Crater Marking and Classification Using Computer Vision

Alejandro Flores-Méndez

LIDETEA, Universidad la Salle
México, D.F. México
aflores@ci.ulsa.mx

Abstract. In the last three years NASA and some other Space Agencies have draw some interest to date Mars surface, mainly because the relationship between its geological age and the probable presence of water beneath it. One way to do this is by classifying craters on the surface attending to their degree of erosion. The naïve way to solve this problem would let a group of experts analyze the images of the surface and let them mark and classify the craters. Unfortunately, this solution is unfeasible because the number of images is huge in comparison with the human resources any group can afford. Different solutions have been tried [1], [2] over this period of time. This paper offers an autonomous *Computer Vision System* to detect the craters, and classify them.

1   Introduction

In the past three years a number of studies under the supervision of NASA and some other Research Institutes and Space Agencies have been done to date the geological age of some celestial bodies. An important number of works over this issue center on Mars surface, mainly because of the relationship between its relative geological age and the probable presence of water beneath the surface.

One way to assign a planet or a moon its relative geological age is by dating craters on it. The first naïve approach one might think to solve this problem would dedicate a group of experts analyze a series of images from Mars surface and let them classify the craters that appear on them. Nevertheless there are not enough human resources available to dedicate a team to do this. In order to solve this problem, a number of different solutions have been studied. One project, under the name of *clickworkers*, proposed by NASA investigators [1] put a group of grayscale images from the NASA database of Viking Orbiter Mission to Mars on the Internet (for a further look see http://clickworkers.arc.nasa.gov/top). The project let anyone who was willing to help, after doing a very basic instruction, signal the position (mark) and classify craters within a series of images that are presented to the collaborator, who received the title of *clickworker*. The system presents every image it has, not to one, but to many clickworkers who give their opinion about the position and class the craters have. The system collects these opinions and obtains the consensus of them. Finally it *colors* a map of Mars based on these information. Some of the results obtained from the clickworkers project (CP) were quite similar when compared with the solutions of the

experts. The problem with this project is that it still depends on the human factor. Some projects have tried to automate this labor, being one example the work done by Negrete [2]. On his work, Negrete gave an autonomous system based basically on the Hough Transform to detect and mark craters over an image, and after that, the system used ontologies to classify the craters that were detected. Sadly, this investigation was not able to detect more than 60% of the craters. The purpose of this paper is to offer yet another computer vision system [3] for the marking and classification of the craters using a number of different image processing techniques [4] and some others, like Fuzzy Logic [5]. A scheme of the proposed system is presented in the following figure.
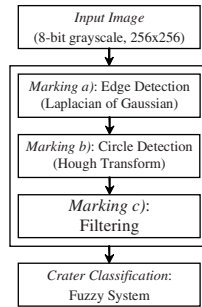
```
┌─────────────────────────────┐
│        Input Image          │
│   (8-bit grayscale, 256x256)│
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│ Marking a): Edge Detection  │
│   (Laplacian of Gaussian)   │
├─────────────────────────────┤
│ Marking b): Circle Detection│
│     (Hough Transform)       │
├─────────────────────────────┤
│       Marking c):           │
│        Filtering            │
└─────────────────────────────┘
              │
┌─────────────────────────────┐
│   Crater Classification:    │
│       Fuzzy System          │
└─────────────────────────────┘
```

Fig. 1. Scheme of the system proposed for crater marking and classification

A detailed explanation of the system and the results obtained from it are presented in the following sections.

2   Computer Vision System

The system proposed here is very simple, and it was intended to be this way, since the algorithm should be as fast as possible so it can manipulate a large volume of data. That is why a number of processes like equalization of the image, mathematical morphology and clustering among others are omitted at this point. In spite of the system's simplicity, the results obtained were in general satisfactory. Moreover, some techniques were ruled out because the system was not truly enhanced when these were added. For example, equalization was omitted because the images captured by the satellite are almost equalized. Nevertheless, when this is not the case, the classification system absorbed most of these differences. In the following subsections a deeper description of the marking and classification subsystems will be presented.

