A simple new method to accurately measure the size of polyps during colonoscopy

William Otero Regino, MD, 1 Liliana Oino Embus, MD, 2 Lina Parra Otero, MD 3

Received: 20-03-18
Accepted: 24-04-18

Abstract
Introduction: Traditionally, colon polyps’ measurements have been empirically estimated visually and with biopsy forceps, but neither method is inaccurate. Other methods have been studied but have not had the accuracy expected. The research reported here was undertaken to address this issue by building an algorithm for measure polyps from photographs taken through a colonoscope. Materials and methods: The study was done in three phases. First, an algorithm was built in MATLAB, and photos taken with a colonoscope were stored in the JPG format. In the second phase, images of objects with known sizes were checked against the algorithm to verify its accuracy. After verification of the algorithm’s accuracy, photographs of colon polyps were measured using the algorithm. In phase 3, images of polyps previously sent to three experts were measured with the algorithm. All photographs were taken with an Olympus Exera II Colonoscope. Results: For objects smaller than 5 mm, the algorithm overestimated sizes by 0.11 to 0.08 mm. For those greater than 5 mm, it overestimated sizes by 0.25 mm to 1.76 mm in those of 22 mm. The experts seriously overestimated sizes. They estimated that 7 mm polyps measured 12 mm, 8 mm polyps measured 15 mm, and 9 mm polyps measured 18 mm. Conclusion: The algorithm developed is sufficiently accurate for measuring colon polyps and is easy to obtain and relatively easy to use. It could become a tool for overcoming the difficulty of measuring polyps during a colonoscopy.

Keywords
Algorithm, MATLAB, colon polyps, biopsy forceps, pixels.

INTRODUCTION

Worldwide, colorectal cancer (CRC) is the third most common cancer in men and the second most common cancer in women. (1) In 2012, there were 1.4 million new cases and 694,000 deaths and was the third most common cause of cancer death in the world. (2) In Colombia, CCR is the fourth most frequent cause of cancer death in both men and in women. (3) In recent decades, CRC incidence has increased in developed countries unlike the underdeveloped countries where it has remained stable or even decreased. (1, 3, 4) Incidence is highest in Australia, New Zealand and Western Europe and lowest in Central Africa and Central and South Asia. (1) Most CRCs originate from adenomatous polyps of the colon following the polyp-adenocarcinoma sequence. (5) Depending on their size, the genome of these lesions may present mutations that start in the APC gene (Adenomatous Polyposis Coli) and then occur in the KRAS and p53 genes. (6) The synergism of this mutation determines that small adenomatous polyps and larger polyps become CRC. (6) This knowledge is the basis for screening colonoscopy for people 50 years of age or older to reduce CRC mortality by detection and resection of polyps found. (7, 8, 9) The intervals between colonoscopic follow-up of patients are determined by the size of the polyps. (10, 11)
Nevertheless, current methods for estimating polyp size are inaccurate. Most colonoscopists “measure” the size of polyps by “visual estimation”. This method, even when used by experts, is inaccurate and can overestimate or underestimate true polyp size. (12, 13) Measurements of resected specimens taken in pathology laboratories are smaller than these in-vivo estimates. (14, 15, 16, 17) When the size is underestimated, very long surveillance intervals can be recommended. These increase the risk of CRC appearing during the interval. (18) Numerous researchers have looked for methods to measure polyps, including through computerized axial tomography (CAT). (13, 14, 15, 16, 17, 18) Among the various methods, the simplest and most frequently used is to open a biopsy clip near the polyp to compare its size to that of the polyp. Since the size of the open clamp is known to be 5 to 6 mm, it is assumed that this will result in a better estimate. However, the margin of error of this “method” is close to 50%, up or down, even among experts. (13) A recent study that used this method found that it correctly determined the size of the polyp in 37% of the cases. The measurement was too large in 34% of cases, and too small in 29%. (13) Other methods include use of a calibrated “hood” placed on the tip of the colonoscope. (19) A study of this method demonstrated that visual estimation is more inaccurate than the hood method, but also found that the hood method is difficult to use. Recently, a complex and sophisticated new method that uses an optical probe and a personal computer has been proposed. It generates a gray scale that allows determination of the size of polyp from the endoscopic image and the distance between the objects. (20) In this investigation, the amplitude of the gray scale was adjusted constantly according to the distance between the tip of the colonoscope and the lesion, since the size of the lesion varies according to the distance. The distance between the objects was calculated according to the amount of laser light reflected from the lesion through an optical probe inserted in the endoscope’s channel. The results of this method led to the conclusion that this system can be used to achieve a more exact determination of polyp size with a conventional endoscope. Nevertheless, cost and difficulty of implementation prevent it from being viable.

