equals $x_{ij}(\omega_2)$ and both $v_{ij}(\omega_1)$ and $v_{ij}(\omega_2)$ exist. In this case we say that $v_{ij}(\omega_1)$, resp. $v_{ij}(\omega_2)$, is a common node of ω_1 and ω_2. Formally $r_{ij} = R(\{(\omega_1, \omega_2) \mid (x_{ij}(\omega_1) = x_{ij}(\omega_2)) \wedge (v_{ij}(\omega_1) \text{ exists }) \wedge (v_{ij}(\omega_2) \text{ exists })\})$. The additional property of R mentioned above is, that the probability that the image $x_{ij}(\omega_1)$ equals $x_{ij}(\omega_2)$ and both $v_{ij}(\omega_1)$ and $v_{ij}(\omega_2)$ exist is the same for all nodes at the same level. I. e., the equalities $r_{ij_1} = r_{ij_2}$ hold for all $i \in \{0, \ldots, n\}$ and $j_1, j_2 \in \{1, \ldots, 4^{n-i}\}$.

In this case we define $r_i := r_{i1}$ for $i \in \{0, \dots, n\}$.

Theorem 2.2 *The expected number $E_R[K]$ of common nodes of two class-n quadtrees obeying the probability measure R as specified above follows the equation*

$$E_R[K] = r_n + \sum_{i=1}^{n} 4^i r_{n-i}.$$

Proof: Let $K : \Omega \times \Omega \to \mathbf{N_0}$ ($\mathbf{N_0}$ denotes the set of the non-negative integers) define the random variable which maps the pair (ω_1, ω_2) to the number of common nodes of ω_1 and ω_2. Further, for $i \in \{0, \dots, n\}$ and $j \in \{1, \dots, 4^{n-i}\}$ let $K_{ij} : \Omega \times \Omega \to \{0, 1\}$ define the random variable with

a. $K_{ij}(\omega_1, \omega_2) = 1$, if $x_{ij}(\omega_1) = x_{ij}(\omega_2)$ and both $v_{ij}(\omega_1)$ and $v_{ij}(\omega_2)$ exist, or with

b. $K_{ij}(\omega_1, \omega_2) = 0$, otherwise.

This implies the relation

$$K = \sum_{i=0}^{n} \sum_{j=1}^{4^{n-i}} K_{ij},$$

and it follows

$$\begin{aligned} E_R[K] &= E_R[\sum_{i=0}^{n} \sum_{j=1}^{4^{n-i}} K_{ij}] \\ &= \sum_{i=0}^{n} \sum_{j=1}^{4^{n-i}} E_R[K_{ij}] \\ &= E_R[K_{n1}] + \sum_{i=0}^{n-1} \sum_{j=1}^{4^{n-i}} E_R[K_{ij}] \\ &= r_n + \sum_{i=0}^{n-1} 4^{n-i} r_i. \end{aligned}$$

$\square$

2.4 The first model

Now, we want to specify the probability measure $R : \Omega \times \Omega \to [0,1]$ in two different ways. First, we want to investigate how well overlapping works, when our images have no intended relation to each other, that is, we assume the stochastic independence of the projections π_1 and π_2 (this is the mathematical description of the idea to choose two quadtrees independently). Further we require $p_1 = p_2 \,(=: p)$. Let R_1 denote this particular measure. It follows that R_1 is the product measure $P_p \otimes P_p$ (see [5], p. 51f). This leads to

Theorem 2.3 *The expected number $E_{R_1}[K]$ of common nodes of two class-n quadtrees obeying the probability measure R_1 follows the equation*

$$\begin{aligned} E_{R_1}[K] &= (p^2+(1-p)^2)^{4^n} + \sum_{i=1}^{n} 4^i \Big(b_{n-i+1}(p)^2 - 2b_{n-i}(p)b_{n-i+1}(p) + w_{n-i+1}(p)^2 \\ &\qquad -2w_{n-i}(p)w_{n-i+1}(p) + (p^2+(1-p)^2)^{4^{n-i}} \Big). \end{aligned}$$

