

# Study of the impact of pre-processing applied to images acquired by the Cygno Experiment<sup>\*</sup>

I. Abrita Costa<sup>1,6</sup>, E. Baracchini<sup>2</sup>, F. Bellini<sup>1,3</sup>, L. Benussi<sup>4</sup>, S. Bianco<sup>4</sup>,  
G. Cavoto<sup>1,3</sup>, E. Di Marco<sup>1</sup>, G. Maccarrone<sup>4</sup>, M. Marafini<sup>5</sup>, G. Mazzitelli<sup>4</sup>,  
A. Messina<sup>1,3</sup>, D. Piccolo<sup>4</sup>, D. Pinci<sup>1</sup>, F. Renga<sup>1</sup>, F. Rosatelli<sup>4</sup>, and  
S. Tomassini<sup>4</sup>

<sup>1</sup> Istituto Nazionale di Fisica Nucleare  
Sezione di Roma, I-00185, Italy

<sup>2</sup> Gran Sasso Science Institute L'Aquila, I-67100, Italy

<sup>3</sup> Dipartimento di Fisica  
Sapienza Università di Roma, I-00185, Italy

<sup>4</sup> Istituto Nazionale di Fisica Nucleare  
Laboratori Nazionali di Frascati, I-00040, Italy

<sup>5</sup> Museo Storico della Fisica e Centro Studi e Ricerche "Enrico Fermi"  
Piazza del Viminale 1, Roma, I-00184, Italy

<sup>6</sup> Federal University of Juiz de Fora  
Electrical Engineering Department, Juiz de Fora, Brazil  
`davide.pinci@roma1.infn.it`

**Abstract. Keywords:** Image Analysis · DBSCAN · Particle physics experiment.

## 1 Introduction

Digital image processing techniques have emerged in the at the end of the 1960s, being applied in fields such as medical science, observation of earth resources and astronomy. Since then, its field of application has grown considerably and can now be applied to the experimental study of physics of the particles[ref1]

Currently, gas detectors have proven to be a choice for detecting particles with low emission of energy. In these detectors, the light produced by the de-excitation of the gas molecules during the multiplication process of electrons can be captured by a camera based on CMOS technology[ref2]. These images can be processed for better pictorial information and for a more effective human interpretation. In addition, additional information can be extracted and processed by classifiers for applications in pattern recognition and machine learning.

For improving the detection and classification of patterns in images, it's necessary a good pre-processing, mainly on images that have a low signal-noise ratio where it's need to choose carefully the algorithms that will be used to improve

---

<sup>\*</sup> This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001".

this relation. To eval the best techniques, a truth dataset it's very important to guide and understand more about the problem and the algorithms used.[citar]

This works proposes a tool for simulating the images of the experiment CYGNO, that aims to find evidences of dark matter[?], with the intention to improve the tracks detection with a low background detection. For a first approach It will test spatial filters.

The present paper is organized as follows: Section 2 describes the environment where the work takes place. The proposed algorithm is explained in Section 3, Section 4 shows the relevant results and discusses the applicability of the method, and Section 5 concludes this work.

## 2 Methodology

In this section, we present a overview of the environment where this work is done, show the used data set and yours characteristics. In addition, we expose the problematic and the steps of our analysis.

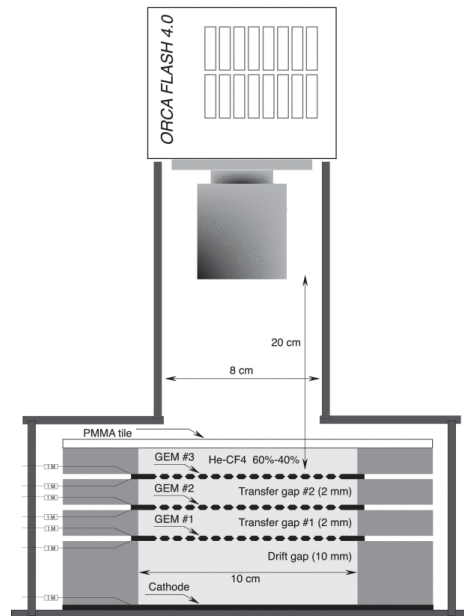
### 2.1 Cygno

The Cygno experiment aims to developed a triple-GEM structure read-out by a CMOS based sensor with high-resolution tracking for low energy particles to search for Dark Matter (DM) massive particles. The detector is developed to read the light produced by the de-excitation of gas molecules during the processes of electron multiplication.