Correct measurement of polyp size has two implications: It allows more accurate choice of the size of the loop and polypectomy technique, and improves recommendations for follow-up intervals. If one or more adenomas smaller than 10 mm are found, follow-up colonoscopy should be done every ten years as in screening. However, if there are three to four small polyps, or at least one that measures 10-19 mm, follow up should be done in three years. If there are more than five small polyps, or one that measures 20 mm or more, the next colonoscopy should be in one year. (21)

Taking into account the need to correctly measure polyps and the absence of a good, widely available, method, we decided to carry out this work with the following objectives.

**OVERALL OBJECTIVE**

Our overall objective was to design a method to measure colon polyps during a routine colonoscopy using an algorithm in MATLAB mathematical software.

**SPECIFIC OBJECTIVES**

1. Use the algorithm developed to measure objects with known sizes to determine algorithm accuracy.
2. Use the algorithm to estimate sizes of colon polyps from images.
3. Compare visual estimates of colonic polyp sizes from three experts with the MATLAB algorithm estimates.
4. Install the specific part of MATLAB containing the algorithm in the computers of the Gastroenterology Unit of the Clínicas Fundadores for use in routine practice.

**MATERIALS AND METHODS**

This study was carried out in the Gastroenterology Unit of Clínicas Fundadores in Bogotá, Colombia from August 2014 to August 2016. In the first phase an algorithm was developed in MATLAB software and photos were captured in JPG format with an Olympus Exera II colonoscope. In the second phase, the algorithm was used to estimate sizes of objects which had been previously measured so that the algorithm’s accuracy could be verified. Once correct functioning of the algorithm had been verified, sizes of polyps that had been photographed were estimated with the algorithm. The third phase consisted of estimating the sizes of photographed polyps that had previously been sent to three experts in gastroenterology for visual size estimates. The purpose of this phase was to determine the accuracy of the experts’ visual estimations by comparing them to the algorithm’s measurements. All the photographs used with the MATLAB algorithm were taken with an Olympus Exera II colonoscope in JPG format using the standard protocol. This protocol consists of capturing the images at the distance reached when an Olympus biopsy forceps protruding from the tip of the colonoscope and three rings on the cover of the forceps could be seen. This distance was chosen because we found that it allows an appropriate view of lesions. The biopsy forceps were reusable Olympus steel forceps with a ringed cover. Appropriate positioning was essential and occurred when three of these rings were displayed on the monitor. The standardized distance we chose...
was 13 mm. Any other device can be used to standardize a distance which can then be entered into the mathematical algorithm thereby maintaining accuracy of measurement. To develop the algorithm, the professional services of an expert MATLAB engineer were contracted. We informed the engineer of our objectives and interest in using this mathematical software containing tools for this type of design (One of us, LOE, is a biomedical engineer).

Patient preparation for colonoscopy was done in split doses in the usual way. (9) Briefly, the day before the test, the patient’s diet was normal diet until lunch but afterwards was limited to a liquid diet. At 4:00 PM, the patient started colon cleansing by taking two envelopes of polyethylene glycol with electrolytes (Nulytely™). Each packet was dissolved in one liter of water. Over two hours, the patient ingested a glass of the solution (250 mL) every 15 minutes, consuming one liter per hour for two hours. At 7:00 AM on the day of the exam, the patient repeated this process with two more packets dissolved in water and consumed over two hours. Preparation was finished at 9:00 AM. Two to three hours later, the colonoscopy was performed without sedation and with the patient in left lateral decubitus or supine decubitus. The procedures were executed by an expert colonoscopist, university professor (WOR). The colonoscopies were performed with an Olympus 180 Exera II.

Ethical Considerations

According to Resolution 008430 of 1993, this study is a low risk investigation since data external to patients are used and some data are taken from routine examinations independent of the investigation. Given the difficulty that terminology related to computer programs can cause, we consider the following basic definitions useful.

