Article

Modeling and Numerical Validation for an Algorithm Based on Cellular Automata to Reduce Noise in Digital Images

Karen Vanessa Angulo, Danilo Gustavo Gil and Helbert Eduardo Espitia*

Facultad de Ingeniería, Universidad Distrital Francisco José de Caldas, Bogotá 110231, Colombia; kvangulos@correo.udistrital.edu.co (K.V.A.); dggils@correo.udistrital.edu.co (D.G.G.)

* Correspondence: heespitiac@udistrital.edu.co

Abstract: Given the grid features of digital images, a direct relation with cellular automata can be established with transition rules based on information of the cells in the grid. This document presents the modeling of an algorithm based on cellular automata for digital images processing. Using an adaptation mechanism, the algorithm allows the elimination of impulsive noise in digital images. Additionally, the comparison of the cellular automata algorithm and median and mean filters is carried out to observe that the adaptive process obtains suitable results for eliminating salt and pepper type-noise. Finally, by means of examples, the result of the algorithm are shown graphically.

Keywords: analysis; automaton; cellular; digital images; processing; noise

1. Introduction

The concept of the cellular automata (CA) was first introduced by Stan Ulam and John von Neumann in the 1940s while they were working on the Manhattan Project at Los Alamos National Laboratory [1]. A remarkable contribution to the development of cellular automata raised from the research of Slam in the study of crystal growth and the interest of von Neumann in self-replicating systems [1]. The initial idea was focused on developing a two-dimensional cellular automaton comprising a grid of square cells. Considering the neighboring cells of a certain cell, it can take a black or white state [1]. According to John von Neumann’s approach, for a certain cell, the neighborhood is composed of four adjacent squares, that is, the cellular automaton can be seen as a cross composed of square cells [1,2]. Based on the above, it is stated, in general terms, that an automaton consists of a set of cells delimited by a finite number of states, the neighborhood, and local transition rules [3].

Therefore, cellular automata are considered as discrete dynamic systems of finite spatial and temporal state composed of a finite set of cells that evolve in parallel in discrete time steps [1,4]; in other words, they are discrete in space and time, allowing the description of local interactions employing transition rules for each cell in the space. In this way, cellular automata are structures that can be used for modeling and studying complex non-linear dynamic systems [1,2,5].

Regarding the diversity of applications in which the concept of a cellular automaton is immersed, and with the inherent ease that these present to model systems, researchers and scientists extrapolated the concept of cellular automata to digital images.

Digital images are considered as a two-dimensional representation of a real image from a numerical matrix that records values for each element at a given point. Each position within the matrix is called a pixel, and it takes discrete values according to properties such as brightness (the intensity of light) or color [6].

Regarding characteristics, digital images have some called “basic”. One is the type of image: for example, there are black and white images that only record the intensity of the light that falls on the pixels. Besides, there are non-optical images such as ultrasound or X-rays in which the intensity of sound or X-rays is recorded. Resolution is also a characteristic
expressed by the number of pixels per inch (ppi), that is, the higher the resolution, the more
detailed the image is. Another is the color depth (of a color image), or “bits per pixel” is the
number of bits that describe brightness or color for one pixel. More bits allow recording
more shades of gray or more colors. Finally, the image format provides more details on
how the numbers are arranged in the image file and includes the type of compression
employed [6,7].

Digital image processing arises from the need to improve appearance, and thereby
make certain details more evident [8]. As a result, techniques were implemented to treat
different image characteristics (brightness, sharpness, contrast, intensity, noise, among oth-
ers), but particularly, methods that allow the information contained to be highlighted or
suppressed selectively in an image [8].

In this sense, one frequently covered defect is noise, since, through different devices
capable of capturing digital images, errors or interferences can be generated when transmit-
ting information bits. Theoretically, different types of noise have been defined and classified
according to their characteristics. Consequently, treatment methods and techniques have
been developed for each one [9].

Conventional techniques include filters in the space and frequency domains (high pass,
low pass, average, etc.), which compared to new techniques have shown lower results [10];
for this reason, different alternatives have been developed aiming at getting efficient and
visually aesthetic results. Therefore, the combination of methods, techniques, and various
mathematical models have been useful to achieve these objectives. Likewise, edge detection
plays a relevant role in image processing, since through these algorithms it is possible to
identify objects, define patterns, or segment information within images [9].

scheme to couple MOD (Method of Optimal Directions) and the Approximate K-Singular
Value Decomposition (AK-SVD). This approach integrates a reconstruction and learning
process into a model to removal multiplicative and additive noise. In this way, a sparse
term is used with the purpose of reducing non-Gaussian outliers associated with multi-
plicative noise. Additionally, a Laplacian Schatten norm is employed to capture the global
structure information. Other related work is presented in [12], introducing an approach on
a discriminative ridge regression to supervised classification called Discriminative Ridge
Machine (DRM). The focus here is to determine a representative model while defining
class discriminations of categorical information. The model is also extended considering
other existing models like lasso and group lasso, incorporating discriminative information.
For implementation, the authors consider a quadratic model that allows analytical solutions
in a closed-form.