2.1   Crater Marking Subsystem

Marking of the crater is a three step process (see Fig. 1). First, we must obtain the borders of the image. A number of different techniques are available for this matter. An analysis of the nature of the images suggested that the Laplacian of the Gaussian Method (LOG) is a good choice in this case. Some other methods were tried, among

them, Sobel (which is suggested by Negrete in his work [2]), but our experiments favored LOG in most of the cases. The better results for LOG were attained when using a threshold of 0.006 and a neighborhood of 2. These borders were feed as an input to the Hough Transform [6] to detect the circles (craters) on the image.

Hough Transform. As already has been stated, the Hough Transform (HT) is used to obtain the probable locations of the craters. The general case of the HT was developed to detect lines, but the generalization of the Transform to detect some other geometric bodies is straightforward in many cases. An example of the pseudocode[1] for the HT that detects the circles within an image, given the set of coordinates of the borders, a radio r, and the number of samples to be taken over the perimeter of a circle would be as follows:

```
HT_circles (borders, r, sample_rate)
initialize (acum)
for i ← 1 to length [borders]
   (x, y) ← borders [i]
   angle = 0;
   repeat
      angle = angle + sample_rate
      a ← round (x – r * cos (angle))
      b ← round (y – r * sin (angle))
      acum [a, b] = acum [a, b] + 1
   until angle > 2π
return acum
```

Nevertheless, a number of problems arise when the HT is used for detection of a geometrical shape such as incomplete borders or deformed shapes. For the project, both things must be accounted for. One way to deal with the geometric body deformations is by using a variation of the HT called sliding window (HTSW). The HTSW uses a neighborhood around the point being analyzed, which enhances the local maxima stored in the accumulator. For the experiments an n by n neighborhood was selected with the evaluated point as its center. The value of n was obtained as a function of the radio (r) through the following expression:

$$n = \begin{cases} 2 & 15 \leq r \\ 1 & r < 15 \end{cases}. \tag{1}$$

To solve the problem of incomplete borders, the system takes into account a threshold value $0 < \theta \leq 1$ which measures the percentage needed to detect a crater. The best results were obtained for $\theta = 0.3$. The marking process using a biased HTSW system is exemplified in the following figure.

---

[1] The pseudocode follows the conventions proposed on Cormen's book [7]
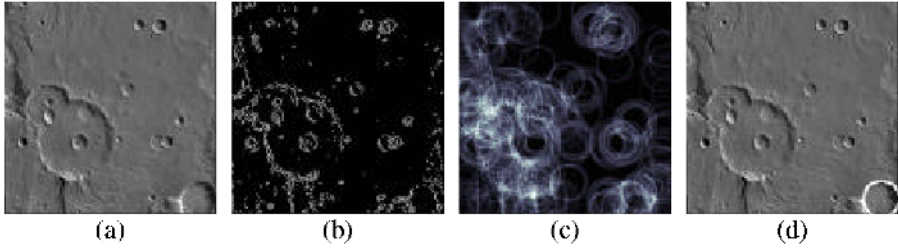
Fig. 2. The marking process takes an image of Mars surface (a) obtains its borders with the LOG (b) and then uses this borders for the HTSW (c) to detect craters of a particular radio (d)

Filtering. Once a crater has been marked, the borders associated with this crater are deleted (since two different craters will have their own rims). By eliminating these borders, not only the process is accelerated, but this also reduces the number of false detections. After deletion is done there are still two problems that the system handles. The first problem was that some craters were detected because the region selected was not a crater, but the borders from a group of different geological accidents. To avoid this, the system sums the borders in that area. If the number of borders not used in the detection of the crater is greater than a threshold, the detected crater is discarded, unless there is clear evidence of the presence of the crater (the accumulator shows that at least 80% of the elements were accounted). An example of how deletion and filtering contributed in the detection process is presented in the following figure.
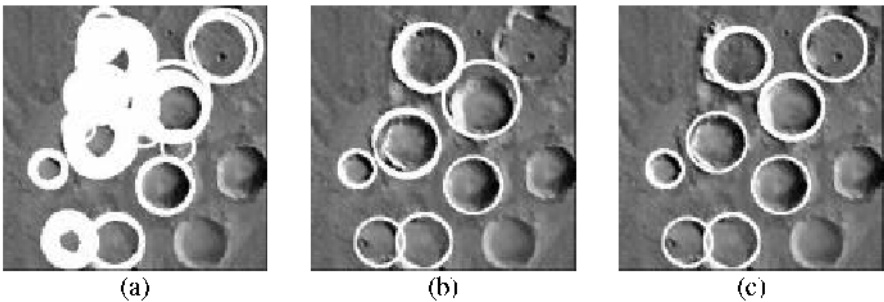