Before giving the proof of the theorem, we will specify some events in $\Omega \times \Omega$ which we need several times in the following. For $i \in \{0, \ldots, n\}, j \in \{1, \ldots, 4^n\}$ and $k = 1, 2$ we define

$$\begin{aligned} A_{ij} &:= \{(\omega_1,\omega_2) | x_{ij}(\omega_1) = x_{ij}(\omega_2)\} \\ B_{ij}^k &:= \{(\omega_1,\omega_2) | x_{ij}(\omega_k) \text{ is black } \} \\ W_{ij}^k &:= \{(\omega_1,\omega_2) | x_{ij}(\omega_k) \text{ is white } \} \end{aligned}$$

Proof of Theorem 2.3: The equality $A_{0j} = (B_{0j}^1 \cap B_{0j}^2) \cup (W_{0j}^1 \cap W_{0j}^2)$ and the independence of π_1 and π_2 lead to

$$\begin{aligned} R_1(A_{0j}) &= R_1(B_{0j}^1)R_1(B_{0j}^2) + R_1(W_{0j}^1)R_1(W_{0j}^2) \\ &= p^2 + (1-p)^2 \end{aligned}$$

for all $j \in \{1, \ldots, 4^n\}$. One can derive that the assumptions made for R_1 imply the independence of the events $A_{01}, \ldots, A_{04^n}$. Thus we obtain $r_n = R_1(A_{n1}) = R_1(\bigcap_{j=1}^{4^n} A_{0j}) = \prod_{j=1}^{4^n} R_1(A_{0j}) = (p^2+(1-p)^2)^{4^n}$. To compute r_{i1} for $i < n$ we divide the event

$$C_i = \{(\omega_1,\omega_2) \mid (x_{i1}(\omega_1) = x_{i1}(\omega_2)) \wedge (v_{i1}(\omega_1) \text{ exists}) \wedge (v_{i1}(\omega_2) \text{ exists})\}$$

into three pairwise disjoint events $C_i = C_{i1} \cup C_{i2} \cup C_{i3}$ with

$$\begin{aligned}
C_{i1} = \{(\omega_1,\omega_2) \mid\ & (x_{i1}(\omega_1) = x_{i1}(\omega_2) \text{ are completely black}) \\
& \wedge(v_{i1}(\omega_1) \text{ exists}) \wedge (v_{i1}(\omega_2) \text{ exists})\} \\
= \{(\omega_1,\omega_2) \mid\ & (x_{i1}(\omega_1) = x_{i1}(\omega_2) \text{ are completely black}) \\
& \wedge(x_{i2}(\omega_1) \cup x_{i3}(\omega_1) \cup x_{i4}(\omega_1) \text{ is not completely black}) \\
& \wedge(x_{i2}(\omega_2) \cup x_{i3}(\omega_2) \cup x_{i4}(\omega_2) \text{ is not completely black})\}, \\
C_{i2} = \{(\omega_1,\omega_2) \mid\ & (x_{i1}(\omega_1) = x_{i1}(\omega_2) \text{ are completely white}) \\
& \wedge(v_{i1}(\omega_1) \text{ exists}) \wedge (v_{i1}(\omega_2) \text{ exists})\} \\
= \{(\omega_1,\omega_2) \mid\ & (x_{i1}(\omega_1) = x_{i1}(\omega_2) \text{ are completely white}) \\
& \wedge(x_{i2}(\omega_1) \cup x_{i3}(\omega_1) \cup x_{i4}(\omega_1) \text{ is not completely white}) \\
& \wedge(x_{i2}(\omega_2) \cup x_{i3}(\omega_2) \cup x_{i4}(\omega_2) \text{ is not completely white})\}, \\
C_{i3} = \{(\omega_1,\omega_2) \mid\ & (x_{i1}(\omega_1) = x_{i1}(\omega_2) \text{ are neither completely black} \\
& \text{nor completely white}) \wedge (v_{i1}(\omega_1) \text{ exists}) \wedge (v_{i1}(\omega_2) \text{ exists})\} \\
= \{(\omega_1,\omega_2) \mid\ & x_{i1}(\omega_1) = x_{i1}(\omega_2) \text{ are neither completely black} \\
& \text{nor completely white}\}.
\end{aligned}$$

