# Machine assisted histogram classification

Paper ID: 193

**B. Benyó [1], C. Gaspar [2], P. Somogyi [1,2]**

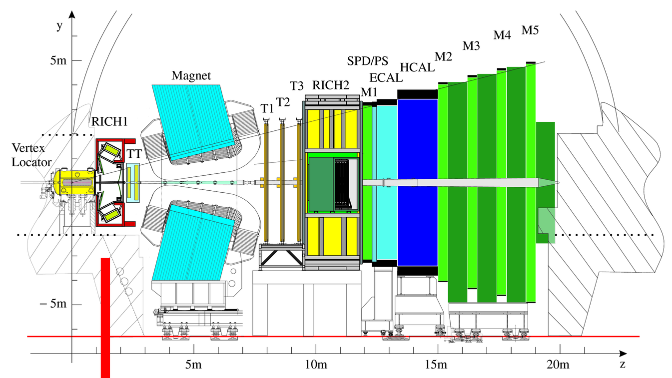
[1] Budapest University of Technology and Economics, [2] European Organization for Nuclear Research

## Introduction

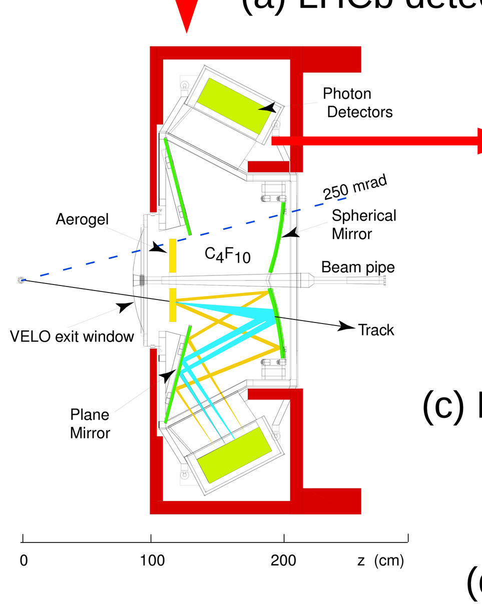
LHCb is one of the four major experiments under completion at the Large Hadron Collider (LHC). Monitoring the quality of the acquired data is important, because it allows the verification of the detector performance. Anomalies, such as missing values or unexpected distributions can be indicators of a malfunctioning detector, resulting in poor data quality.

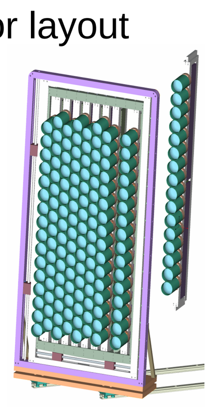
Spotting faulty or aging components can be either done visually using instruments such as the LHCb Histogram Presenter, or by automated tools. In order to assist detector experts in handling the vast monitoring information resulting from the sheer size of the detector, a graph-theoretic based clustering tool, combined with machine learning algorithms is proposed and demonstrated by processing histograms representing 2D event hitmaps. The concept is proven by detecting ion feedback events in the LHCb experiment's RICH subdetector.

## Visual monitoring of data quality histograms produced by subdetectors

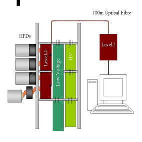


(a) LHCb detector layout





(c) RICH plane of HPD-s



(d) HPD connections

(b) RICH subdetector schematic

Figure 1: LHCb detector, RICH subdetector design

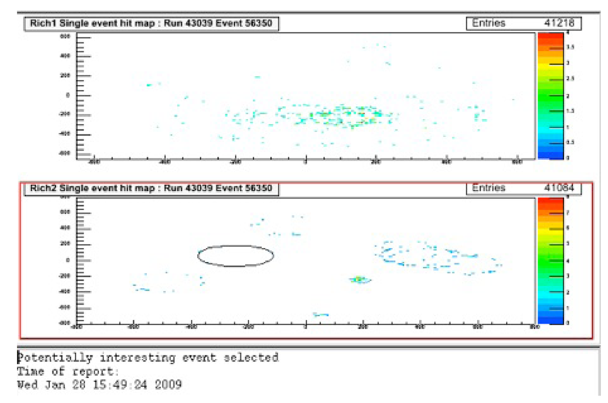
The LHCb detector is placed in one of the interaction points of the LHC, and its layout is shown in Figure 1(a). A detailed description of the detector can be found in [1,2]. In addition to trackers and to other particle identifiers, an important type of particle identifying subdetector within LHCb is the Ring Imaging Čerenkov* detector (RICH). The overall combination of radiators employed in RICH1 and RICH2 allow for the measurement in the 2-100 GeV momentum range, lending precision to the whole experiment by helping with a good pion-kaon separation capability, needed to distinguish similar B decay modes.