According to [13], image denoising can be addressed as an inverse problem where
an approach is sparse decomposition over redundant dictionaries. In a sparse representa-
tion form, the signals correspond to a linear combination of redundant dictionary atoms.
Considering this orientation, in [13] is presented an algorithm for image denoising based
on Non Negative Matrix Factorization (N-NMF), and sparse representation over redund-
ant dictionary. In this proposal, the dictionary is trained, rooted in samples from noised
image, and then it searches for the best representation using the Approximate Matching
Pursuit (AMP). Another advancement in this orientation is presented in [14], proposing a
non-negative matrix factorization method that learns both clustering and local similarity in
a suitable form. Such portrayal exposes data’s inherent geometric property. Applying the
representation in the kernel space allows boosting the capability of the model to identify
nonlinear structures associated to the data.

Regarding applications based on Robust Principal Component Analysis (RPCA),
in [15] is exposed the issues task to eliminate both mixed types and heavy noises from
Hyperspectral Images (HSI). Authors address such issue proposing a non-convex develop-
ment in RCPA for denoising application of hyperspectral images. This approach takes the
log-determinant rank approximation and the $L_2, log$ norm to restrain the sparse properties
of column-wise or low-rank for the component matrices. Another related work can be
observed in [16] that proposes a RCPA model based on matrix tri-factorization that uses SVDs computation of small matrices. Thus, such an approach diminishes the complexity of RCPA to make it linear and completely scalable.

Regarding cellular automata and image processing, an approach to CA-based image segmentation is the GrowCut algorithm proposed in [17]. Since this is an interactive image segmentation algorithm, the user selects the set of seed pixels that signalizes the sections of interest in the image. Regarding GrowCut algorithm, in [18] is presented a procedure of interactive image segmentation designed to reduce specific image segmentation problems for identifying regions of interest. The feasibility to automatically generate seeds for GrowCut is shown; besides, authors suggest a method to automate seed generation for the segmentation task in heart images. In addition, a conventional GrowCut cellular automaton using chaotic features is enhanced in [19]. This development employs an extended, stochastic neighborhood, where randomly-selected remote neighbors reinforce the conventional local neighbors. The authors state that according to the results, by having small changes in the initial conditions in the process, major changes can be induced in the segmentation result.

Regarding others works, in [20] is proposed an algorithm for determining the optimal outdoor evacuation routes in hills. The system uses web services to obtain geographic information from Google Image. The routes are determined using graph theory with geographic information (latitude, longitude, and elevation), and cellular automata in 3D.

Meanwhile, in [21], cellular automata rules are optimized for edge detection employing Particle Swarm Optimization (PSO). In this work, it is exposed that cellular automata provides fast computation, the optimization rule, and the adaptability to target images. According to authors, the method is tunable in medical images to identify structures such as cardiac cavities.

Considering the integration of cellular automata with other techniques, in [22] is designed an automaton that could be part of a more complex system to make bio-computers that can be used for teachers in multidisciplinary education. Meanwhile, in [23] is considered a Hybrid Cellular Automata (HCA) architecture for modeling the cardiac cell–cell membrane resistance. This work shows that the modeling proposal reproduces important and complex spatio-temporal properties that can be used in future models. Besides, authors show how the GPU-based technology can accelerate the simulation and analysis of these systems. Finally, in [24], random walk theory and lattice gas cellular automata are used to generate a mathematical model for gluing wood particles.

Given the above, researchers and those interested in the subject have implemented methods, algorithms, or various techniques that combine basic theories with cellular automata to image processing and model applications. Although outstanding results have been obtained, the possibility of conducting research in this area to improve digital image processing using cellular automata is observed.

Regarding previous works, reference [25] presents an algorithm based on cellular automata to reduce impulsive noise in digital images. Meanwhile, the papers [26,27] describe the integration of a cellular neural network and an adaptive cellular automaton for the impulsive noise reduction and edge detection in digital images; finally, the document [28] displays a process for digital images edge detection employing cellular automaton.

As it can be observed, there are different approaches to eliminate noise in digital images, the cellular automata being a suitable alternative given the direct relationship between a cellular automaton and a digital image given its representation in the form of matrices. In this way, explorative research can be carried out to include different behaviors of cellular automata for digital image processing.

Proposal Approach and Document Organization

This document aims at displaying the mathematical model for an algorithm based on cellular automata to eliminate noise in digital images. The considered algorithm is
implemented in [25–27]; however, the mathematical description of the automata dynamics operation for filtering process is not performed, which is the object of study in this paper.

In this way, the contributions of the article are the presentation of the mathematical model of the CA operation considering the description of the adaptation process for the cellular automaton, and the methodology for statistical analysis using synthetic images.

The motivation in this work is that the cellular automata have a very direct relationship with digital images given their structure, then CA behaviors can be included to process images; particularly, an adaptive behavior of CA modifying the neighborhood of the cells is considered. In this context, it is important to describe this type of behavior mathematically, which can serve as a reference for further works.

The document is organized as follows. First, Section 2 introduces the background; Section 3 displays the cellular automata algorithm to eliminate impulsive noise in digital images. Section 3.2 details the model for the algorithm based on cellular automata. Next, Section 4 presents the simulation results to observe the algorithm characteristics. Finally, the discussion and conclusions are presented in Sections 5 and 6.