The independence requirements lead to

$$\begin{aligned}
R_1(C_{i1}) &= b_i(p)^2(1-b_i(p)^3)^2 \qquad \text{and} \\
R_1(C_{i2}) &= w_i(p)^2(1-w_i(p)^3)^2.
\end{aligned}$$

A similar argument as made above for the computation of r_n gives $R_1(\{(\omega_1,\omega_2)|x_{i1}(\omega_1) = x_{i1}(\omega_2)\}) = (p^2+(1-p)^2)^{4^i}$, thus

$$R_1(C_{i3}) = (p^2+(1-p)^2)^{4^i} - b_i(p)^2 - w_i(p)^2.$$

It follows

$$\begin{aligned}
r_{i1} &= R_1(C_i) \\
&= R_1(C_{i1}) + R_1(C_{i2}) + R_1(C_{i3}) \\
&= b_i(p)^2(1-b_i(p)^3)^2 + w_i(p)^2(1-w_i(p)^3)^2 + (p^2+(1-p)^2)^{4^i} \\
&\quad -b_i(p)^2 - w_i(p)^2 \\
&= b_{i+1}(p)^2 - 2b_i(p)b_{i+1}(p) + w_{i+1}(p)^2 - 2w_i(p)w_{i+1}(p) + (p^2+(1-p)^2)^{4^i}
\end{aligned}$$

with making use of $b_{i+1}(p) = b_i^4(p)$ and $w_{i+1}(p) = w_i^4(p)$. Because of the symmetrical properties of R_1 we obtain $r_{ij_1} = r_{ij_2}$ for all $j_1, j_2 \in \{1, \ldots, 4^{n-i}\}$, that is, we can apply theorem 2.2 and it follows

$$\begin{aligned} E_{R_1}[K] &= (p^2 + (1-p)^2)^{4^n} + \sum_{i=1}^{n} 4^i \Big(b_{n-i+1}(p)^2 - 2b_{n-i}(p)b_{n-i+1}(p) + w_{n-i+1}(p)^2 \\ &\qquad -2w_{n-i}(p)w_{n-i+1}(p) + (p^2 + (1-p)^2)^{4^{n-i}} \Big) . \end{aligned}$$

□

2.5 The second model

Now we want to specify the probability measure R of section 2.3 in a second way, denoting it by R_2. We introduce this model to reflect the idea, that an image corresponding to $\omega_2 \in \Omega$ results from an image corresponding to $\omega_1 \in \Omega$ through changing the color of some pixels. Note, that for $j \in \{1, \ldots, 4^n\}$ the value $R_2(\overline{A_{0j}})$ gives the probability that the pixel $x_{0j}(\omega_1)$ changes its color.

We make two assumptions for R_2. The first is, that this color changing probability $R_2(\overline{A_{0j}})$ is the same for all pixels, formally $R_2(\overline{A_{0j}}) = c$ for all $j \in \{1, \ldots, 4^n\}$, where c is a given number such that $0 < c < 1$. Furthermore we require the stochastic independence of the events $A_{01}, \ldots, A_{04^n}$, $B_{01}^1, \ldots, B_{04^n}^1$, which means that

a. the color of a pixel and the occurrence that the color of this pixel changes are independent for all pixels, and further, that

b. the events of changing the color of two or more pixels are independent.

The relation $B_{0j}^2 = (B_{0j}^1 \cap A_{0j}) \cup (\overline{B_{0j}^1} \cap \overline{A_{0j}})$ implies

$$\begin{aligned} R_2(B_{0j}^2) &= R_2(B_{0j}^1)R_2(A_{0j}) + R_2(\overline{B_{0j}^1})R_2(\overline{A_{0j}}) \\ &= p_1(1-c) + (1-p_1)c \end{aligned}$$

for all $j \in \{1, \ldots, 4^n\}$. Thus, we obtain $p_2 = p_1(1-c) + (1-p_1)c$.