In this work we are using the data acquired by the ORANGE (Optically ReAdout GEm) prototype, which is described in details in [?,?,?]. A  $10 \times 10 \text{cm}^2$  Triple GEM structure, with a  $1 \text{cm}$  high drift gap, using a binary gas mixture  $\text{He}/\text{CF}_4$  (60/40), was readout by an Orca Flash 4 CMOS-based camera<sup>7</sup> equipped with a large aperture ( $f/0.95$ ) lens, as shown in the Figure 1.

---

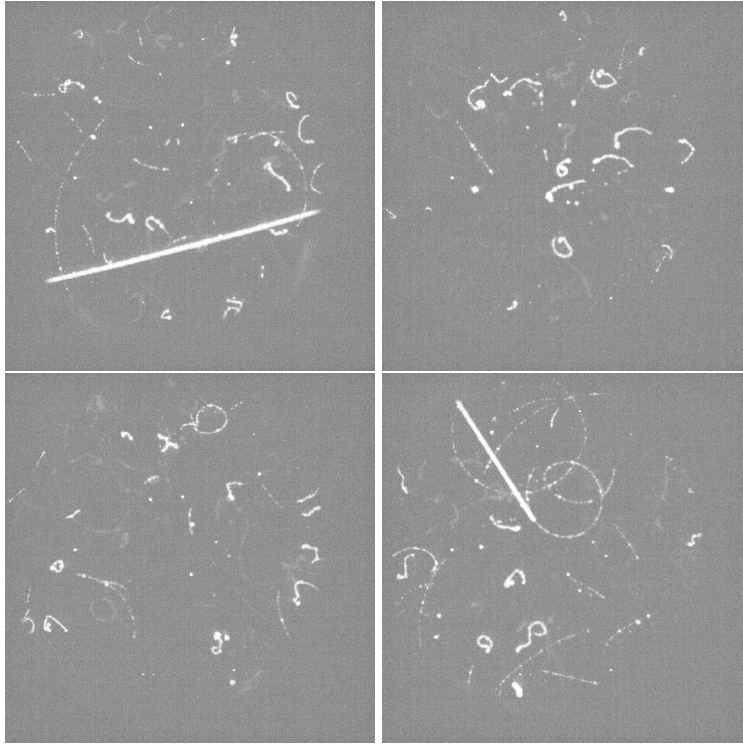
<sup>7</sup> For more details visit the site [www.hamamatsu.com](http://www.hamamatsu.com)



**Fig. 1.** Drawing (not to scale) of the triple-GEM stack with the lens and the CMOS camera. The drift and transfer gaps are shown.

## 2.2 Data set

The simulation is based on dataset generated by ORANGE where were acquired in free running mode, without any trigger and a  $AmBe$  neutron source was placed near to detector. Images produced by the GEM were recorded with an exposure of 2s and all data were saved without any selection. In figures ref It can see the output images generated by detector.



**Fig. 2.** Images acquired by the ORANGE prototype.

How it can be observed, the tracks to be identified have different shapes and different brights. Possibly only one step in the pre-processing is not enough, mainly because of the tracks bright. Then to prove it, we developed a simulation environment where it's possible to generate tracks, put them on images, add noise and change the tracks bright, being able to know what are the best algorithms for each bright region for a certain noise level. To be close to the experiment, It was extracted the noise parameters of the images and eval these algorithms on this noise level. The same was done for tracks, getting parameters like width and length of tracks.

