Integration and development of directional clustering for high pile-up

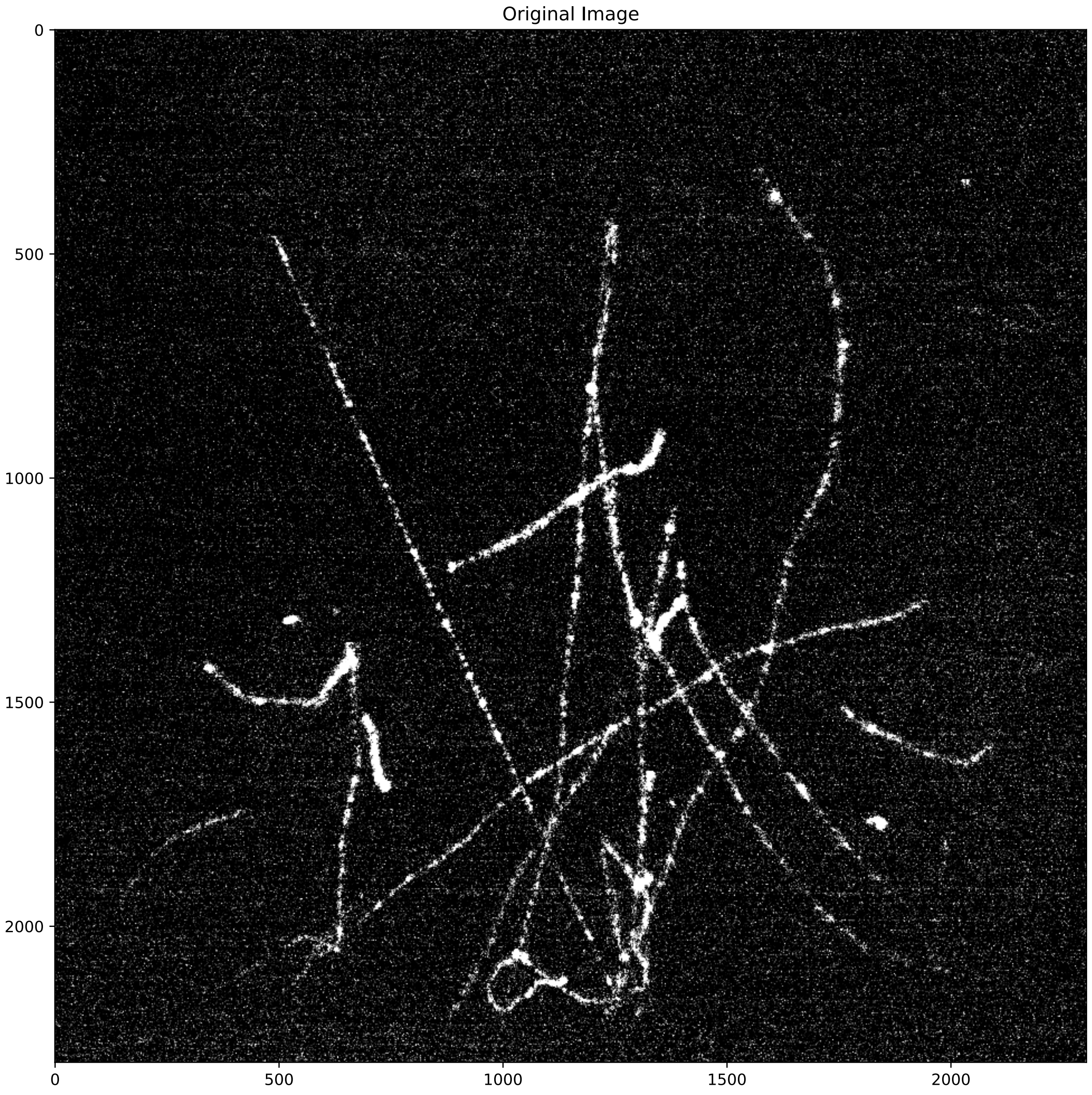
Cavoto, Di Marco, Pinci

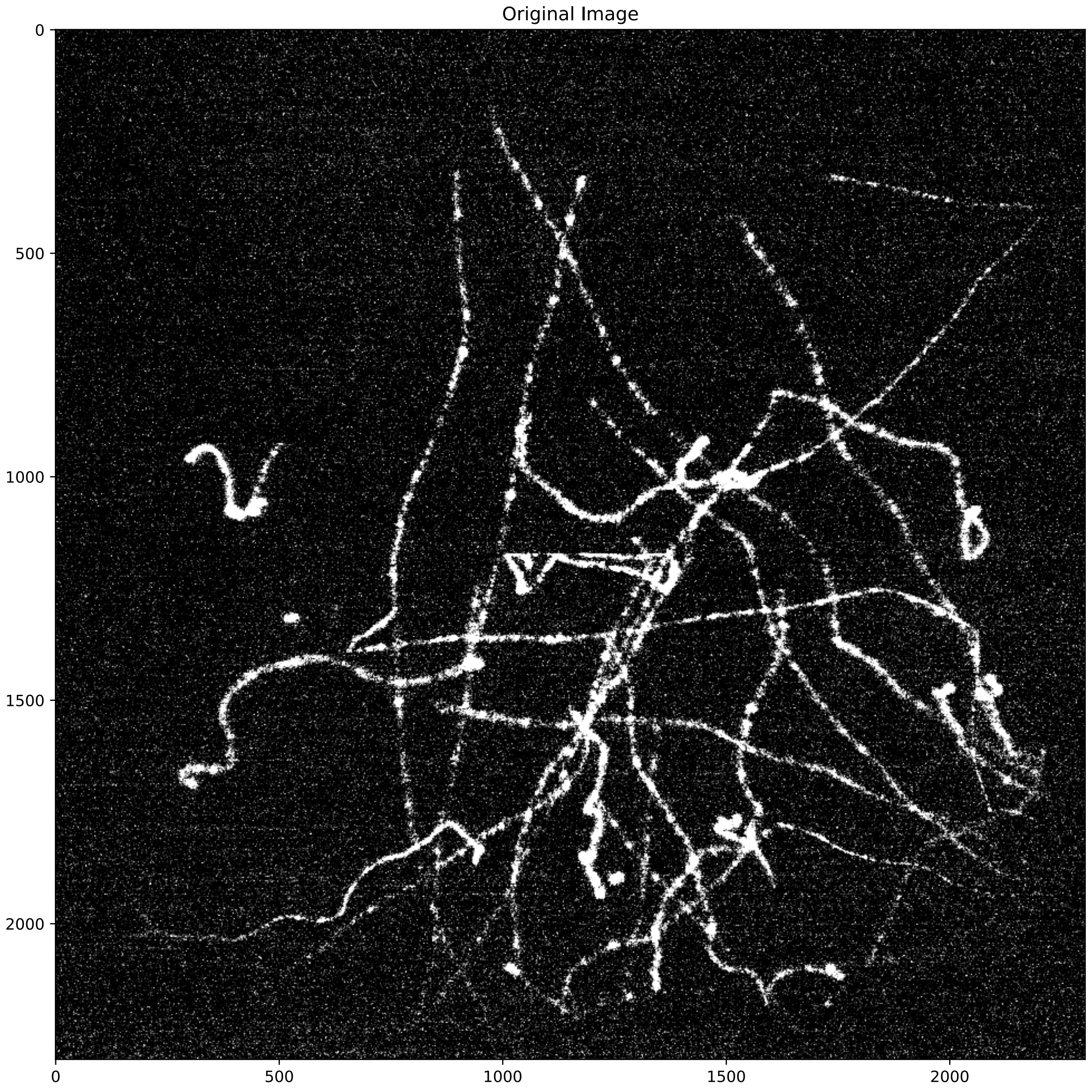
CYGNO reco and analysis meeting,
18 March 2021

Cs-137 in LIME

Data taken with LIME and exposure = 500 ms have high pileup of tracks in data taken with Cs-137

Typical (not extreme) pictures:





High pileup pictures

Maybe it is impossible / useless to disentangle these balls of wool, but it can be an extreme case to test different reconstruction algorithms

Up to now (AmBe paper = "lime_2020" GIT branch):

- iterative - DBDSCAN used to seed high interest regions
- GAC used to make superclusters looking at high interest regions
- Works well (apart start of tracks for directionality studies), but cannot cope with overlapping tracks: they are merged

Refresh: why the need to cluster together a long track if the expected signals from NRs/ERs are short?