Fig. 3. Series of images showing the craters detected with the HTSW (a), detection using deletion of the borders used (b) and detection with deletion and filtering (c)

2.2   Crater Classification Subsystem

The crater classification, as stated in the CP proposes that there are three basic categories called *fresh*, *degraded* and *ghost* related with crater aging. Depending on the aging, each crater will have a number of features that can be looked upon. The definition given by CP states that "…a fresh crater displays a sharp rim, distinctive ejecta blanket, and well-preserved interior features (if any). In a degraded crater the surrounding ejecta blanket is gone, interior features are largely or totally obliterated

and the rim is rounded or removed. Finally, the ghost craters are faintly visible through overlying deposits". The next figure shows an example of these classes.
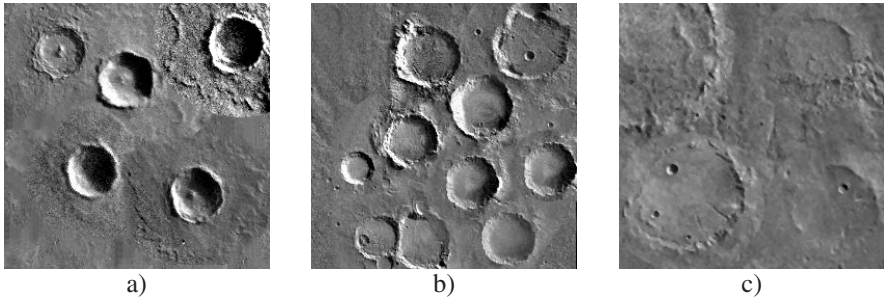

a)   b)   c)

Fig. 4. The basic classification proposed by the CP states that there are three categories based on crater aging. The craters could be fresh (a), degraded (b) and ghost (c)

A common problem with automated classification arises when the rules that define a category are ambiguous or there is some level of "spatial vagueness". On these cases a different approach like a fuzzy logic system is needed. Fuzzy systems have been applied for a number of different problems of classification such as medical applications [8], soil classification [9], fish grading [10], etc. For the crater classification process, a number of problems are inherited from the ambiguity and vagueness of the definitions, and because of this, the use of a fuzzy system was considered. The system proposed in fact is a supervised fuzzy system.

Fuzzy System. In the preceding lines, the reader learned that crater classification is related with a number of features to be looked upon the image. Nevertheless, for an untrained eye, trying to determine how recent the crater is might be equivalent to simply say how deep the crater is. Luckily this very simple feature proved to be enough for the vast majority of the cases. A simple observation related with this feature (how deep a crater is) is that the number of bright pixels for a fresh crater (Fig. 5(b)) is considerably larger than the same number for a ghost crater (Fig. 5(d)). Another way to say this is that the distribution of the graylevels for a fresh crater tends to be uniform while Normal distribution is better suited for ghost craters.
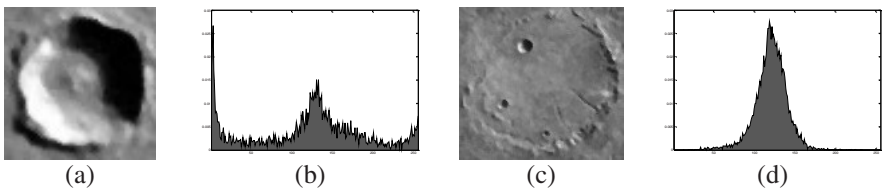