### 2.3 Simulation Environment

The proposed simulation environment is showed in the figure ref. This simulation can be divided in 4 steps:

- Track Generation

This stage defines the parameters of the track and the bounds of this track in the image.



- Image Generation  
 On this step the tracks generated are add into image together with the background and the noise extracted from the CYGNUS data set. How the pixels where there are tracks are known, It's possible to control their intensity about the background. This operation is called contrast reduce.
- Image preprocessing  
 After to generate the image with tracks the next step is apply techniques to improve the signal noise ratio for any track(high contrast or no). Then the technique that was choosen for do this is filters. In this work we used some the most used spatial filters(mean filter, median filter and gaussian filter). Behind to filter the images, the filtered images are binarized to separated the tracks from the background.
- Evaluation  
 The last one step is to compare the pixels of the tracks detected after threshold operation with the truth pixels that have been generated and because of this they are known. The metrics used to do this is simply get the efficiency of signal detection and background rejection.

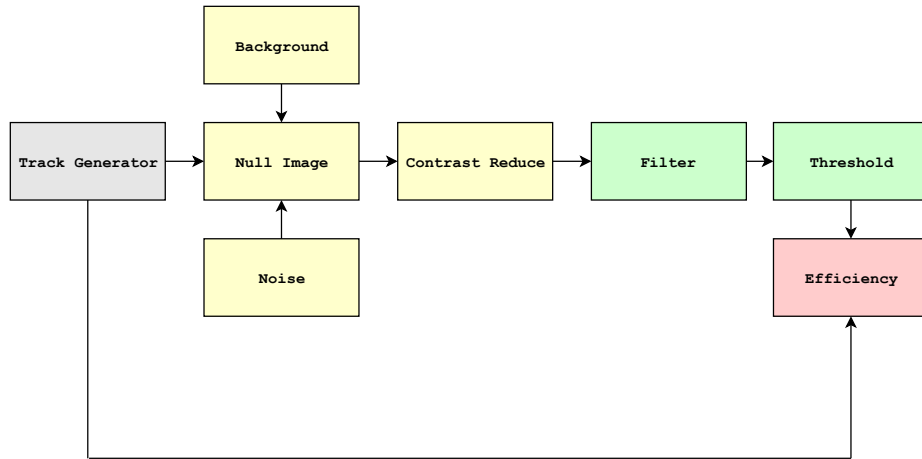


Fig. 3. Simulation environment diagram

### 3 Development

In this section, it will be presented the steps taken to obtain the simulation model for the images of CYGNUS experiment, it will show the theory about the filters used and the evaluation method. This steps are based on the blocks already described in Section 2. It will be presented in three subsections: the first one describes the procedure used to generate the tracks; the second describes the

generation of image and noise, together with the contrast reduction operation. The third describe the filters that were used on this work and last one shows the evaluation method.

### 3.1 Track Generation

How can it be observed in Figure 2 there are many possibilities of track's shape. The tracks also can be present some holes in their trajectory. This holes happen because the system detection and it needs to be studied at some moment, this effect has been called of degradation.

For a first approach, It has been characterized only tracks that their shape is straight and for now, the degradation was disregarded. To became the generation near to reality, It was used the information about the track's length and width, they were the parameters used as input in the generator.

The pdfs(probability density functions) of these measures are showed in Figure 4. When it looking into the jointly pdf it's possible to realize the correlation between measures is low, so these variables can be generated singly. It has been used the KDE(Kernel Density Estimator)[referencia] to estimate the pdfs of the data and generate them. The KDE estimator is defined as 1, where  $K$  is the kind of kernel used and  $h$  is the bandwidth kernel. The gaussian kernel has been choosen and the bandwidth value can be determined using the equation 2, where  $s$  is the lenght of data and  $A$  is the minimum between the interquartile divided by 1.34 and standard deviation of data[referencia].