Theorem 2.4 *The expected number $E_{R_2}[K]$ of common nodes of two class-n quadtrees obeying the probability measure R_2 follows the equation*

$$\begin{aligned} E_{R_2}[K] &= c_n(c) + \sum_{i=1}^{n} 4^i c_{n-i}(c) \Big(1 - (b_{n-i+1}(p_1) + w_{n-i+1}(p_1))(1 - c_{n-i}(c)^3) \\ &\qquad -(p_1(1-c) + (1-p_1)c)^{3\cdot 4^{n-i}} b_{n-i}(p_1) \\ &\qquad -(1 - p_1(1-c) - (1-p_1)c)^{3\cdot 4^{n-i}} w_{n-i}(p_1)\Big) \end{aligned}$$

with $c_i(c) = (1-c)^{4^i}$ for $0 \leq i \leq n$.

Proof: The event A_{ij} for $i \in \{0, \ldots, n\}$ and $j \in \{1, \ldots, 4^{n-i}\}$ is the intersection of 4^i events of $A_{01}, \ldots, A_{04^n}$. This and the independence assumption lead to $R_2(A_{ij}) = (1-c)^{4^i} =: c_i(c)$, especially $r_n = R_2(A_{n1}) = c_n(c)$.

Using the same terminology as in the proof of Theorem 2.3 we obtain for $i < n$

$$r_{i1} = R_2(C_i) = R_2(C_{i1}) + R_2(C_{i2}) + R_2(C_{i3}).$$

For computing $R_2(C_{i1})$ we define $(k = 1, 2)$

$$\begin{aligned} D_i^k &= \overline{B_{i2}^k \cap B_{i3}^k \cap B_{i4}^k} \\ &= \{(\omega_1, \omega_2) | x_{i2}(\omega_k) \cup x_{i3}(\omega_k) \cup x_{i4}(\omega_k) \text{ is not completely black}\}. \end{aligned}$$

This leads to

$$C_{i1} = B_{i1}^1 \cap A_{i1} \cap D_i^1 \cap D_i^2.$$

One can show that the independence assumption leads to the independence of B_{i1}^1, A_{i1} and $D_i^1 \cap D_i^2$, thus we get

$$\begin{aligned} R_2(C_{i1}) &= R_2(B_{i1}^1) R_2(A_{i1}) R_2(D_i^1 \cap D_i^2) \\ &\quad b_i(p_1) c_i(c) R_2(D_i^1 \cap D_i^2). \end{aligned}$$

Computing

$$\begin{aligned} R_2(D_i^1 \cap D_i^2) &= R_2(\overline{\overline{D_i^1} \cup \overline{D_i^2}}) \\ &= 1 - R_2(\overline{D_i^1} \cup \overline{D_i^2}) \end{aligned}$$

$$
\begin{aligned}
&= 1 - R_2(\overline{D_i^1}) - R_2(\overline{D_i^2}) + R_2(\overline{D_i^1} \cap \overline{D_i^2}) \\
&= 1 - R_2(B_{i2}^1 \cap B_{i3}^1 \cap B_{i4}^1) - R_2(B_{i2}^2 \cap B_{i3}^2 \cap B_{i4}^2) \\
&\quad + R_2(B_{i2}^1 \cap B_{i3}^1 \cap B_{i4}^1 \cap A_{i2} \cap A_{i3} \cap A_{i4}) \\
&= 1 - b_i(p_1)^3 - b_i(p_2)^3 + b_i(p_1)^3 c_i(c)^3 \\
&= 1 - b_i(p_1)^3(1 - c_i(c)^3) - b_i(p_2)^3.
\end{aligned}
$$