(a)   (b)   (c)   (d)

Fig. 5. Histograms (b), (d) of a fresh (a) and a phantom crater (c)

The fuzzy system proposed here, will use as inputs the maximum over the first and the second quartiles ($\max Q_1$, $\max Q_2$) of the histogram of the square region containing a crater detected from the marking process. For simplicity, the fuzzy input

variables will be labeled after the values they receive, *i.e.* $\max Q_1$, $\max Q_2$. Each of these fuzzy inputs will have three different sets called *small*, *medium* and *large*; which were determined experimentally. As output, a single variable is used. This variable receives the name of *category* and it contains three fuzzy sets (also determined experimentally) called *fresh*, *degraded* and *ghost*. The sets of the input variables are trapezoidal, while the sets for the classes are triangular as can be seen in Fig. 6. The inputs variables are related with the category simply by an AND operator as reflected in the following table.

Table 1. AND Relationship between fuzzy input variables $\max Q_1$ and $\max Q_2$ and the fuzzy output variable *category* for the classification system

$\max Q_2$ $\max Q_1$	*small*	*medium*	*large*
small	*degraded*	*ghost*	*ghost*
medium	*degraded*	*degraded*	*degraded*
large	*fresh*	*fresh*	*degraded*

For the defuzzification, the mean of maximum method was selected. The combination of this method together with the type of sets choose for the classes allowed to determine, by using a simple hard limiter nonlinear function, in which category the crater was. As an example, the following figure shows the results of the fuzzy system when it was feed with the pairs (0.0267,0.0124) and (0.0009,0.0276). These pairs represent the input $\left(\max Q_1, \max Q_2\right)$ for the craters in Fig. 5(a) and (c).

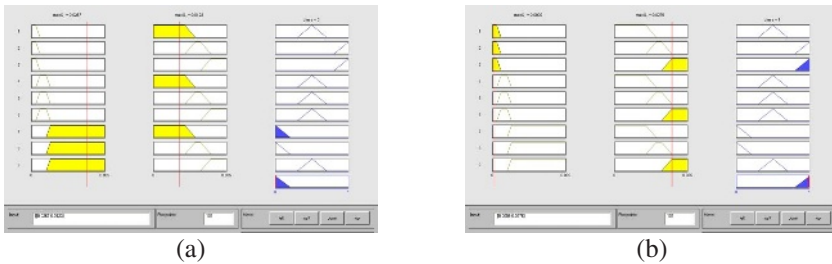(a)   (b)

Fig. 6. When the system was feed with the inputs for the craters shown in Fig. 5(a) (a) and the ghost crater shown in Fig. 5(c) (b), it correctly classified the craters

The results obtained by this system correctly classified more than 90% of the craters marked.

3   Results

To determine the factors used for every part of the system a number of experiments were conducted over a set of 100 grayscale images of 256 x 256 pixels containing little less than 300 craters. The experiments were carried out to obtain first the

parameters related with the marking process. After the best set of parameters was determined from these results the performance of the classifier was obtained.

3.1   Experiments for the Marking

For the marking process, a group of 30 images containing approximately 100 craters were used. A number of border techniques were tested. In combination with the border techniques, different approaches were tested as the HT, HTSW, HTSW with border deletion (HTSW/D) and HTSW with deletion and filtering (HTSW/DF). The following table (Table 2) contains the results for some schemes that were studied. The first number of the combination refers to the percentage of craters detected, while the second number refers to the percentage of the false detections. For example, when the HTSW with Deletion and Filtering (HTSW/DF) was selected to detect craters with a threshold $\theta = 0.3$ using the Sobel Technique with a threshold of 0.06, 53.12% of the craters were detected, but 23.52% of the crater detections were false. In some cases, the number of false detections for a particular combination was above 50%. When this is the case, the data from the experiments is not presented in the table.