$$f_h(\hat{x}) = \frac{1}{nh} \sum_{i=1}^n K_h\left(\frac{x - x_i}{h}\right) \quad (1)$$

$$h = 0.79 * A * s^{-0.2} \quad (2)$$

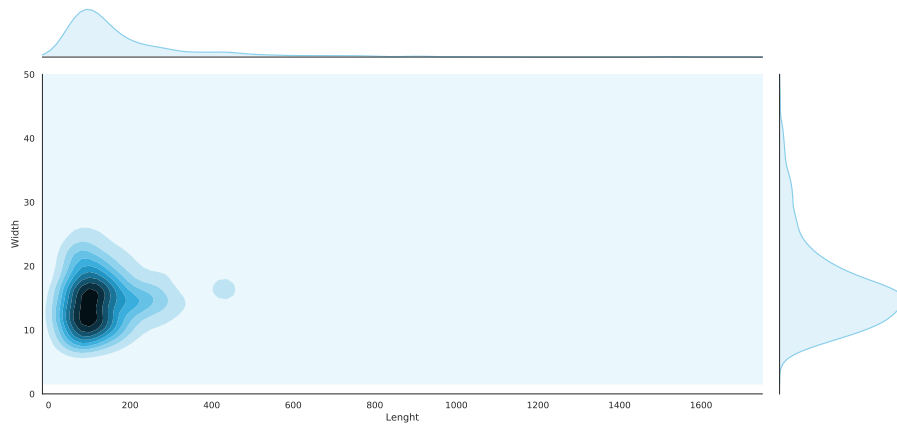


Fig. 4. Simulation environment diagram

The next step is to generate the tracks from the length and width extracted by estimated pdfs. The track generation is divided in 3 parts. The first is to divide a straight segment with the length obtained from pdf's in  $n$  points at the Cartesian plane, how is showed in Figure 5.

The second part is to rotate this points by a random angle and centralize it in a random position. This is necessary because of the tracks have a random orientation and center in images. How we don't have the pdfs of this parameters, for now, we assume that they are an uniform random variables. Then, after to sort a angle from 0 to  $2\pi$ , we apply on the points the rotation matrix that is showed in equation 3. The Figure 6 shows the rotation and translation of the straight segment of the Figure 5

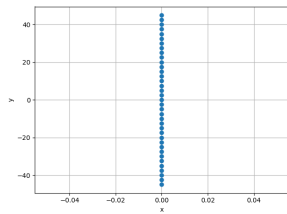
$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & x_0 \\ \sin(\theta) & \cos(\theta) & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (3)$$

The third part is to join these points to get a straight but this must be done in a way that resembles the tracks found in reality. It could be seen the straight rounded formats around you and to reproduce this behavior it was used a Gaussian radial base function  $K(\mathbf{x}, \mathbf{x}')$  in the equation 4, for each point at the straight.

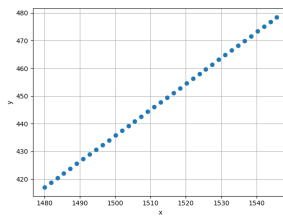
$$K(\mathbf{x}, \mathbf{x}') = e^{-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\sigma^2}} \quad (4)$$

For this work, the vector  $\mathbf{X}$  is the Cartesian plane  $(x, y)$  and the vector  $\mathbf{X}'$  is the coordinates of the  $n$  over the straight. The output of function  $K$  are values from 0(far from  $\mathbf{X}'$ ) to 1(near to  $\mathbf{X}'$ ). Each point has one function  $K$ , so the output is the sum off all functions as can be showed in equation 5. How the Gaussian function is infinite a threshold it's necessary to define the region of tracks. This threshold was set as the Gaussian output when the input is  $4\sigma$ , the value is 0.000335.