(using the independence of B_{i2}^2, B_{i3}^2 and B_{i4}^2 that we obtain from the independence assumption) gives

$$R_2(C_{i1}) = b_i(p_1)c_i(c)(1 - b_i(p_1)^3(1 - c_i(c)^3) - b_i(p_2)^3).$$

Analogous arguments lead to

$$R_2(C_{i2}) = w_i(p_1)c_i(c)(1 - w_i(p_1)^3(1 - c_i(c)^3) - w_i(p_2)^3).$$

Further we get the independence of A_{i1} and $B_{i1}^1 \cup W_{i1}^1$, that is

$$
\begin{aligned}
R_2(C_{i3}) &= R_2(A_{i1} \cap (\overline{B_{i1}^1 \cup W_{i1}^1})) \\
&= R_2(A_{i1})(1 - R_2(B_{i1}^1) - R_2(W_{i1}^1)) \\
&= c_i(c)(1 - b_i(p_1) - w_i(p_1))
\end{aligned}
$$

It follows

$$
\begin{aligned}
r_{i1} &= R_2(C_{i1}) + R_2(C_{i2}) + R_2(C_{i3}) \\
&= b_i(p_1)c_i(c)\left(1 - b_i(p_1)^3(1 - c_i(c)^3) - b_i(p_2)^3\right) \\
&\quad + w_i(p_1)c_i(c)\left(1 - w_i(p_1)^3(1 - c_i(c)^3) - w_i(p_2)^3\right) \\
&\quad + c_i(c)(1 - b_i(p_1) - w_i(p_1)) \\
&= c_i(c)\Big(1 - (b_{i+1}(p_1) + w_{i+1}(p_1))(1 - c_i(c)^3) \\
&\qquad - b_i(p_1)b_i(p_2)^3 - w_i(p_1)w_i(p_2)^3\Big)
\end{aligned}
$$

with $b_i(p_2) = (p_1(1-c) + (1-p_1)c)^{4^i}$ and $w_i(p_2) = (1 - p_1(1-c) - (1-p_1)c)^{4^i}$.

Arguments about the symmetry of R_2 lead to $r_{ij_1} = r_{ij_2}$ for all $i \in \{0, \ldots, n\}$ and $j_1, j_2 \in \{1, \ldots, 4^i\}$; thus we can apply Theorem 2.2 and obtain

$$
\begin{aligned}
E_{R_2}[K] &= c_n(c) + \sum_{i=1}^{n} 4^i c_{n-i}(c)\Big(1 - (b_{n-i+1}(p_1) + w_{n-i+1}(p_1))(1 - c_{n-i}(c)^3) \\
&\qquad - (p_1(1-c) + (1-p_1)c)^{3 \cdot 4^{n-i}} b_{n-i}(p_1) \\
&\qquad - (1 - p_1(1-c) - (1-p_1)c)^{3 \cdot 4^{n-i}} w_{n-i}(p_1)\Big).
\end{aligned}
$$

□

Before concluding this section we want to make some remarks.

1. For $c = \frac{1}{2}$ the relation $p_2 = p_1(1-c)+(1-p_1)c$ leads to $p_2 = \frac{1}{2}$, that is, p_2 doesn't depend of p_1, and one obtains the stochastic independence of π_1 and π_2 in this case. We would expect this intuitively, because in this setting the probability that a pixel changes its color equals the probability that it doesn't, and that independently for all pixels.

2. For $p_1 = \frac{1}{2}$ and $c = \frac{1}{2}$ the second model reflects the first with $p = \frac{1}{2}$ (we get $p_2 = p_1(1-c)+(1-p_1)c = \frac{1}{2}$ and π_1 and π_2 become independent); one can check this by proving that $E_{R_1}[K]$ and $E_{R_2}[K]$ compute to the same value in this setting.

3. For $p_1 = \frac{1}{2}$ in the second model we obtain $p_2 = p_1 + c - 2p_1c = \frac{1}{2}$, that is p_2 doesn't depend of c. But only in the case $c = \frac{1}{2}$ the projections π_1 and π_2 are independent.