**Digital Image**

A digital image is a photograph, a drawing, an artistic work or any other “image” that is converted into a computer file. A digital image consists of a collection of orderly rows of values. (22)

**Image Format**

Digital images can be saved in various formats. Each format has a specific extension of the file that contains it. Currently, the most commonly used are BMP, GIF, JPG, TIF and PNG files. (23)

**Pixel (24)**

A pixel is the smallest visual unit in a digital image. Digital cameras and scanners capture images in the form of a pixel grid. Pixels are the points that make up digital images. Each point is called a pixel. Observing them all together creates an image. The amount of pixels that an image has indicates the quality of its resolution. In simple terms, they are the “little points” with which the images are made in the world of computing. To store the information of an image, each pixel is coded by a set of bits of a certain length (called depth of color). For example, a single pixel can be encoded with a color depth of 8 bits (1 byte). This allows up to 256 (2^8) color variants. In photographic images, 3 bytes (24 bits) are usually used to define the color of each pixel. On this basis, 16,777,216 colors can be represented. These types of images are called true color.

Pixels are also used as a unit to measure the resolution of a screen, an image and devices such as digital cameras (which use megapixels). The width and length of an image can be measured in pixels; for example, an 800 x 600 image is made up of 480,000 pixels. (25)

**Algorithm (26)**

In mathematics, logic, computer science and related disciplines, an algorithm (from the Greek and Latin dixit algorithmus, in turn, from the Persian mathematician Al-Juarismi) is a prescribed set of well-defined, ordered and finite instructions or rules that allow an activity to be carried out through successive steps that generate no doubts in an agent that performs this activity. Given an initial state and an entry, following the successive steps, a final state is reached and a solution is obtained. Algorithms are the object of study of algorithmics. (26)

**MATLAB**

It is the abbreviated name of MATrix LABoratory which was developed by MathWorks. (27) MATLAB is a very versatile software package of mathematical applications. It is one of the many sophisticated computer packages available on the market and can even be downloaded online. Like other software packages such as Maple, Mathematica and MathCad, MATLAB can be used to solve mathematical problems. (28) MATLAB offers an integrated development environment with its own programming language and can be used to perform numerical calculations with vectors and matrices. It can also work with real and complex scalar numbers, with strings of characters and with other more complex information structures. One of its most attractive capabilities is that it can make a wide variety of two and three dimensional graphics. MATLAB was originally intended for matrix algebra but over time became multi-part programming environment containing three-dimensional graphics, graphic user interfaces, interactions with other downloads and toolboxes. The toolboxes are the program packages that provide additional functionality to MATLAB. (28)
In general, medical images are stored as DICOM files (Digital Imaging and Communications in Medicine). DICOM files use the .dcm file extension. The MathWorks company offers an image toolbox that can read these files so that your data are available for processing in MATLAB. The toolbox for images also includes a wide range of functions, many of which are especially appropriate for medical images. A limited set of MRI data that has already been converted to a MATLAB compatible format is included with the standard MATLAB program. This data set allows you to test some of the image generation features available with both the standard MATLAB installation and the expanded image toolbox. MATLAB has multiple tools, with specific functions that include imdistline, subplot, uigetfile and imread.

RESEARCH PHASES

Phase 1

This first step in this phase was photographing reference objects with the colonoscope using the JPG format. These objects had known sizes that were supplied by the manufacturer. The reference objects were rubber rings which were made specifically for this investigation by the Discorreas, Mangueras and Empaques S.A. factory in Bogotá. Each ring had its own code and certification printed on the external and internal surfaces. The sizes of the reference objects are shown in Table 1.