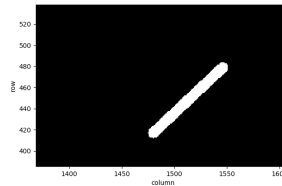
$$\sum_{i=1}^n K_n(\mathbf{x}, \mathbf{x}'_n) = e^{-\frac{\|\mathbf{x}-\mathbf{x}'_n\|^2}{2\sigma^2}} \quad (5)$$



**Fig. 5.** Straight segment divide



**Fig. 6.** Rotation of points

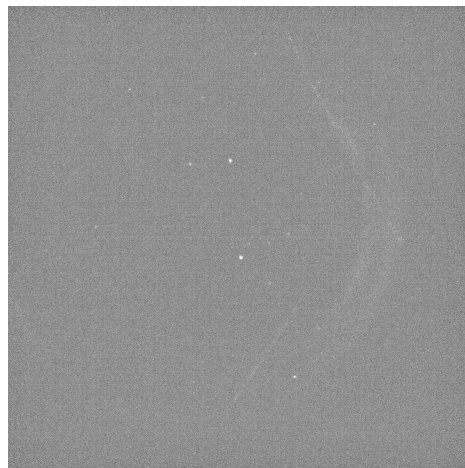


**Fig. 7.** Track generated

The last one part is to grid the function of equation 5 in the image dimensions. The step of grid is 1 for each dimension( $x$  and  $y$ ). The result is showed in Figure 7.

### 3.2 Image Generation

How can be observed in Figure 2, the output images of CYGNUS experiment beyond have tracks, they also have a background and noise inherent in its capture process. Then, to reproduce this effect, It was need to characterize and get their parameters. For do it, It was used a dataset without tracks, an example of image from this data set is showed in Figure 11.



**Fig. 8.** Cygnus image without tracks

### 3.3 Filters theory

### 3.4 Evaluation method

falar que foi um track por imagem, dividiu o L em 3 partes..