2. Background

Within the field of digital image processing, a large number of researchers who over time have sought to optimize different techniques, methods and methodologies that allow to have greater approximations [8,28]. However, despite the fact that many of these advances have had interesting results, in most cases, they lack support for an evaluation in a more objective sense [28].

Under this approach, the development of supports based on scientific elements for any technique developed gives greater validity not only from the visual aspect it can be evaluated, but also from its fundamental perspective.

2.1. Mathematical Modeling

Commonly in the research field, when carrying out approximations to find the solution to a problem, tools like empirical knowledge, experience in the field, and intuition are employed. However, as suggested by Pyt’ev [29], this can generate all kinds of difficulties that limit its progress, consequently, the use of tools of scientific nature and formalized knowledge allow the development to be widely supported and its validity guaranteed.

The construction, understanding and application of mathematical concepts according to Ouvrier-Buffet [30] convert informal experimentation into properly grounded scientific knowledge. Likewise, thanks to the above, it is possible for a researcher to obtain a series of expected results that can be replicated in different environments without compromising the construction.

Additionally, according to Menskii [31], the construction of a model requires three essential steps; the first is observation, where the behavior of a phenomenon is analyzed to identify its relevant characteristics. Later, it advances to the formulation stage, where a postulate is prepared and associated with such behavior; finally, the equation of the model is developed accordingly with the formulation made.

2.2. Noise Elimination

In this research field, approaches are continually being developed using different methods with interesting results, such as Sahin, Uguz and Sahin in [32], who presented an algorithm made up of two components, fuzzy logic and a CA used for the restoration of digital images contaminated by impulsive noise. The algorithm employs a local fuzzy transition rule that assigns a membership value to the corrupt pixel (in the neighborhood) and allocates the next state value as the center pixel value.

On the other hand, Chui and He in [33] designed a filter using statistical values estimated from the differences between pixel intensities compared to the most similar neighbors. This filter allows eliminating impulsive and Gaussian noise; nevertheless, when the noise level is high, the filter performance decreases.
Likewise, Chan, Ho, and Nikolova [34], propose a method divided into two phases: First, a median filter is applied to identify the candidate pixels affected by noise. In the second, the digital image is restored employing a specialized regularization method over the noisy pixels. This method has suitable results; however, when the size of the window increases, the processing resources also increase.

As previously seen, image processing has been approached from many perspectives reflected in different research points. However, the support of mathematical models provides investigative elements, since it allows to analyze the results from quantitative and qualitative perspectives.

3. Algorithm Based on Cellular Automata

Considering [35], cellular automata is a class of discrete dynamic system that allow to model complex systems. The CA is composed of a set of cells with a dimension $D$, taking 1, 2 or 3. Image processing employs CA in a two-dimensional space [36]. According to [37], a cellular automaton is composed of three parts:

(a) Cells or lattice.
(b) Neighborhood or adjacent neighbors.
(c) Rules for cell transitions.

Regarding the functioning of a cellular automaton, the state of a cell in time $t + 1$ is calculated using its current state and the values of its neighbors in time $t$ [38,39]. A $D$-dimensional cellular automata can be defined as a 4-tuple $A = \{Z^D, S, N, \delta\}$, where:

- $Z^D$: $D$-dimensional space of integers.
- $S$: finite set of to the states of $A$.
- $N$: finite ordered subset of $Z^D$ that corresponds to the neighborhood of $A$.
- $\delta$: local rule (transition function) of $A$.

The neighborhood allows us to obtain environmental information. As shown in Figure 1, Moore neighborhood has a configuration of $3 \times 3$ matrix that covers a larger number of pixels than the von Neumann neighborhood. The Moore neighborhood is described by Equation (1), where $r$ is the range, for an 8 neighbors $r = 1$.

$$N^M(x_0, y_0) = \{(x, y) \in L : |x - x_0| \leq r, \text{ and } |y - y_0| \leq r\} \quad (1)$$

![Figure 1. Two-dimensional neighborhood.](image)

In order to implement the algorithm, each cell of the CA corresponds to a pixel inside the digital image, where the value of each cell corresponds to an intensity value of the image, in a grayscale image, the values are from 0 to 255 [40]. It is noticeable that the proposed algorithm performs an adaptive modification of the neighborhood where both the first and the last row and column of the image are extended.
3.1. Cellular Automata Algorithm to Eliminate Noise in Digital Images

The mathematical model developed to describe noise elimination in digital images (using CA) is a synthesis of the algorithm presented in [25] called DK (Dynamic-Knowledge), which through the use of cellular automata and an adaptive property, eliminates noise salt and pepper present in a digital grayscale image.

A two-dimensional digital image can be represented as a matrix of size \( m \times n \), where each element \( a_{ij} \) found in the \( i \)-th row and the \( j \)-th column is a pixel within the digital image. The values assigned to the element \( a_{ij} \) corresponds to the luminosity information presented in a pixel [25]. Additionally, a cellular automaton can be defined as a discrete dynamic system composed of a set of cells with a dimension \( D \). The most used to model natural or artificial systems are the two-dimensional ones to easily recreate a collection of simple objects locally interacting with each other [2,3,25].