Figure 1(b) shows the layout of one of the RICH detectors. As particles pass through the radiators, Čerenkov light is emitted, which is directed out of the detector's acceptance area using mirrors, to allow the photons to reach the Hybrid Photon Detectors (HPD-s). The silicon sensor arrays of the HPD-s give a readout image of pixels spreading across the whole detection plane illustrated in Figure1(c), allowing for the reconstruction of rings as shown in Figure 2(a). The radius of each reconstructed rings is a measure for the Čerenkov angle found in Figure 1(b) used to identify the given particle.

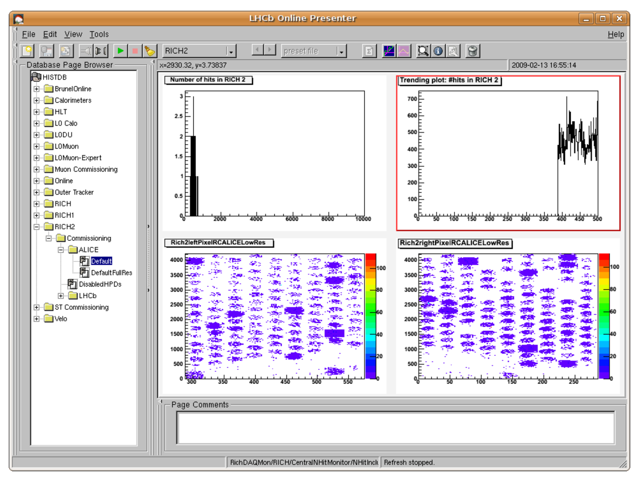
An example of online visual monitoring during data acquisition can be seen in Figure 2(b), showing plots of the number of hits and their trend in RICH2. The two coloured hitmaps below show the occupancy of the individual HPD tubes in the two detection planes. Should two HPD-s under each other in a column go totally blank, the subdetector expert would suspect that a Level-0 interface board is in failure, due to design of the connections shown in Figure 1(d).

Checking for known or anticipated anomalies can be automated in some cases. Machine assisted detection of events can lighten the burden of monitoring for the operators, when the behaviour of a large number of HPD-s need to be followed.

* When a charged particle passes through a medium, and its speed exceeds the speed of light in that given insulator medium, then Čerenkov radiation is produced. By measuring the angle between the Čerenkov photons and the track of the particle as shown in Figure 1(b), the velocity of the particle can be estimated when knowing the refractive index of the radiator medium, which the particle traverses. Large refractive index is needed to identify particles having a low momentum, while a small refractive index necessary for taking measurements in the high momentum range.



(a) Čerenkov light rings reconstructed using HPD readouts



(b) Presenter monitoring RICH HPD-s

Figure 2: Presentation software

## Machine assisted detection of ion feedback events

### Problem statement

The Hybrid Photon Detectors (HPD-s) shown in Figure 4(a) are electro-optical image intensifiers: a combination of a photomultiplier vacuum tube with focusing electrodes, and a multi pixel silicon sensor, providing excellent resolution for both space and energy.
The goal is to create a general purpose tool, which can recognize strange or anomalous histograms on the basis of examples shown and labelled by the subdetector expert, as for example in the case of an histogram shown in Figure4(b), which illustrates an ion feedback event. Photoelectrons ionize residual gas and result in noisy readout behaviour, peaking around the central axis of the tube. The frequency of occurrence of the ion feedback events is a measure of ageing of an HPD vacuum tube, therefore needs monitoring.

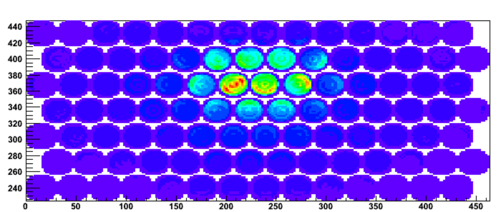
### Approach to the solution

As applied to ion feedback events:

- Read sensor data for each panel (Figure 3.)
- Segment RICH panel into individual HPD pixel readouts (Figure 4.)
- Extract generic features to obtain a vector as a signature of the given event, by creating a graph of the minimal spanning tree, covering every pixel hit in the HPD readout (Figure 5(a))
- Find natural clusters based on cuts on the feature vector elements (Figure 5(b))
- Label cluster categories as event types during assisted learning
- Create a model which separates the clusters according to the labels based on the feature vectors (Figure 6.)
- Automatic event recognition for new hitmaps, based on the output of the model and create statistics on ion feedback events for each HPD vacuum tube (Figure 7.)

