Astron. Nachr./AN **xxx** (xxxx) x, xxx–xxx

A fuzzy logic-based algorithm for cosmic-ray hit rejection from single images

L. SHAMIR

Department of Computer Science, Michigan Technological University,
1400 Townsend Drive, Houghton, MI 49931, USA

Received <date>; accepted <date>; published online <date>

Abstract. An algorithm for cosmic-ray rejection from single images is presented. The algorithm is based on modeling human perception using fuzzy logic. The proposed algorithm is specifically designed to reject multiple-pixel cosmic ray hits that are larger than some of the point spread functions of the true astronomical sources. Experiments show that the algorithm can accurately reject ∼97.5% of the cosmic rays hits, while mistakenly rejecting 0.02% of the true astronomical sources. The major advantage of the presented algorithm is its computational efficiency.

Key words: methods: data analysis — techniques: image processing

1. Introduction

The presence of cosmic ray hits in astronomical CCD frames is frequently considered as a disturbing effect. Cosmic rays add an undesirable signal to astronomical images, and can weaken algorithms for astronomical image processing. For instance, it can decrease the compression factor of astronomical image compression algorithms (Offenberg et al. 1999), and can disturb the operation of autonomous astronomical pipelines (Axelrod et al. 2004; Becker et al. 2004; Otuairisg et al. 2004). Exposures taken at high altitude observatories get more cosmic ray hits than sea level observatories. This becomes even more significant in space-based telescopes located far from a planetary magnetic field (Offenberg et al. 1999).

Several methods for cosmic ray hit rejection have been proposed. One common technique is by comparing several exposures of the same field (Fixsen et al. 2000; Windhorst, Franklin & Neuschaefer 1994; Shaw & Horne 1992). However, exposures of the same fields are not always available. Other approaches, such as (Salzberg et al. 1995; Rhoads 2000; Van Dokkum 2001; Pych 2004), have been proposed in order to perform cosmic ray hit rejection in a single CCD exposure. Rhoads (2000) performs filtering by adapted point spread functions, Van Dokkum (2001) uses Laplacian edge detection, Pych (2004) performs an analysis of the histogram of the image data and Salzberg et al. (1995) use an artificial neural network. The more difficult cases are when some of the multiple-pixel cosmic ray hits are larger than some of the PSFs of true astronomical sources (Van Dokkum 2001). Thus, reasonably trained humans can usually perform this task with a considerable percentage of accuracy. In this paper we present an algorithm that aims to reject cosmic ray hits based on human perception, and implemented using Fuzzy Logic modeling (Zadeh 1965, 1988). In section 2 we discuss the human intuition of detecting cosmic ray hits in an astronomical exposure, in section 3 we describe the fuzzy logic model, in section 4 we describe the computation process, in section 5 we present how the fuzzy logic model is used, and in section 6 we discuss the performance of the algorithm.

Correspondence to: lshamir@mtu.edu

2. Manual detection of cosmic ray hits

Cosmic ray hits in astronomical exposures are usually noticeably different then PSFs of true astronomical sources, and a reasonably trained human can usually tell between the two. One examining an astronomical frame can notice that cosmic ray hits are usually smaller than PSFs of astronomical sources, and their edges are usually sharper. Although many cosmic ray hits are not larger than just one pixel, in some cases they can be larger than some of the point spread functions of astronomical sources (Van Dokkum 2001). An observer trying to manually detect cosmic ray hits in an astronomical frame would probably examine the edges and the surface size of the peaks. For instance, if the surface size of the peak is very small and it has sharp edges, it would be classified as a cosmic ray hit. If the surface size of the peak is larger and its edges are not very sharp, it would be probably classified as a PSF of an astronomical source. Since some of the cosmic ray hits have only one or two sharp edges, it is also necessary to examine the sharpest edge of the PSF.

This intuition can be summarized by a set of intuitive natural language rules such as:

1. If the peak is small and the edges are sharp then the peak is a cosmic ray hit.
2. If the peak is large and the edges are not very sharp then the peak is not a cosmic ray hit.
3. If the peak is medium and most of the edges are not very sharp except from one extremely sharp edge then the peak is a cosmic ray hit.
4. If the peak is small and the edges are moderately sharp then the peak is a cosmic ray hit.
5. If the peak is large and the edges are not sharp except from one edge that is moderately sharp then the peak is not a cosmic ray hit.