The DK algorithm uses a Moore neighborhood, which considers 8 neighbors from the base cell (center). Each cell of the cellular automaton can be seen as a pixel within the image, when the algorithm traverses the image using cellular automata and detects a possible noisy pixel, a function is applied to change the pixel value based on the neighbors value. Moreover, when having insufficient information from the neighbor, the cellular automata extends its neighborhood in a row and a column, that is, forming a 5 \( \times \) 5 matrix to obtain more data that lead toward the best possible decision.

3.2. Cellular Automata Behavioral Model

In order to perform the model description, firstly, the unit step function is defined since this is used to activate or deactivate the algorithm adaptive property. In this order is used function \( U(\alpha - a) \) given by Equation (2). Figure 2 graphically shows the unitary step function.

\[
U(\alpha - a) = \begin{cases} 
0, & \text{if } 0 \leq \alpha < a \\
1, & \text{if } \alpha \geq a 
\end{cases}
\]  

(2)

![Figure 2. Unitary step function.](image)

Considering \( a \) value, function \( U(\alpha - a) \) allows describing a piecewise function via turn on/off the components of a function \( F(\alpha) \). For example, \( F(\alpha) \) Equation (3) is represented as \( F(\alpha) = (1 - U(\alpha - a))f_1(\alpha) + U(\alpha - a)f_2(\alpha) \).

\[
F(\alpha) = \begin{cases} 
f_1(\alpha), & \text{if } 0 \leq \alpha < a \\
f_2(\alpha), & \text{if } \alpha \geq a 
\end{cases}
\]  

(3)

On the other hand, a matrix of dimensions 3 \( \times \) 3 or \( M_{3\times3} \) represents a part of the digital image. Each element within the matrix can be depicted in the form \( M(x, y) \), which allows establishing a relative position within the matrix, as shown in Figure 3.
Each of the eight elements surrounding the central cell \((x, y)\) is part of the Moore neighborhood. In the DK algorithm, this neighborhood is taken with a noisy central pixel and evaluates the number of neighbors that have values 0 or 255 (possible noisy pixels). If the number of neighbors with different values of 0 or 255 is less than 5, the algorithm calculates the average of these values, and based on the neighborhood rules, it changes the value of the central pixel. Otherwise, if the number of cells with values 0 or 255 is greater than or equal to 5, then the algorithm expands the neighborhood to one of size \(5 \times 5\) containing the previous one (see Figure 4), this to collect more information to lead the algorithm to make the best decision.

Considering the above, a unit step function is used to turn off the function that employs the neighborhood \(M_{3x3}\) and a new function using the neighborhood \(V_{5x5}\) is considered to calculate the average of the values of the cells that do not have values 0 and 255. Thus, the function that models the behavior of the DK algorithm is given by Equation (4).
The number of experiments was determined considering what was reported in [41] for hue corresponds to the pixel with noise. Once the number of experiments (population)

\[
DK(x, y) = (1 - U(n - 5)) \left[ \frac{1}{8 - n} \sum_{i=1}^{8} \sum_{j=1}^{8} I(x + i, y + j) - \frac{1}{8 - n} I(x, y) \right] - U(n - 5) \left[ \frac{1}{24 - m} \sum_{i=-2}^{2} \sum_{j=-2}^{2} I(x + i, y + j) - \frac{1}{24 - m} I(x, y) \right]
\]

where \( n \) is the number of pixels with values 0 or 255 in the neighborhood \( 3 \times 3 \) and \( m \) is the number of pixels with values 0 or 255 in the neighborhood \( 5 \times 5 \). Finally, \( I(x, y) \) is the pixel value at position \((x, y)\), and \( x' = x + i, y' = y + j \).

Equation (4) can be expressed as \( DK(x, y) = f_1 + f_2 \), in this order, function \( f_1 \) counts the number \((8 - n)\) of cells without salt and pepper values \( \{0, 255\} \) and calculates the average of these values. Since the summation includes the position \((x, y)\), it is eliminated from the calculated average using \( \left( \frac{1}{8 - n} I(x, y) \right) \). Finally, the result is rounded to the smallest integer employing the floor function \( f(a) = \lfloor a \rfloor \).

\[
f_1 = (1 - U(n - 5)) \left[ \frac{1}{8 - n} \sum_{i=1}^{8} \sum_{j=1}^{8} I(x + i, y + j) - \frac{1}{8 - n} I(x, y) \right]
\]

On the other hand, for \( f_2 \) function, \( U(n - 5) \) is the unit step function activated when the number of pixels with salt and pepper values \((0, 255)\) is equal or greater than 5 in the neighborhood \( 3 \times 3 \) of the cell \((x, y)\).

\[
f_2 = U(n - 5) \left[ \frac{1}{24 - m} \sum_{i=-2}^{2} \sum_{j=-2}^{2} I(x + i, y + j) - \frac{1}{24 - m} I(x, y) \right]
\]

When the number of cells in the neighborhood \( 3 \times 3 \) is equal or greater than 5, the DK algorithm takes a neighborhood of size \( 5 \times 5 \) calculating the average of all the cells in that neighborhood. Given that the sum includes the position \((x, y)\), this is removed from the calculated average employing the term \( \left( \frac{1}{24 - m} I(x, y) \right) \). Note that when function \( U(n - 5) \) turns on, \( f_2 \) is activated and \( f_1 \) deactivated. Finally, the result obtained is rounded to the smallest integer close to the calculated average.