BECAUSE unclustered pieces of long bkg tracks mimick signal!

Directional tracking

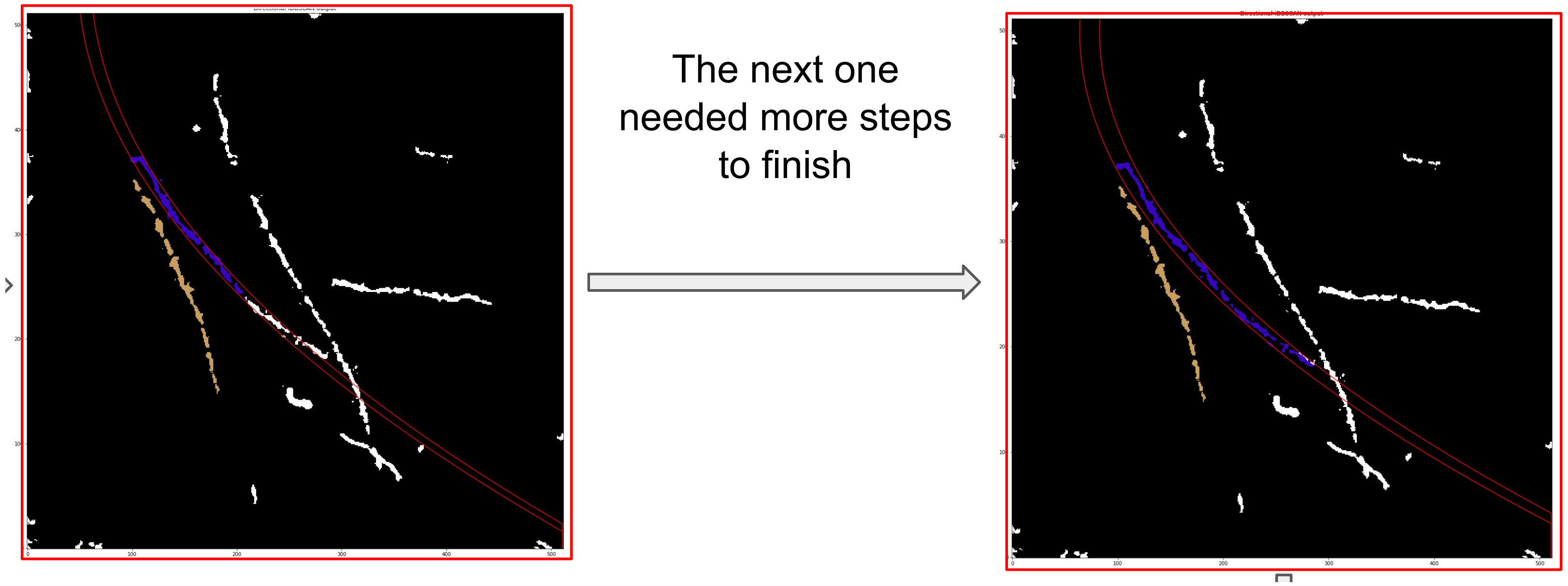
For this I. Pains has developed a clustering that search for patterns compatible with polynomials (line or 3rd order polynomial). Links to presentations here and here.

Reminder of the method:

- starts with DBSCAN with a short radius

- tests if starting from these clusters, one can find other clustered points compatible with a polynomial

    - the polynomial is fitted iteratively until points are added to the supercluster



The next one needed more steps to finish

I. Pains

Limitations in high PU

- As soon as the occupancy increases, everything gets merged when the "seed" cluster is in a crowded region

- it is slow, because of the many fits/seed done

- 3-rd order polynomial sometimes not sufficient, but fitting with higher order can get crazy soon