3. A human perception-based fuzzy logic model

One of the advantages of fuzzy logic is its ability to mathematically describe human intuition (Zadeh 1965, 1988, 1994). The first step in compiling the rules of intuition described above into a fuzzy logic model is to define the antecedent (input) and consequent (output) fuzzy variables. The antecedent variables in this model are the surface size of the peak (in pixels), the sharpness of the sharpest edge (in σ, where σ is the estimated noise) and the average sharpness of the edges (also in σ). The consequent variable is the classification of the peak. The domain of this variable is $\{$Yes/1,No/0$\}$, such that YES/1 means that the peak is classified as a cosmic ray hit and NO/0 means that the peak is classified as a PSF of a true astronomical source. Therefore, the model can be defined by the following function *f*:

f: *surface_size*, *sharpest_edge*, *average_edge* $\longmapsto \{0{,}1\}$

The surface size is determined by counting the pixels around the peak that are at least 3σ brighter than the local background. When a pixel less than 3σ above the local background is reached, the pixel is not counted and the edge sharpness is determined as the difference between the value of that pixel and the value of its neighboring pixel (in the direction of the peak).

The fuzzy sets defined for the surface size of the peak are described in Figure 1, and the membership functions of the fuzzy sets are defined by the following formulae:

$$F_{single}(x) = \begin{cases} 1 - \frac{x-1}{3} & 1 \leq x \leq 4 \\ 1 & 0 < x < 1 \\ 0 & x < 0 \text{ or } x > 4 \end{cases}$$

$$F_{small}(x) = \begin{cases} \frac{x-1}{3} & 1 \leq x < 4 \\ 1 - \frac{x-4}{3} & 4 \leq x \leq 7 \\ 0 & x < 1 \text{ or } x > 7 \end{cases}$$

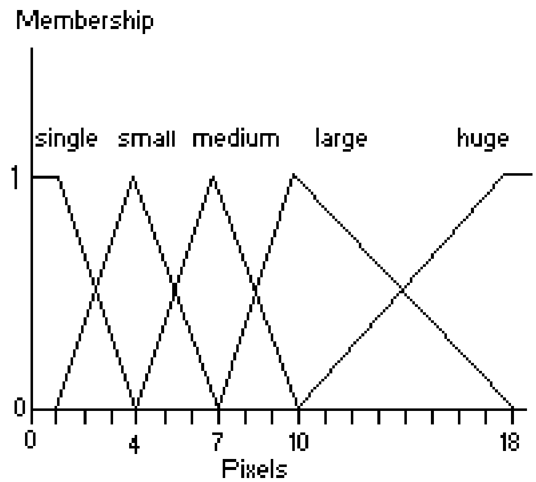

Fig. 1. The membership functions of the fuzzy sets defined for the surface size: single, small, medium, large, huge.

$$F_{medium}(x) = \begin{cases} \frac{x-4}{3} & 4 \leq x < 7 \\ 1 - \frac{x-7}{3} & 7 \leq x \leq 10 \\ 0 & x < 4 \text{ or } x > 10 \end{cases}$$

$$F_{large}(x) = \begin{cases} \frac{x-7}{3} & 7 \leq x < 10 \\ 1 - \frac{x-10}{8} & 10 \leq x \leq 18 \\ 0 & x < 7 \text{ or } x > 18 \end{cases}$$

$$F_{huge}(x) = \begin{cases} \frac{x-10}{8} & 10 \leq x < 18 \\ 1 & 18 \leq x \\ 0 & x < 10 \end{cases}$$

The antecedent variables *sharpest_edge* and *average_edge* use the same fuzzy sets. The fuzzy sets defined for these variables are *low*, *moderate*, *sharp* and *extreme*, as described in Figure 2.