## 4 Results

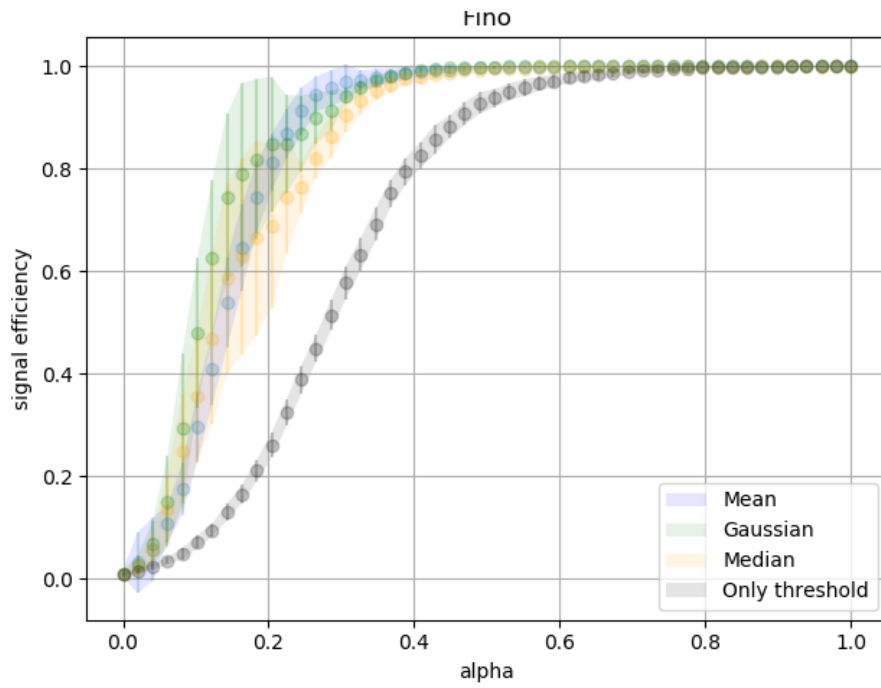


Fig. 9. Fino

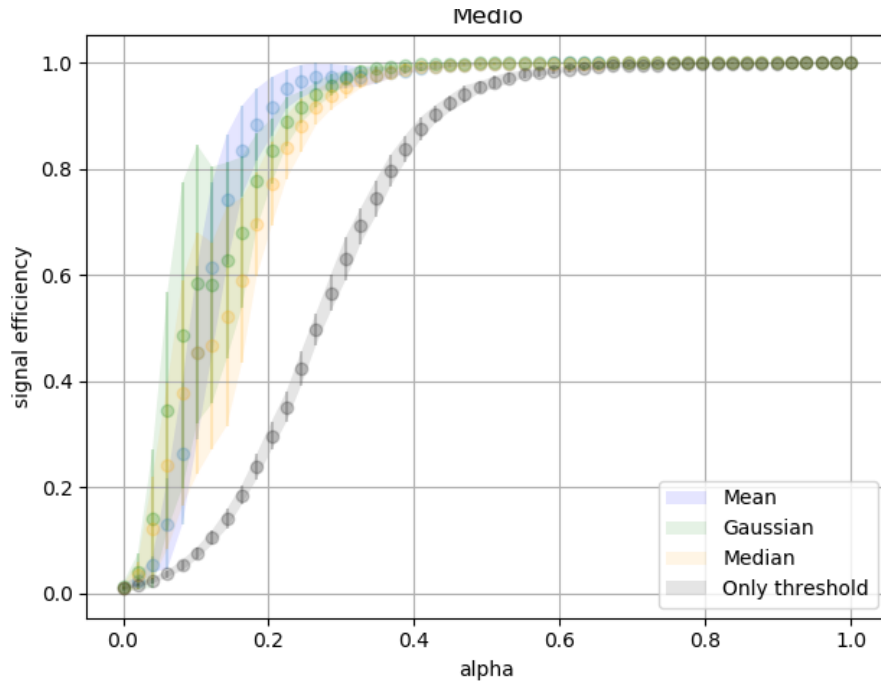


Fig. 10. Medio

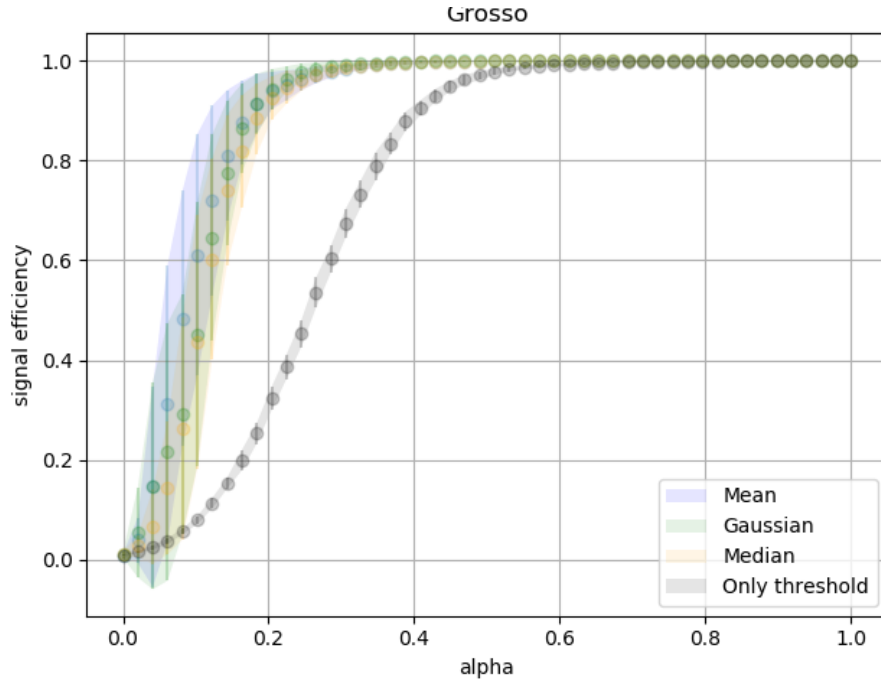


Fig. 11. Grosso

## 5 Conclusions

## Acknowledgment