2.6 Real World Motivation

Images which obey the second probabilistic similarity model do exist in practice. For example in Biology where we have pictures of populations of insects or distant pictures of populations of birds. We may have a couple of such images, where the second image emerges from the first by some insects or birds changing position, or moving out of the picture scene, or entering the picture scene. Overlapping occurs between the two images. The second similarity model could hold for such a couple. Alternatively, we could have one original image and a sequence of images which are similar to this original one. The original image might be an ideal picture of our population and each of the other images might be actual photographs of such populations. In this case, overlapping occurs between the original image and each one of the similar images. Again, the second similarity model could hold for the original image and an image that is overlapped to the original one.

n	p= 0.5	p= 0.45	p= 0.4	p= 0.35	p= 0.3	p= 0.25	p= 0.2	p= 0.15	p= 0.1	p= 0.05
1	35.42	35.24	34.68	33.75	32.44	30.84	29.24	28.45	30.73	43.30
2	33.55	33.37	32.81	31.87	30.52	28.82	26.81	24.57	21.79	18.15
3	33.12	32.93	32.38	31.43	30.08	28.37	26.35	24.08	21.24	17.06
4	33.01	32.83	32.27	31.32	29.97	28.26	26.24	23.96	21.11	16.89
5	32.98	32.80	32.24	31.29	29.95	28.23	26.21	23.93	21.08	16.85
6	32.98	32.79	32.23	31.29	29.94	28.22	26.21	23.92	21.07	16.84
7	32.97	32.79	32.23	31.28	29.94	28.22	26.20	23.92	21.07	16.84
8	32.97	32.79	32.23	31.28	29.94	28.22	26.20	23.92	21.07	16.84
9	32.97	32.79	32.23	31.28	29.94	28.22	26.20	23.92	21.07	16.84

Table 1: The efficiency of overlapping for the first model.

3 Evaluation of overlapping

In order to get an impression of the storage space reduction achieved by the overlapping of quadtrees, we compute the values of our formulae for various parameters. The size of images ranges from class-1 to class-9, the black pixel probability p in the first model (resp. p_1 in the second model) ranges from 0.5 to 0.1 and the change probability c in the second model ranges from 0.2 to 0.01. Because of the symmetry of the formulae (and of our problems, of course) in relation to black and white colors the results for black (resp. white) pixel probability equal to p (resp. $1-p$) are identical to the results for black (resp. white) pixel probability equal to $1-p$ (resp. p). For this reason the black pixel probability ranges only up to 0.5.

Tables 1 and 2 present some of the theoretically calculated overlapping percentages (in other words, the space reduction percentages) for the two models respectively. More precisely Table 1 shows the fraction of $E_{R_1}[K]$ and $E_p[N]$ given by Theorem 2.3 and Theorem 2.1, respectively; Table 2 shows the fraction of $E_{R_2}[K]$ and $E_{p_2}[N]$ given by Theorem 2.4 and Theorem 2.1, respectively. $E_{R_1}[K]/E_p[N]$ and $E_{R_2}[K]/E_{p_2}[N]$ express the fraction of the average number of common nodes of two quadtrees and the average number of nodes of the second of these quadtrees for models one and two, respectively. These quantities give an indicator to the average space reduction percentages.

Based on these tables we can make some interesting observations.

n	p_1= 0.5 & c= 0.2	p_1= 0.3 & c= 0.2	p_1= 0.1 & c= 0.2	p_1= 0.5 & c= 0.15	p_1= 0.3 & c= 0.15	p_1= 0.1 & c= 0.15	p_1= 0.5 & c= 0.1	p_1= 0.3 & c= 0.1	p_1= 0.1 & c= 0.1	p_1= 0.5 & c= 0.05	p_1= 0.3 & c= 0.05	p_1= 0.1 & c= 0.05
1	66.99	60.93	37.34	74.07	68.63	45.74	81.87	77.54	57.33	90.48	87.90	74.10
2	63.61	57.72	33.33	70.56	65.21	40.72	78.54	74.18	51.44	88.04	85.35	68.52
3	62.78	56.95	32.82	69.64	64.31	40.05	77.52	73.15	50.51	86.94	84.19	67.18
4	62.58	56.76	32.70	69.42	64.10	39.89	77.27	72.90	50.28	86.66	83.89	66.84
5	62.53	56.71	32.67	69.36	64.04	39.85	77.20	72.83	50.23	86.59	83.82	66.75
6	62.52	56.70	32.66	69.35	64.03	39.84	77.19	72.82	50.21	86.57	83.80	66.73
7	62.51	56.69	32.66	69.34	64.02	39.84	77.18	72.81	50.21	86.57	83.80	66.73
8	62.51	56.69	32.66	69.34	64.02	39.84	77.18	72.81	50.21	86.57	83.79	66.72
9	62.51	56.69	32.66	69.34	64.02	39.84	77.18	72.81	50.21	86.57	83.79	66.72

Table 2: The efficiency of overlapping for the second model.