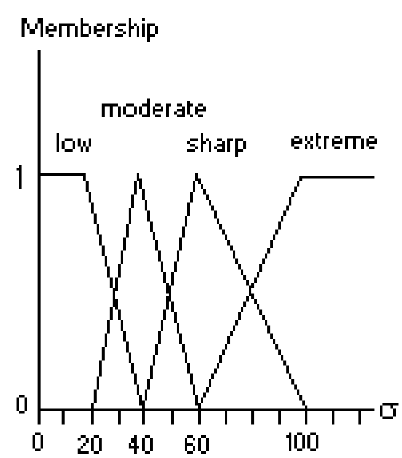

Fig. 2. The membership functions of the fuzzy sets defined for the edge sharpness

The membership functions of the fuzzy sets described in Figure 2 are defined by the following formulae:

$$F_{low}(x) = \begin{cases} 1 - \frac{x-20}{20} & 20 \leq x \leq 40 \\ 1 & 0 \leq x < 20 \\ 0 & x < 0 \text{ or } x > 40 \end{cases}$$

$$F_{moderate}(x) = \begin{cases} \frac{x-20}{20} & 20 \leq x < 40 \\ 1 - \frac{x-40}{20} & 40 \leq x \leq 60 \\ 0 & x < 20 \text{ or } x > 60 \end{cases}$$

$$F_{sharp}(x) = \begin{cases} \frac{x-40}{20} & 40 \leq x < 60 \\ 1 - \frac{x-60}{40} & 60 \leq x \leq 100 \\ 0 & x < 40 \text{ or } x > 100 \end{cases}$$

$$F_{extreme}(x) = \begin{cases} \frac{x-60}{40} & 60 \leq x < 100 \\ 1 & 100 \leq x \\ 0 & x < 60 \end{cases}$$

The fuzzy rules are defined using the membership functions of the antecedent variables and the domain of the consequent variable $\{0,1\}$, and are based on the natural language rules of intuition described in section 2.

single, low, low ↦ 0	single, low, moderate ↦ 1
single, low, sharp ↦ 1	single, low, extreme ↦ 1
single, moderate, low ↦ 1	single, moderate, moderate ↦ 1
single, moderate, sharp ↦ 1	single, moderate, extreme ↦ 1
single, sharp, low ↦ 1	single, sharp, moderate ↦ 1
single, sharp, sharp ↦ 1	single, sharp, extreme ↦ 1
single, extreme, low ↦ 1	single, extreme, moderate ↦ 1
single, extreme, sharp ↦ 1	single, extreme, extreme ↦ 1
small, low, low ↦ 0	small, low, moderate ↦ 0
small, low, sharp ↦ 1	small, low, extreme ↦ 1
small, moderate, low ↦ 1	small, moderate, moderate ↦ 1
small, moderate, sharp ↦ 1	small, moderate, extreme ↦ 1
small, sharp, low ↦ 1	small, sharp, moderate ↦ 1
small, sharp, sharp ↦ 1	small, sharp, extreme ↦ 1
small, extreme, low ↦ 1	small, extreme, moderate ↦ 1
small, extreme, sharp ↦ 1	small, extreme, extreme ↦ 1
medium, low, low ↦ 0	medium, low, moderate ↦ 0
medium, low, sharp ↦ 0	medium, low, extreme ↦ 0
medium, moderate, low ↦ 0	medium, moderate, moderate ↦ 0
medium, moderate, sharp ↦ 0	medium, moderate, extreme ↦ 1
medium, sharp, low ↦ 1	medium, sharp, moderate ↦ 1
medium, sharp, sharp ↦ 1	medium, sharp, extreme ↦ 1
medium, extreme, low ↦ 1	medium, extreme, moderate ↦ 1
medium, extreme, sharp ↦ 1	medium, extreme, extreme ↦ 1
large, low, low ↦ 0	large, low, moderate ↦ 0
large, low, sharp ↦ 0	large, low, extreme ↦ 0
large, moderate, low ↦ 0	large, moderate, moderate ↦ 0
large, moderate, sharp ↦ 0	large, moderate, extreme ↦ 0
large, sharp, low ↦ 0	large, sharp, moderate ↦ 0
large, sharp, sharp ↦ 1	large, sharp, extreme ↦ 1
large, extreme, low ↦ 0	large, extreme, moderate ↦ 1
large, extreme, sharp ↦ 1	large, extreme, extreme ↦ 1
huge, low, low ↦ 0	huge, low, moderate ↦ 0
huge, low, sharp ↦ 0	huge, low, extreme ↦ 0
huge, moderate, low ↦ 0	huge, moderate, moderate ↦ 0
huge, moderate, sharp ↦ 0	huge, moderate, extreme ↦ 0
huge, sharp, low ↦ 0	huge, sharp, moderate ↦ 0
huge, sharp, sharp ↦ 0	huge, sharp, extreme ↦ 0
huge, extreme, low ↦ 0	huge, extreme, moderate ↦ 0
huge, extreme, sharp ↦ 0	huge, extreme, extreme ↦ 0