Table 1. Rubber rings with known diameters (reference standards)

<table>
<thead>
<tr>
<th>Reference Object</th>
<th>Internal diameter (mm)</th>
<th>External diameter (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-001</td>
<td>0.74</td>
<td>2.78</td>
</tr>
<tr>
<td>2-002</td>
<td>1.07</td>
<td>3.61</td>
</tr>
<tr>
<td>2-003</td>
<td>1.42</td>
<td>4.46</td>
</tr>
<tr>
<td>2-004</td>
<td>1.78</td>
<td>5.34</td>
</tr>
<tr>
<td>2-005</td>
<td>2.57</td>
<td>6.13</td>
</tr>
<tr>
<td>2-006</td>
<td>2.90</td>
<td>6.46</td>
</tr>
<tr>
<td>2-012</td>
<td>9.25</td>
<td>12.81</td>
</tr>
</tbody>
</table>

These photos were taken at the same distance from the reference objects as that used to take photos of the polyps (as previously described). Although this distance was not used to make calculations, nor did it influence the development of the algorithm, it was used to standardize captures of images so that each image was taken at the same distance so that they were all comparable and biases were avoided (Figure 1).

To estimate sizes of objects, a simple algorithm was designed by expert engineers using functions contained in the MATLAB toolbox.

First, we verified that the algorithm was efficient using different photographs in JPG format. The program was considered to have run properly when the image could be loaded and manipulated with imdistline. The program successively asks the user to enter a number in pixels and then provides a result. The following functions were used in the sequence of operations: “uigetfile”, “imread”, strcat”, “imshow”, “subplot”, “imdistline”, “Inputdlg”, and “str2double”. Four variables were created on the basis of these functions. Two of them (A and B) had hypothetical values that were provided by the algorithm for calculation of the value of the fourth variable (D). After this, the value of the third variable (C) is calculated by using the rule of three. Variable A stores a fixed amount of data in pixels, variable B stores a fixed amount of data in millimeters, variable C saves a value that the user enters and variable D performs the rule of three with the data stored in variables A, B and C.

The “msgbox” function shows the result in millimeters that was calculated in the previous process. Once the correct functioning of the algorithm was verified, the number of pixels contained in a certain area, in this case 5 mm, is calculated. To find this value, a photograph in JPG format is placed on a millimeter sheet similar to the one used to measure the rubber ring in Figure 1. A rubber ring with a known size used and then its size was estimated with algorithm that had been established. The four quadrants near the position of the biopsy clip were chosen because this is the position where the polyp is visualized. The “imdistline” function is used to draw a horizontal line on each of the quadrants of the lines of the grid with known size (5 mm) and the pixel value is obtained. The photo we took is shown in Figure 2. The upper and lower right quadrants both measure 172.00 pixels, while the upper and lower left quadrants both measure 139.00 pixels. The values of the grids in pixels differ because the image is convex.
The values obtained in pixels of each of the 5 mm quadrants and the average of these were stored in the algorithm as the reference standard. These values are shown in Table 2.

Table 2. Numbers of pixels in 5 mm quadrants

<table>
<thead>
<tr>
<th>Quadrant measured</th>
<th>Pixels</th>
<th>Size of grid in mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper Right Quadrant</td>
<td>172.00</td>
<td>5 mm</td>
</tr>
<tr>
<td>Upper Left Quadrant</td>
<td>139.00</td>
<td>5 mm</td>
</tr>
<tr>
<td>Lower Right Quadrant</td>
<td>172.00</td>
<td>5 mm</td>
</tr>
<tr>
<td>Lower Left Quadrant</td>
<td>139.00</td>
<td>5 mm</td>
</tr>
<tr>
<td>Average pixels</td>
<td>155.5</td>
<td>5 mm</td>
</tr>
</tbody>
</table>

In this first phase, the resolution of the photographs was also verified with respect to the dimensions in pixels and the depth in bits (Shown in Figure 2A and Figure 2B). This verified that all the photos taken in the standardized way contained the information in pixels and depth, regardless of the color of the image and the background color. This contrasts with the method of image segmentation in which the object of study is immersed in the environment and cannot be clearly highlighted by adjusting the number of pixels. With the method chosen in this investigation, the information depends on the distance at which the image is captured during photography since the charge-coupled device (CCD) of the camera of the colonoscope has the property that all the photos taken have the same depth in bits. In this case it was 24 bits, as shown in Figure 3.

We used neither Image processing nor segmentation for these photos, although both can be done in MATLAB because the pixels contained within the object (in this case, the polyps) are similar to or are the same as those in the background of the object (mucosa of the colon).