1. In all cases, for given p (resp. p_1) and c, the overlapping percentage tends to reach a limit as we deal with larger images. This limit is already closely approached for class-3 or class-4 images.

2. The overlapping percentage for statistically independent images (first model) is significant and it depends on the black pixel probability. More specifically overlapping decreases as we deviate from images with the same probability for black or white pixels. However, it is well over 20%, if p is larger than 10%.

3. The overlapping percentage for model two images is very high. As expected it gets higher for smaller change probability. However, again the percentage depends on the original image black pixel probability. For change equal or smaller than 5% and black pixel probability larger than 30% overlapping tends to be higher than 80%. For even smaller change (1-2%) overlapping gets higher than 93-95%.

4 Conclusion

The two similarity models introduced provide very promising indications for the use of overlapping in real applications, even when our images are statistically independent to each other. The analysis presented handles the two models with uniformity and permits comparisons between them when the formulae parameters take certain values, while Theorem 2.2 permits easy definition and analysis of other similarity models that might be defined. Besides, the formulae proved provide a new insight to the behavior of the quadtree when it represents random images (when randomness is based on the black-white probabilities of individual pixels) and extend the work of other researchers on quadtree storage efficiency analysis. We believe that the theoretical analysis could be extended by considering images representing shapes of various sizes, such as rectangles [7] or polygons [2].

Acknowledgements

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. The third author gratefully acknowledges helpful and inspiring discussion with Barbara Grünewald.

References

[1] F. Arcieri and E. Nardelli, An integration approach to the management of geographical information: CARTECH, *Proceedings of the 2nd Inter. Conf. on Systems Integration (ICSI)*, Morristown, New Jersey, (1992), 726-737.

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

[3] F.W. Burton, J.G. Kollias, and V.G. Kollias, A general PASCAL program for map overlay of quadtrees and related problems, *The Computer Journal* **30** (4), 1987, 355-361.

[4] F.W. Burton and J.G. Kollias, Functional programming with quadtrees, *IEEE Software* **6** (1), 1989, 90-97.

[5] K.L. Chung, A course in probability theory, *Harcourt, Brace and World, Inc.*, New York, 1968.

[6] C.R. Dyer, The space efficiency of quadtrees, *Computer Graphics and Image Processing* **19** (4), 1982, 335-348.

[7] C. Faloutsos, Analytical results on the quadtree decomposition of arbitrary rectangles, *Pattern Recognition Letters* **13** (1), 1992, 31-40.

[8] R.A. Finkel and J. L. Bentley, Quad trees: a data structure for retrieval on composite keys, *Acta Informatica* **4** (1), 1974, 1-9.

[9] G.M. Hunter and K. Steiglitz, Operations on images using quadtrees, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **1** (2), 1979, 145-153.

[10] F. Olken and D. Rotem, Sampling from Spatial Databases, *Proceedings of the 9th IEEE Data Engineering Conf.,* (1993), 199-208.

[11] H. Samet, The quadtree and related hierarchical data structures, *ACM Computing Surveys* **16** (2), 1984, 187-260.

[12] C.A. Shaffer, A formula for computing the number of quadtree node fragments created by a shift, *Pattern Recognition Letters* **7** (1), 1988, 45-49.

[13] M. Vassilakopoulos, Y. Manolopoulos and K. Economou, Overlapping quadtrees for the Representation of Similar Images, *Image and Vision Computing*, **11** (5), 1993, 257-262.

[14] M. Vassilakopoulos and Y. Manolopoulos, Analytical results on the Quadtree storage-requirements, *Proceedings of the 5th Inter. Conf. on Computer Analysis of Images and Patterns (CAIP)*, Budapest (September 1993), 41-48.