4. Simulation Validation

In this section, a computational analysis is presented by simulation in R of the mathematical model described to evaluate the behavior of the DK algorithm compared to the conventional methods for noise elimination (median and mean filters).

In order to have simulation results with statistical validity, an experimental design based on a simple random sampling with replacement is employed, in such a way that there is a sufficient number of experiments to evaluate the performance of the algorithms. The number of experiments was determined considering what was reported in [41] for simple random sampling with replacement.

The denoising algorithms are evaluated using \( 3 \times 3 \) arrays (simulating a portion of a picture), which are randomly generated. In these experiments, the central point of this hue corresponds to the pixel with noise. Once the number of experiments (population)
is defined, the noise elimination algorithms (mean, median and DK) are executed. Then, with the results obtained, the calculation of the absolute error is made to subsequently compare the results.

The experiments were carried out via a simple sampling with replacement, that is, $3 \times 3$ matrix were randomly generated in the discrete interval of pixel values $[a, b]$; in this case, the size of the population is known [41]. An integer value of the interval is placed in each element of the array, except for position $M_{22}$, which can take the values 0 or 255. The size of the population is given by Equation (7), where $N$ is divided by 8 given the symmetries of the square (according to groups theory [42]). In this way, the sample size (number of experiments) is given by Equation (8).

$$N = \frac{(b - a + 1)^9}{8}$$  \hspace{1cm} (7)

$$n = \frac{NZ^2pq}{d^2(N - 1) + Z^2pq}$$  \hspace{1cm} (8)

In Equation (8), the respective variables are:

- $Z$: 95% confidence level (1.96).
- $p$: Probability of success (0.5).
- $q$: Probability of failure $(1 - p)$.
- $d$: Accuracy 5% (0.05).

Simulations were made in MATLAB and the samples obtained were statistically analyzed. As an example, it is taken the pixel interval $[80, 84]$. In this case, the space of all possible experiments is 48,828.

It is relevant to mention that when the number of noisy pixels within the $3 \times 3$ matrix is greater than half the number of non-noisy pixels (i.e., noisy pixels greater than 5), the matrix expands to a $5 \times 5$ size to encompass a larger number of pixels. In this way, after determining the correct or closest value of the central pixel, the matrix returns to its original size of $3 \times 3$.

With the sampling formula, using the same values taken for the parameters, 381 masks were simulated. The absolute errors between the real value of pixel $M_{22}$ and that obtained with each of the algorithms were calculated. In this order, the histogram of the absolute error displayed in Figure 5 is obtained, where the results of each algorithm are shown separately. Figure 6 shows the histogram set of data obtained in a simulation, presenting in the same scale the results for the median, mean, and DK algorithms. In these results, the DK algorithm presents smaller values.

Regarding Figures 5 and 6, it is observed that for the DK algorithm, the largest number of values of the absolute difference are close to zero. This implies that, from the total of samples obtained when executing the DK algorithm exists a greater quantity of successes at the moment of identifying the real value of the pixel. In other words, it is possible to recover the original value of the image (or a sufficiently close value) so as not to present a significant difference with the figure without noise.

In the case of the mean algorithm, although it exhibits a distribution with many values close to zero, it also presents a considerable number of pixels that are far from the real value, which implies that there are notable differences in the results after noise removal. On the other hand, the median algorithm presents results with a particular distribution (with three groups), it is observed that there is a group where the original value is recovered; however, there are two other groups far from the real value pixel, resulting in an image where impulsive noise is not satisfactorily removed.
The data were saved in three vectors and processed using the statistical software R obtaining the mean, variance, among other relevant statistical values as presented in Table 1. The basic statistics are the number of values considered in the sample (nbr-val), number of null values within the sample (nbr-null), number of missing values within the sample (nbr-na), minimal value obtained (min), maximal value obtained (max), range equal to difference between max and min (range), and the sum of all non-missing values (sum). Meanwhile, the descriptive statistics are: the median (median), the mean (mean), the standard error on the mean for a given variable (SE-mean), the confidence interval for the arithmetic mean (CI-mean) at the respective $p$-level, the variance (var), the standard deviation (std-dev), and the variation coefficient (coef-var) equal to the standard deviation divided by the mean. Finally, the normal distribution statistics are: the skewness coefficient $g_1$ (skewness), kurtosis coefficient $g_2$ (kurtosis), statistic of a Shapiro–Wilk test of normality (norm-test-SW), and the respective associated probability (norm-test-p) [43].
Table 1. Comparison of the statistics of the median and mean filters versus the DK algorithm.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Median</th>
<th>Mean</th>
<th>DK</th>
</tr>
</thead>
<tbody>
<tr>
<td>nbr-val</td>
<td>381.000</td>
<td>381.000</td>
<td>381.000</td>
</tr>
<tr>
<td>nbr-null</td>
<td>35.000</td>
<td>6.000</td>
<td>15.000</td>
</tr>
<tr>
<td>nbr-na</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>min</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>max</td>
<td>175.000</td>
<td>147.000</td>
<td>175.000</td>
</tr>
<tr>
<td>range</td>
<td>175.000</td>
<td>147.000</td>
<td>175.000</td>
</tr>
<tr>
<td>sum</td>
<td>26,233.000</td>
<td>18,353.000</td>
<td>8951.000</td>
</tr>
<tr>
<td>median</td>
<td>80.000</td>
<td>41.000</td>
<td>15.000</td>
</tr>
<tr>
<td>mean</td>
<td>68.853</td>
<td>48.171</td>
<td>23.493</td>
</tr>
<tr>
<td>SE-mean</td>
<td>3.679</td>
<td>1.930</td>
<td>1.238</td>
</tr>
<tr>
<td>CI-mean (0.95)</td>
<td>7.233</td>
<td>3.794</td>
<td>2.435</td>
</tr>
<tr>
<td>var</td>
<td>5155.663</td>
<td>1418.552</td>
<td>584.277</td>
</tr>
<tr>
<td>std-dev</td>
<td>71.803</td>
<td>37.664</td>
<td>24.172</td>
</tr>
<tr>
<td>coef-var</td>
<td>1.043</td>
<td>0.782</td>
<td>1.029</td>
</tr>
<tr>
<td>skewness</td>
<td>0.447</td>
<td>0.794</td>
<td>1.994</td>
</tr>
<tr>
<td>kurtosis</td>
<td>−1.426</td>
<td>−0.237</td>
<td>6.947</td>
</tr>
<tr>
<td>norm-test-SW</td>
<td>0.767</td>
<td>0.920</td>
<td>0.807</td>
</tr>
<tr>
<td>norm-test-p</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