4. The computation process

The computation process is based on *product* inferencing and *weighted average* defuzzification (Takagi & Sugeno 1983, 1985), and can be demonstrated by the following example: Suppose that *surface_size*=5, *average_edge*=35 and *sharpest_edge*=60. The membership of the value of the fuzzy variable *surface_size* in the fuzzy set *small* is $1 - \frac{5-4}{3} = \frac{2}{3}$ and the membership in the fuzzy set *medium* is $\frac{5-4}{3} = \frac{1}{3}$. The membership in the fuzzy sets *single*, *large* and *huge* is 0.

Similarly, the membership of the value of the fuzzy variable *average_edge* in the fuzzy set *moderate* is $\frac{35-20}{20} = \frac{3}{4}$ and in the fuzzy set *low* it is $1 - \frac{35-20}{20} = \frac{1}{4}$. 60 is the point where the membership function of the fuzzy set *sharp* reaches its maximum of unity, so the membership of 60 in *sharp* is $\frac{60-40}{20} = 1$, while its membership in all other fuzzy sets is 0.

In the inference computation stage, the first rule $single, low, low \mapsto 0$ is dependent on the fuzzy sets *single*, *low* and *low*, so the strength of this rule is $0 \cdot \frac{1}{4} \cdot 0 = 0$. When using *product* inferencing, if one of the fuzzy variables has a membership level of 0 (no membership) in one of the fuzzy sets referred by the rule, the rule does not have any effect on the final result of the computation. In this example, the only fuzzy rules that have non-zero membership values in all of 3 fuzzy sets are:

1. small, low, sharp ↦ 1
2. small, moderate, sharp ↦ 1
3. medium, low, sharp ↦ 0
4. medium, moderate, sharp ↦ 0

The membership of the values of the antecedent variables in the fuzzy sets of rule 1 (*small*, *low* and *sharp*) are $\frac{2}{3}$, $\frac{1}{4}$, 1 respectively. Similarly, the membership of the antecedent variables in the fuzzy sets of rule 2 are $\frac{2}{3}$,$\frac{3}{4}$,1, the membership in the fuzzy sets of rule *3* are $\frac{1}{3}$, $\frac{1}{4}$, 1, and the membership in the fuzzy sets of rule *4* are $\frac{1}{3}$, $\frac{3}{4}$, 1. Since the defuzzification is performed using the *weighted average* defuzzification method (Takagi & Sugeno 1983, 1985), the computed value of the consequent variable is:

$$\frac{\frac{2}{3}\cdot\frac{1}{4}\cdot 1\cdot 1+\frac{2}{3}\cdot\frac{3}{4}\cdot 1\cdot 1+\frac{1}{3}\cdot\frac{1}{4}\cdot 1\cdot 0+\frac{1}{3}\cdot\frac{3}{4}\cdot 1\cdot 0}{\frac{2}{3}\cdot\frac{1}{4}\cdot 1+\frac{2}{3}\cdot\frac{3}{4}\cdot 1+\frac{1}{3}\cdot\frac{1}{4}\cdot 1+\frac{1}{3}\cdot\frac{3}{4}\cdot 1} = \frac{2}{3}$$

The value of the consequent variable is handled such that values greater than 0.5 are classified as cosmic ray hits. Otherwise, the values are classified as non-cosmic ray hits. Since in the above computation the output value is $\frac{2}{3}$, the input values in this example are classified as a cosmic ray hit.

5. Using the fuzzy logic model

The fuzzy logic model is used in order to classify peaks in the frame as cosmic ray hits or non-cosmic ray hits. Each cosmic ray hit or PSF has one (or several) brightest pixels that can be considered as the center of the peak. In the presented algorithm, searching for peaks in a FITS frame is performed by comparing the value of each pixel with the values of its

8 neighboring pixels. If the pixel is equal or brighter than its 8 neighboring pixels, it is considered as a center of a peak. After finding the peaks in the frame, the fuzzy logic model is applied on the peaks in order to classify them as cosmic ray hits or non-cosmic ray hits.