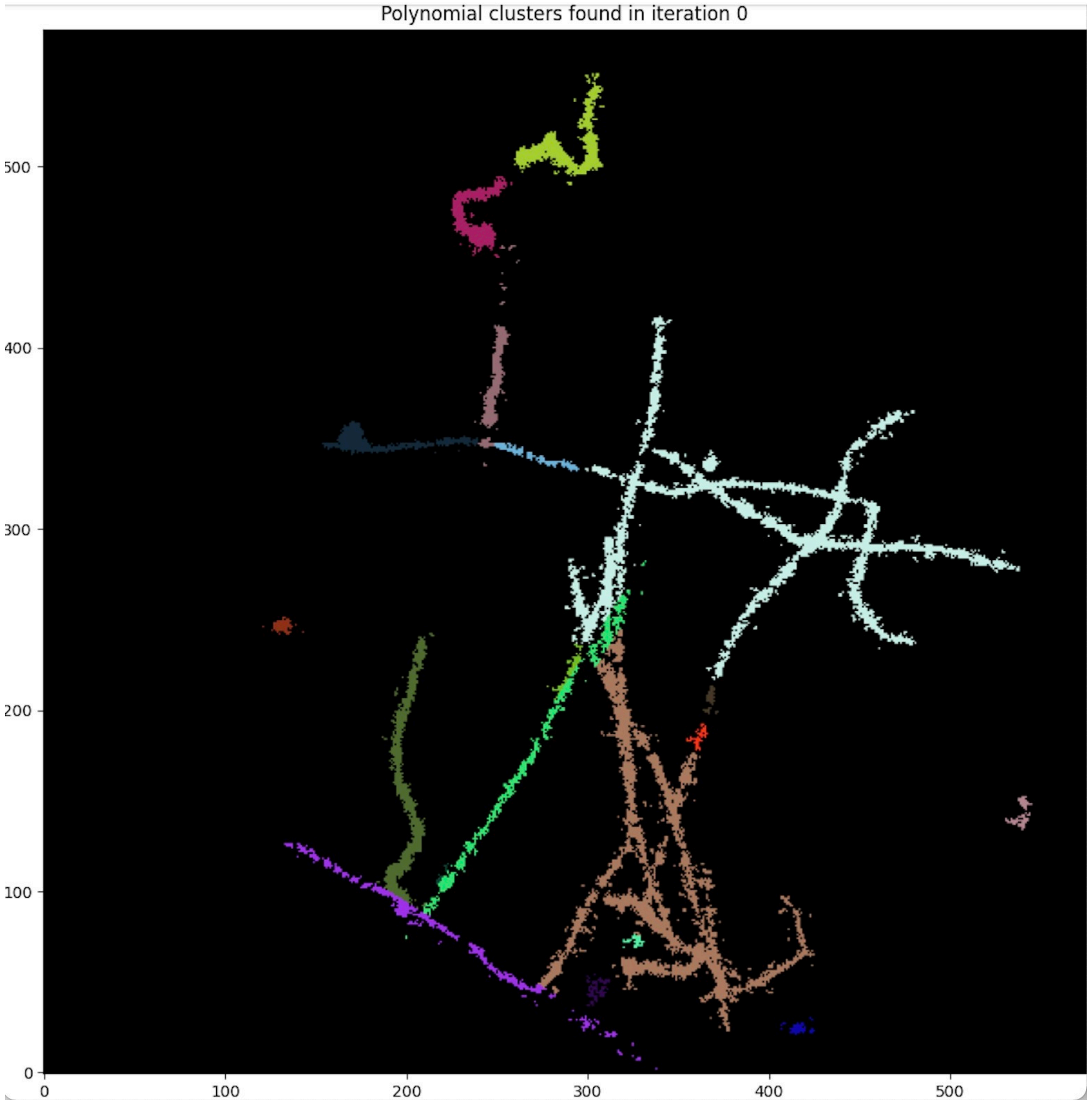
ATTEMPTs to improve:

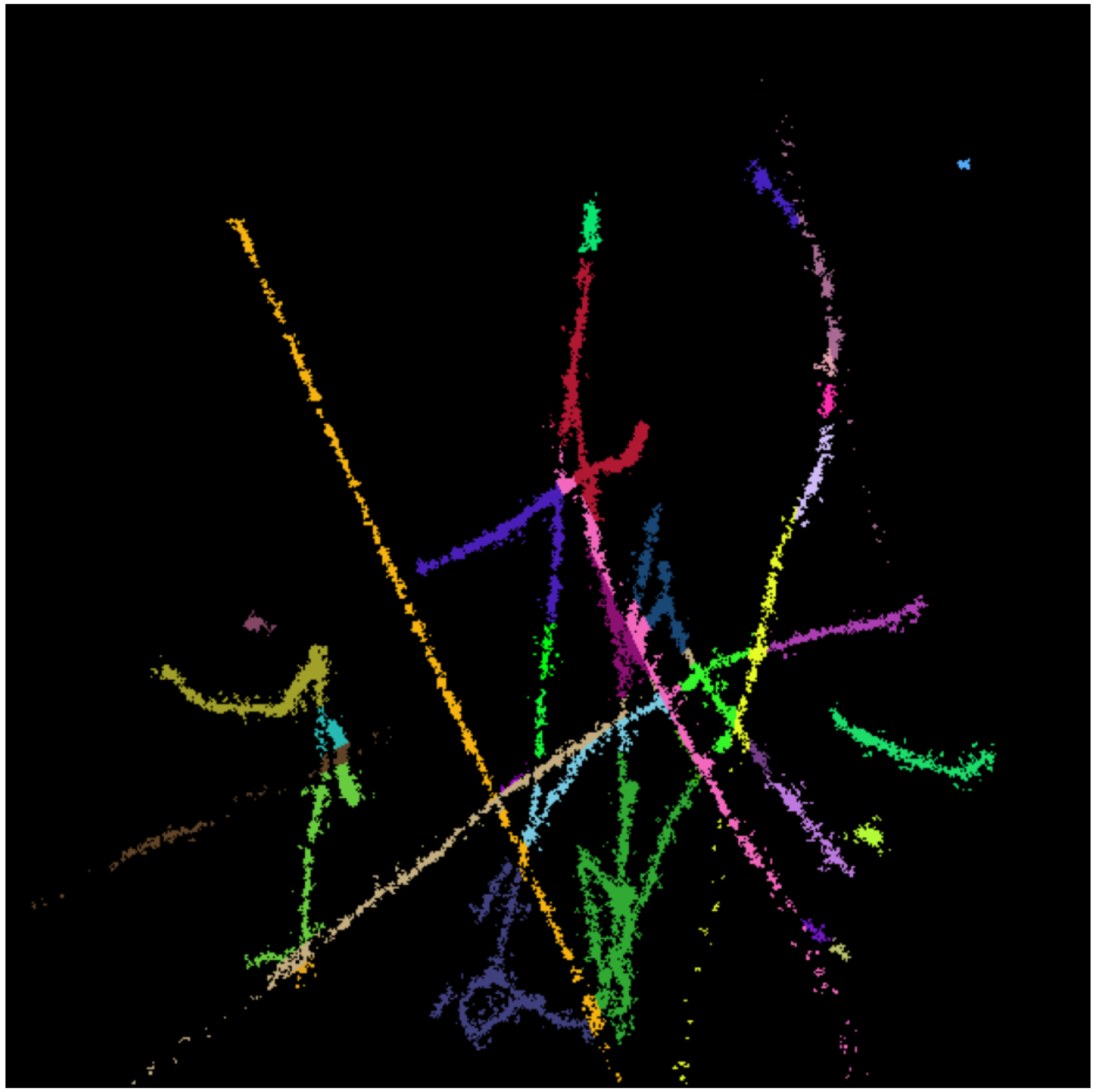
1. Use "isolated" seeds to start directional search, i.e. with the miniminum $I = \sum_{i}^{\Delta R = 200} A_i$. ($A_i$ = i-pixel amplitude). If >1 has I=0, then sort by the best linear fit X^2.

2. Use Bernstein polynomials to approximate the curve, to improve stability

Examples (Cs-137)

N.B. This is after a lot of tuning of parameters (isolation definition, clustering metric, etc.)