In addition, the statistical summary of the median, mean filters and DK algorithm is presented in Table 2. This information can be seen more compactly in the box and whisker plot shown in Figure 7.

Table 2. Median, mean filters and DK algorithm statistical data summary.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>0.00</td>
<td>2.00</td>
<td>81.00</td>
<td>76.45</td>
<td>171.00</td>
<td>175.00</td>
</tr>
<tr>
<td>Mean</td>
<td>0.00</td>
<td>16.00</td>
<td>40.00</td>
<td>48.32</td>
<td>69.00</td>
<td>147.00</td>
</tr>
<tr>
<td>DK</td>
<td>0.00</td>
<td>7.00</td>
<td>15.00</td>
<td>23.67</td>
<td>35.00</td>
<td>171.00</td>
</tr>
</tbody>
</table>

In these results, a stronger trend is observed by the DK algorithm towards zero. Thus, on average, the data obtained with the algorithm tend to be more similar to the real values of the mask. However, as more simulations were done using fewer points with salt and pepper, the behavior of the DK algorithm and the average algorithm behaved similarly.

Additionally, calculations were made to determine the correlation of the data obtained and are presented in Table 3. A positive correlation was observed between the median and mean algorithms (filters), while DK algorithm showed very little correlation between them.
Figure 7. Box and whisker plot.

Table 3. Correlation table for the simulation result.

<table>
<thead>
<tr>
<th></th>
<th>Data Median</th>
<th>Data Mean</th>
<th>Data DK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Median</td>
<td>1.0000000</td>
<td>0.7158127</td>
<td>0.2821712</td>
</tr>
<tr>
<td>Data Mean</td>
<td>0.7158127</td>
<td>1.0000000</td>
<td>0.2301503</td>
</tr>
<tr>
<td>Data DK</td>
<td>0.2821712</td>
<td>0.2301503</td>
<td>1.0000000</td>
</tr>
</tbody>
</table>

Finally, to quantitatively and qualitatively observe the algorithm performance, Figure 8 shows the elimination of noise using the mean, median and DK algorithms, where different noise values are considered for each row having 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90% of noise level, which indicates the number of pixels contaminated in the image. Meanwhile, the first column presents the original image; later, the second column the noisy image, in the third, the image processed with the median filter; the fourth column the result with the mean filter, and finally the fifth column the filtering process with DK algorithm. As can be seen from these results, in most cases DK algorithm presents a better result than the other algorithms considered.

Considering [25,44], the Peak Signal-to-Noise Ratio (PSNR), the Signal-to-Noise Ratio (SNR), and the Structural Similarity Index Measure (SSIM) are calculated using the images in Figure 8 to evaluate the performance of DK algorithm quantitatively.

The PSNR in decibels (dB) is calculated using Equation (9), which employs the Mean Squared Error (MSE) given in Equation (10), where, \( I_r(x,y) \) is a pixel of the original image (reference image), \( I_f(x,y) \) is a pixel of the reconstructed image (filtered image), and \( B \) the number of bits employed for representing each pixel (8 bits).

\[
PSNR = 10 \log_{10} \left( \frac{2^B - 1}{MSE} \right) \text{dB} \tag{9}
\]

\[
MSE = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} [I_r(x,y) - I_f(x,y)]^2 \tag{10}
\]

The performance index SNR in dB is determined by Equation (11). This metric characterizes the quality of an image with the relationship between image power and the noise it presents [44].