Table 2. Statistcs obtained from some of the different combinations tried for crater marking

	Canny 0.175, 0.05	Sobel 0.05	Sobel 0.06	LOG 0.0055	LOG 0.006	LOG 0.0065
HTSW/DF $\theta = 0.25$			65.62% 41.83%		81.25% 43.48%	66.67% 42.11%
HTSW/DF $\theta = 0.3$		47.75% 46.14%	53.12% 23.52%	68.75% 42.42%	71.88% 22.58%	65.63% 19.23%
HTSW/DF $\theta = 0.35$	37.50% 38.09%	21.50% 31.17%	48.48% 11.76%	46.87% 22.22%	43.75% 14.29%	40.62% 4.17%

It is important to point out that the output for the HT and HTSW in general detected more than 50% of false detections. From the results presented in the table, it can be said that the best ratio between true and false recognition was obtained for the HTSW/DF when the threshold was set at 30% and the LOG Method was choose for the borders with a threshold of 0.006.

3.2   Results for the Classification

For the classification process we let the best system (HTSW/DF and LOG with proper parameters) mark the craters of 100 images which contain little less than 300 craters. After marking was done 200 images containing craters (no false detection was selected) were manually classified using the CP criteria. The classification subsystem was feed with the images previously selected and its results were compared with the ones manually obtained. The results generated in this way agreed in 91.5% of the cases.

4   Conclusions and Further Work

The time used by the system to mark and classify the craters (less than 5 seconds per image in average using a Pentium 4, 2 GHz, 512 MB RAM computer) makes it possible to think that this process can be used in practice. Nevertheless, if the process is to be useful, the marking subsystem needs to be further developed to obtain at least an 85% of recognition while the false recognition is kept below 7%. There are some signals from some research lines being studied at this moment that both percentages can be achieved, but further work will be needed. Sadly, recognition close to 100% has already been discarded, because the recognition of some "hard to identify" craters (usually phantom craters) implies almost certainly an increment in the number of false recognitions. For the fine tuning of the detection, is probable that Mathematical Morphology will be helpful. Still, more experiments in this sense are to be accomplished.

On the other hand, the classification subsystem proved not only to be a good classifier but also it was very robust because it was capable of correctly classify craters even when the area contained only partially the crater, the area was bigger than the one from the crater or the area contained some other elements besides the crater. The success of this part of the system, in my opinion, resides mainly on the selection of the histogram based feature, and because of this, some other techniques different from fuzzy logic might be used with this feature as its input to classify the craters. Still, I do not believe that great improvements can be done in this line since the classification remain being a subjective process, and that's why some results in this area might vary but not meaningfully.

References

1. Kanefsky, B., Barlow, N. G., Gulick, V. C.: "Can distributed volunteers accomplish massive data analysis task?", Lunar and Planetary Science XXXII, paper 1272 (2001).
2. Negrete, V.: Crater Image Classification using Classical Methods and Ontologies, M. Sc. Thesis, University of Houston (2002).
3. Parker, J. R.: Algorithms for Image Processing and Computer Vision, John Wiley & Sons (1996).
4. Gonzalez, R. C., Woods, R. E.: Digital Image Processing, Addison-Wesley Pub Co., 2nd edition (2002).
5. Zadeh, L A.: A theory of approximate reasoning. Machine Intelligence 9, ed. J. Hayes. D. Michie, L. I. Mikulich, 149–194. New York, Halstead Press. (1979).
6. Sonka M., Hlavac V., Boyle R.: Image Processing, Analysis and Machine Vision, Chapman and Hall Computing, London, (1993).
7. Cormen T. H. (ed.): Introduction to Algorithms, 2nd. Edition, MIT Press, (2001).
8. Belacel, N., Boulassel, M. R.: Multicriteria Fuzzy Classification Procedure PROCFTN: Methodology and Medical Application, Fuzzy Sets and Systems (2003) in press.
9. McBratney, A. B., deGruijter, J. J.: A continuum approach to soil classification by modified fuzzy k-means with extragrades. Journal of Soil Science 43: 159–75, (1992).
10. Hu, B. G., Gosine, R. G., Cao, L. X., de Silva, C. W.: IEEE Trans. on Fuzzy Systems, Vol.6, No.1, February (1998).