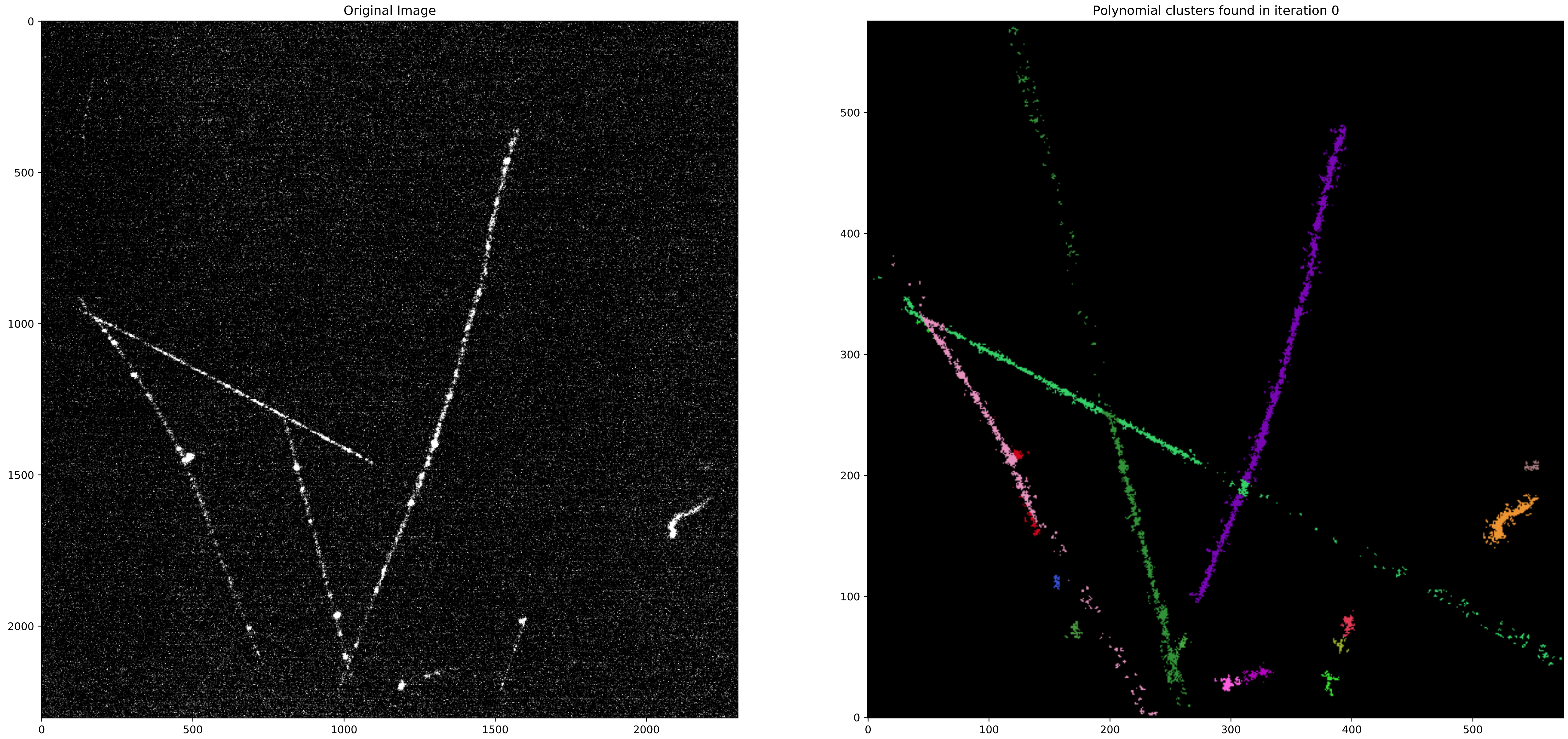




The eagerness of the directional is on purpose exaggerated because it is better to eat some piece of another track that leave a disjoint piece around (signal fake!)

And AmBe?

Luckily, data taken with Cs-137 is a bit extreme. E.g. AmBe seems much quieter



N.B. The situation is worsened by the light entering in the top/bottom regions of the detector

Practicalities

This is slow. Time dominated by fitting many times track candidates: 1 event can take 3/4 mins, depending by the occupancy.

Possible speed-up: CYTHON (compiled C from python version) or Numba (machine code from python version) tested

=> can speed up by ~20%, not more (because most is using numpy, which is C++ compiled behind the scenes)

Makes the development slow, but maybe we can survive using batch at LNGS (not many events to process)

Clusters removed from the ntuples, everything done on superclusters (cluster shapes, etc) are done exactly as before.

GIT Branch: "lime21" (bleeding edge, not stable right now)

Applications

1. basic check: on SIM ERs, NRs, look at the containment, resolution, and compare to standard

2. on bkg-only data in LIME, check that the energy of cosmics is the one expected, after Fe calibration

3. look at the number of "short" clusters compatible with a NR in bkg-only data with this or the standard supercluster

Running on Summer data taken with LIME to check these points right now. Results not ready for today…