\[
SNR = 10 \log_{10} \left[ \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} |I_r(x,y)|^2}{\sum_{x=1}^{M} \sum_{y=1}^{N} |I_r(x,y) - I_f(x,y)|^2} \right] \text{dB} \tag{11}
\]
On the other hand, the SSIM is calculated via Equation (12), it is used to establish the similarity between two images, allowing us to determine how different the original image $I_r$ is from the distorted image $I_f$. Values of luminance $l(I_r, I_f)$, contrast $c(I_r, I_f)$, and structure $s(I_r, I_f)$ are used for SSIM considering $I_r$ and $I_f$ [10,25,32]. The SSIM measures between two images to compare a value between 0 and 1, where one is the absolute similarity and zero the total loss of similarity.

$$SSIM(I_r, I_f) = [l(I_r, I_f)]^a \cdot [c(I_r, I_f)]^b \cdot [s(I_r, I_f)]^r$$  \hspace{1cm} (12)$$

![Figure 8. Results processing.](image)

Table 4 shows the results of the metrics with different noise levels for the algorithms evaluated using the images displayed in Figure 8, as can be seen, the noise was effectively reduced in most of the levels using DK algorithm, and the similarity was always above 0.93, which means that the filtering images with the proposed method was successful and with broad capacity to restore the image.
Table 4. Performance metric results, PSNR and SNR in dB.

<table>
<thead>
<tr>
<th>Noise %</th>
<th>Mean PSNR</th>
<th>Mean SNR</th>
<th>Mean SSIM</th>
<th>Median PSNR</th>
<th>Median SNR</th>
<th>Median SSIM</th>
<th>DK PSNR</th>
<th>DK SNR</th>
<th>DK SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>17.202</td>
<td>9.981</td>
<td>0.222</td>
<td>23.496</td>
<td>16.573</td>
<td>0.720</td>
<td>42.612</td>
<td>32.896</td>
<td>0.995</td>
</tr>
<tr>
<td>30</td>
<td>17.430</td>
<td>8.192</td>
<td>0.187</td>
<td>23.474</td>
<td>14.237</td>
<td>0.757</td>
<td>38.251</td>
<td>29.013</td>
<td>0.988</td>
</tr>
<tr>
<td>40</td>
<td>15.740</td>
<td>6.503</td>
<td>0.144</td>
<td>18.522</td>
<td>9.285</td>
<td>0.478</td>
<td>36.438</td>
<td>27.200</td>
<td>0.982</td>
</tr>
<tr>
<td>50</td>
<td>14.799</td>
<td>9.901</td>
<td>0.098</td>
<td>14.925</td>
<td>7.892</td>
<td>0.089</td>
<td>36.088</td>
<td>31.190</td>
<td>0.957</td>
</tr>
<tr>
<td>60</td>
<td>11.683</td>
<td>6.624</td>
<td>0.096</td>
<td>12.952</td>
<td>7.892</td>
<td>0.089</td>
<td>34.078</td>
<td>29.018</td>
<td>0.939</td>
</tr>
<tr>
<td>70</td>
<td>13.684</td>
<td>6.524</td>
<td>0.065</td>
<td>9.935</td>
<td>2.776</td>
<td>0.028</td>
<td>37.862</td>
<td>30.703</td>
<td>0.938</td>
</tr>
<tr>
<td>80</td>
<td>10.731</td>
<td>4.695</td>
<td>0.049</td>
<td>7.346</td>
<td>1.310</td>
<td>0.015</td>
<td>32.654</td>
<td>26.617</td>
<td>0.935</td>
</tr>
<tr>
<td>90</td>
<td>12.148</td>
<td>7.636</td>
<td>0.049</td>
<td>6.504</td>
<td>1.992</td>
<td>0.007</td>
<td>30.095</td>
<td>25.584</td>
<td>0.931</td>
</tr>
</tbody>
</table>

5. Discussion

Even though the considered cellular automata was employed in previous works [25–28], a detailed mathematical model was not addressed. Therefore, the main aspect in this paper corresponds to the mathematical description and statistical validation.

The proposed mathematical model of cellular automata can be used in a later work to carry out the respective dynamic analysis (not addressed in this work). In order to observe the algorithm features, a statistical validation for the functioning of the cellular automaton to eliminate impulsive noise is carried out.

In this paper are performed the mathematical description of the algorithm and a comparison with two very well-known algorithms, which can be considered as a comparison standard for salt and pepper noise removal. It is also relevant to mention that the algorithm is designed to eliminate salt and pepper noise, for other types of noise, it is expected to carry out the respective research to adjust the proposed algorithm.

In order to have an experimental validation, a comparison is made with two standard algorithms for impulsive noise elimination observing that DK algorithm displays a better performance; however, a broader comparison with other algorithms can be made in a further work. In this way, new strategies can also be considered to incorporate into the DK algorithm. To perform a suitable algorithm comparisons, the following aspects must be taken into account:

- Selection of the type of algorithms to be compared considering: reported performance, actuality, available code, number of citations, proposed approach of the algorithm. Some algorithms to consider consist on cellular automata-based algorithmic approaches for noise removal in digital images as Outer Totalistic Cellular Automata (OTCA) [45], and other developments like the presented in [10,46–52]; likewise, hybrid methods that incorporate cellular automata and fuzzy logic [32,53], as well as modifications and improvements of median filter as Unsymmetric Trimmed Median Filter (UTMF) [54], median-type noise detectors [34], and implementations using local image statistics [33]. Other approaches could also be considered, including algorithms based on dictionary learning methods [11,12], non-negative matrix factorization [13,14], and robust principal component analysis [15,16].