Phase 2

The final reference pattern of the algorithm, 155.5 pixels contained in a 5 mm grid (see Table 2), allows calculations of the sizes if images entered in JPG format. In this phase, photos of objects with known sizes were used to test the estimates of the algorithm. These objects included photographs of the rubber rings, coins of different denominations, and a skin mole of a volunteer. After loading the image in the algorithm, the “imdistline” function was used and the diameter of the object was measured by drawing a line from one point to another point on the object under.
study (Figures 4, 5, 6 and 7). The “inputdlg” function opens in a separate window in which the number of pixels previously seen can be entered (Figures 4C, 5C and 6C). Then, the value in millimeters is produced (Figures 4D, 5D and 6D). Other elements with known diameters were also used with the algorithm to establish its accuracy. These included coins of various denominations, open biopsy forceps, a skin mole and colon polyps. These objects are shown in Figures 4, 5 and 6.

In this phase several polyps were also measured using the algorithm. Figure 7 shows one of these, a pedunculated polyp.

Phase 3

After finding that the algorithm built in MATLAB had excellent concordance, it was used as the reference test for the third phase. In this phase, photographs in JPG format, taken using the methodology described above, were sent to three expert gastroenterologists for visual estimates of their sizes. The sizes they estimated visually were compared with those produced by the algorithm.

RESULTS

The values in pixels and in millimeters that were calculated by the MATLAB algorithm are shown in Figures 5, 6 and 8. Their consolidated values are shown in Table 3. The same table compares the MATLAB algorithm values of each element with each object’s real measurements. It also shows the variations of these two measurements. The signs (+) and (-) indicate that the size supplied by MATLAB is “more than” or less than” the actual value.

This table shows that the algorithm gives a slightly larger size (one tenth of 1 mm), overestimating sizes by between 0.11 and 0.08 mm, when the object measures less than 5 mm. On the other hand, it underestimates sizes by 0.25 mm when objects are larger than 5 mm. When the object measures 22 mm, the error is 1.76 mm. With these minimal variations, the kappa match is almost perfect.

After verifying the accuracy of the algorithm with objects whose real measurements were known, the size of a polyp found in a colonoscopy was calculated. Its value of 11.09 mm is shown in Figure 8.

Table 4 shows a comparison of visual measurements by experts and measurements estimated by the algorithm.

The experts estimated that the 7, 8 and 9 mm polyps measured 12, 15 and 18 mm, respectively.

DISCUSSION

An algorithm designed with MATLAB tools for this study was used to determine the size of real objects with a high
Figure 4. Calculation of the size of a known object in millimeters. A. Photo of a known object (rubber ring), that had already been entered into the algorithm. B. The "imdistline" function is used to measure an object whose size is already known. The line is drawn and the measurement in pixels is obtained. C. The pixel measurement is entered into the algorithm. By clicking "OK", the measurement is translated into millimeters.

Figure 5. Calculation of the size of a known object in millimeters. A. Object with known size (100 peso coin). B. Measurement in pixels with "imdistline". C. Window for measurement in pixels. D. Result window showing diameter of the object in millimeters.
Figure 6. A. Skin lesion with known size. B. Measurement in pixels by "imdistline". C. Window to enter measurement in pixels. D. Result window showing diameter of the object in millimeters.

Figure 7. Colon polyp to be measured with the algorithm. Photograph (JGP) taken with colonoscope during a routine colonoscopy.

Table 3. Comparisons of objects with known diameters (reference patterns) and algorithm

<table>
<thead>
<tr>
<th>Reference object</th>
<th>Diameter according to algorithm (mm)</th>
<th>Real external diameter (mm)</th>
<th>Differences between measurements (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ring 2-001</td>
<td>2.86</td>
<td>2.78</td>
<td>+0.08</td>
</tr>
<tr>
<td>Ring 2-002</td>
<td>3.65</td>
<td>3.61</td>
<td>+0.04</td>
</tr>
<tr>
<td>Ring 2-003</td>
<td>4.66</td>
<td>4.46</td>
<td>+0.2</td>
</tr>
<tr>
<td>Skin lesion</td>
<td>4.11</td>
<td>4</td>
<td>+0.11</td>
</tr>
<tr>
<td>Ring 2-004</td>
<td>5.09</td>
<td>5.34</td>
<td>-0.25</td>
</tr>
<tr>
<td>Ring 2-013</td>
<td>13.39</td>
<td>14.40</td>
<td>-1.01</td>
</tr>
<tr>
<td>Colombian 100 peso coin</td>
<td>20.24</td>
<td>22</td>
<td>-1.76</td>
</tr>
</tbody>
</table>