If a background pixel happens to be brighter than its 8 neighboring pixels, it will be mistakenly considered as a center of a peak. The probability of that event is $0.5^8 = 0.00390625$. For instance, in an astronomical frame of 1024×1024, 4096 background pixels are expected to be mistakenly considered as peaks. However, since the computation process is relatively fast, these additional peaks do not significantly slow down the algorithm.

6. Performance of the algorithm

Measurements of the performance of the algorithm were taken using 24 *Night Sky Live* (Perez-Ramirez, Nemiroff & Rafert 2004) all-sky exposures. Each NSL frame contains an average of 6 noticeable cosmic ray hits brighter than 20σ, and around 1400 astronomical sources brighter than 20σ over their local background. Out of 158 cosmic ray hits that were tested, the algorithm did not reject 4, and mistakenly rejected 6 true astronomical sources out of a total of 31,251 PSFs. These numbers do not present better accuracy than other reported cosmic ray rejection algorithms such as (Rhoads 2000; Van Dokkum 2001; Pych 2004), but the proposed algorithm has a clear advantage in terms of computational efficiency. While some of the above algorithms are relatively slow, the presented algorithm can process a 1024×1024 integer FITS frame in less than 4 seconds, using a system equipped with an Intel Pentium IV 2.66 MHZ processor and 512 MB of RAM. This advantage can be significant in systems such as the *Night Sky Live* all-sky survey, in which each node of the network takes an exposure every 3 minutes and 56 seconds.

7. Conclusions

In this paper we presented an algorithm for cosmic ray hits rejection that is based on fuzzy logic modeling of human intuition. Experiments show that the algorithm rejects ∼97.5% of the cosmic ray hits, but also rejects 0.02% of the PSFs of true astronomical sources. The algorithm is implemented and used regularly by the Night Sky Live all-sky survey, in which the system automatically searches for transients and monitors known bright astronomical sources.

References

Axelrod, T., Connolly, A., Ivezic, Z., Kantor, J., Lupton, R., Plante, R., Stubbs, C., Wittman, D.: 2004, AAS Meeting 205, 108.11

Becker, A.C., Rest, A., Miknaitis, G., Smith, R.C., Stubbs, C.: 2004, AAS Meeting 205, 108.12

Fixsen, D.J., Offenberg, J.D., Hanisch, R.J., Mather, J.C., Nieto-Santisteban, M.A., Sengupta, R., Stockman, H.S.: 2000, PASP 112, 1350.

Offenberg, J.D., Sengupta, R., Fixsen, D.J., Stockman, H.S., Nieto-Santisteban, M.A., Stallcup, S., Hanisch, R.J., Mather, J.C.: 1999, in ASP Conf. 172, 141

Otuairisg, S., Golden, A., Butler, R.F., Shearer, A., Voisin, B.: 2004, Ground-based Telescopes 5493, 467

Perez-Ramirez, D., Nemiroff, R.J., Rafert, J.B.: 2004, AN 325, 568

Pych, W., 2004, PASP 116, 148

Rhoads, J.E.: 2000, PASP 112, 703

Salzberg, S., Chandar, R., Ford, H., Murphy, S.K., White, R.: 1995, PASP 107, 279

Shaw, R.A., Horne, K.: 1992, ASP Conf. 25, Astronomical Data Analysis, Software and Systems

Takagi, T., Sugeno, M.: 1983, Proc. of the IFAC Symp. on Fuzzy Information, Knowledge Representation and Decision analysis, 55

Takagi, T., Sugeno, M.: 1985, IEEE Trans. Syst. Man & Cybern. 20, 116

Van Dokkum, P.G.: 2001, PASP 113, 1420

Windhorst, R.A., Franklin, B.E., Neuschaefer, L.W.: 1994, PASP 106, 798

Zadeh, L.A.: 1965, Information and Control 8, 338

Zadeh, L.A.: 1988, Fuzzy Logic, Computer 21, 83

Zadeh, L.A.: 1994, IEEE Software 11, 48