- Type of noise to eliminate considering different algorithms approaches. It can be considered noise additive, multiplicative, impulse static and dynamic noise [55]. The associated probability distribution can also be considered as: uniform, Gaussian, Poisson, Rayleigh, Speckle, Gamma, White, Brownian, and other noise characteristics like periodic and structural [56].

- Performance metrics considering the operation of the algorithms, in a way that the advantages of each algorithm, can be observed as: processing time, amount of noise removed, image distortion, Mean Absolute Error (MAE), Root Mean Square Error
(RMSE), Signal-to-Noise Ratio (SNR), Image Enhancement Factor (IEF), and Structural Similarity Index Measure (SSIM), that is a perceptual metric that quantifies image quality degradation caused by the processing; also the Peak Signal-to-Noise Ratio (PSNR) corresponding to the relationship between the maximum possible energy of a signal and the noise that affects it [44,45,53].

- Statistical tests to carry out the comparisons (ANOVA, Kruskal–Wallis, Bonferroni, etc.), considering assumptions of normality and equality of variance to establish the type of test (parametric and non-parametric), and in this way perform a fair comparison between algorithms [57,58].

Finally, the limitations of this work included to carry out the statistical tests the images used are synthetic considering only impulsive noise; besides, the algorithm operates on grayscale images, and no wide comparison with other types of algorithms is made.

6. Conclusions

In this work, the mathematical description of the algorithm based on a cellular automata to eliminate (reduce) noise in digital images is obtained. Various functions are incorporated into this model to complete this description.

The proposed model can be used to adapt the operation of the algorithm to carry out other processes on the digital image such as edge separation, equalization, and pattern identification.

A statistical validation of automaton cellular functioning to eliminate impulsive noise is carried out considering an experimental design based on a simple random sampling with replacement using a sufficient number of experiments to evaluate the algorithms performance.

Comparison of DK algorithm with other well-known techniques for noise removal (salt and pepper) is made. These results show that the algorithm obtains a suitable performance compared to other techniques. The denoising technique based on cellular automata can be considered as a non-linear type filtering.

According to the statistical analysis results carried out, it is observed from the correlation table that the DK algorithm presents an approximate percentage relationship of 28% with respect to the median algorithm, and 23% to the mean algorithm. This means that, even though the DK algorithm is based on a behavior similar to median and mean algorithms, its adaptive feature gives it the ability to expand the information with which it makes decisions and obtains effective results.

Limitations to overcome in other works are dynamic analysis of the model, also the filtering of other types of noise; operation in color images, and a wide comparison with other types of algorithms.

In a further work, the generalization of this model can be considered to be applied in color images, it can be also used to acquire or modify other color image characteristics. In addition, the application of the algorithm can also be extended to other types of noise. Besides, it could also include additional strategies in DK algorithm as neural networks, neuro-fuzzy systems, and support vector machines.

Author Contributions: Conceptualization, K.V.A., D.G.G. and H.E.E.; methodology, K.V.A., D.G.G. and H.E.E.; project administration, H.E.E.; supervision, H.E.E.; validation, K.V.A. and D.G.G.; writing—original draft, K.V.A. and D.G.G.; writing—review and editing, K.V.A., D.G.G. and H.E.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: In this work, no test were carried out on individuals (humans).

Informed Consent Statement: The study presented in this work does not involve human beings.

Data Availability Statement: No external data were required for this work.

Acknowledgments: The authors express gratitude to the Universidad Distrital Francisco José de Caldas.

Conflicts of Interest: The authors declare no conflict of interest.
References

3. Das, D. A Survey on Cellular Automata and Its Applications; Springer: Berlin/Heidelberg, Germany, 2011; Volume 269. [CrossRef]
20. Velasquez, W.; Alvarez-Alvarado, M.S. Outdoors Evacuation Routes Algorithm Using Cellular Automata and Graph Theory for Uphills and Downhills. Sustainability 2021, 13, 4731. [CrossRef]
23. Tremi, L.M.; Bartocci, E.; Gizzi, A. Modeling and Analysis of Cardiac Hybrid Cellular Automata via GPU-Accelerated Monte Carlo Simulation. Mathematics 2021, 9, 164. [CrossRef]
26. Gil-Sierra, D.G.; Angulo-Sogamoso, K.V.; Espitia-Cuchango, H.E. Integration of an adaptive cellular automaton and a cellular neural network for the impulsive noise suppression and edge detection in digital images. In Proceedings of the 2019 IEEE Colombian Conference on Applications in Computational Intelligence (ColCACI), Barranquilla, Colombia, 5–7 June 2019; pp. 1–6. [CrossRef]


39. Rosin, P.; Adamatzky, A.; Sun, X. *Cellular Automata in Image Processing and Geometry*; Springer: Cham, Switzerland, 2014. [CrossRef]