degree of agreement with their real measurements. Once adequate performance of the algorithm had been verified, the sizes of colon polyps were estimated. As far as we know, this is the first study that uses MATLAB to estimate the size of colon polyps. With the method developed in this research using an algorithm based on MATLAB mathematical software, measurements are practically exact. For objects smaller than 5 mm, it overestimated sizes by bet-
A simple new method to accurately measure the size of polyps during colonoscopy. The experts gave values of 12 mm, 12 mm and 13 mm, respectively, with variations above the real diameter of 50% and 80%. For 11.2 mm polyps, the overestimation was 33% to 40%. If we relied on the measurements of these three experts, patients with polyps that had diameters of 7 and 8 mm would have been scheduled for follow-up colonoscopy within three years since all of the expert estimations were over 10 mm. With these errors of overestimation, costs increase and endoscopy services become congested due to multiplication of unnecessary procedures generated by requests based on erroneously reported sizes over 10 mm. (10, 11) This also represents unjustified costs due to the large number of colonoscopies indicated throughout the country. The results obtained in this investigation allow us to conclude that the accuracy achieved is superior to the other methods now in use.

Table 4. Polyp size estimates by expert gastroenterologists compared with MATLAB algorithm estimates

| Table 4. Polyp size estimates by expert gastroenterologists compared with MATLAB algorithm estimates |
|----------------------------------------|----------------|----------------|----------------|----------------|----------------|----------------|
| Algorithm estimate in mm | Estimate by expert 1 in mm | Difference in mm between algorithm and expert 1 | Estimate by expert 2 in mm | Difference in mm between algorithm and expert 2 | Estimate by expert 3 in mm | Difference in mm between algorithm and expert 3 |
| Polyp 1 | 8.0 | 12.0 | +4.0 | 15.0 | +7.0 | 12.0 | +4.0 |
| Polyp 2 | 10.4 | 8.0 | -2.4 | 8.0 | -2.4 | 15.0 | +4.6 |
| Polyp 3 | 7.0 | 12.0 | +5.0 | 12.0 | +5.0 | 13.0 | +6.0 |
| Polyp 4 | 9.4 | 15.0 | +5.6 | 18.0 | +8.6 | 13.0 | +3.6 |
| Polyp 5 | 11.2 | 15.0 | +3.8 | 16.0 | +4.8 | 15.0 | +3.8 |

Input a value

Input the value in pixels

The diameter value is 11.0932 mm

Figure 8. Algorithm polyp measurement.

When the visual estimations of experts was compared with the estimations of the algorithm, the experts measurement varied between 50% and 80% above 8 mm polyps with values of 12 mm, 15 mm and 12 mm, respectively. For 7 mm polyps, the experts gave values of 12 mm, 12 mm and 13 mm, respectively, with variations above the real diameter of 50% and 80%. For 11.2 mm polyps, the overestimation was 33% to 40%. If we relied on the measurements of these three experts, patients with polyps that had diameters of 7 and 8 mm would have been scheduled for follow-up colonoscopy within three years since all of the expert estimations were over 10 mm. With these errors of overestimation, costs increase and endoscopy services become congested due to multiplication of unnecessary procedures generated by requests based on erroneously reported sizes over 10 mm. (10, 11) This also represents unjustified costs due to the large number of colonoscopies indicated throughout the country. The results obtained in this investigation allow us to conclude that the accuracy achieved is superior to the other methods now in use.
This original research was designed to solve the universal problem of accurately measuring colon polyps. At this time, we have already begun using our system to measure esophageal and stomach polyps.

**Recommendations**

Implement this method in endoscopy rooms.

**Conflicts of interest**

None.

**Financing**

The costs of this investigation were borne entirely by the researchers.

**Acknowledgement**

Although he was hired as an outside consultant, we greatly appreciate Anuar Steven García Gutiérrez, our systems and computing engineer and MATLAB expert from Antonio Nariño University for his dedication, constant willingness to resolve our doubts, and his enthusiasm for all of our concerns.

**REFERENCES**


A simple new method to accurately measure the size of polyps during colonoscopy