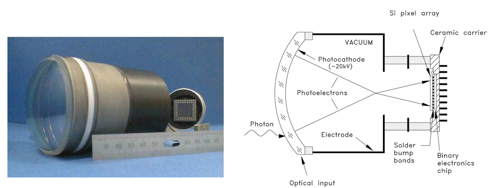
Sensor → Segmentation → Feature extraction → Classification → Post processing

(a) RICH Panel photo

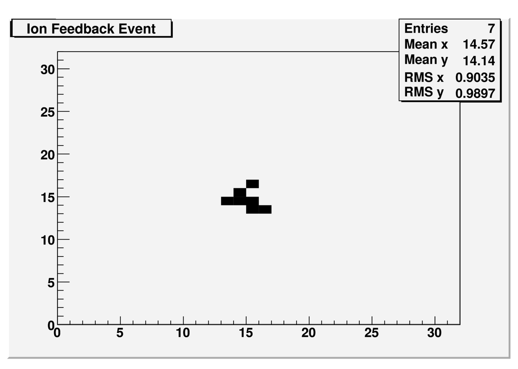


(b) RICH Panel data readout

Figure 3: Sensor array

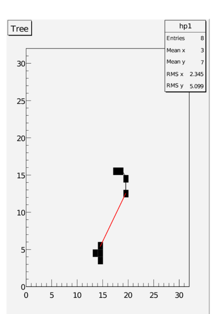


(a) HPD photo and schema

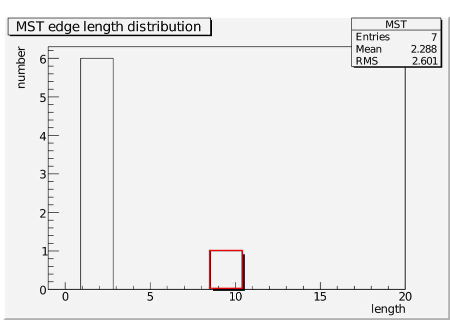


(b) HPD data readout, Ion feedback event

Figure 4: Individual HPD



(a) Minimal spanning tree of the pixels

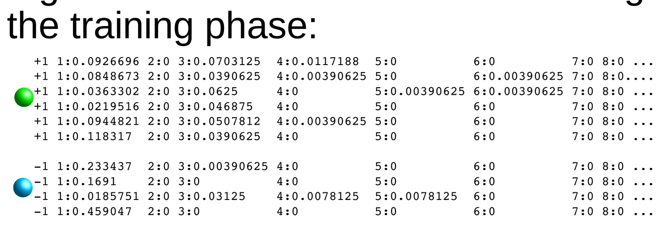


(b) Histogram of tree edge lengths

Resulting vector of the illustrating histogram (b):

[0 6 0 0 0 1 0 0 0 0 0]

Subdetector experts can label histograms as positive or negative samples, according to the event type, e.g. ion feedback event or not during the training phase:



A Support Vector Machine Classifier is used [3]:
•Input feature vectors are mapped into a higher dimensional feature space by a $\Phi$ kernel function
•A separating plane is calculated, which optimally separates the training vectors

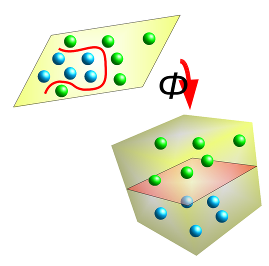


Figure 6: Classification

Comparison with the current approach, cutting on any cluster bigger than 5 pixels for:

- 1000 events 66.2%
- 3400 events 72.9706%
- 4000 events 77.975%

cross validation accuracy with 4-fold cross validation (50% positive and 50% negative samples of IFB)

•Statistics for ion feedback rate of each individual HPD vacuum tube

Figure 7: Post processing

## Further Developments

All libraries and code are compatible with LHCb's software stack, and the LCG software distribution also includes [3]. Some further integration is still required, and speed optimization with the introduction of parallelism is also needed, because of the computationally intensive full-graph construction, when the minimal spanning tree is created. Furthermore other studies would be interesting, with more events, teaching samples and also other classes of events.

## References

[1] Collaboration, The LHCb, "LHCb : Technical Proposal", CERN, 1998.
[2] Collaboration, The LHCb, LHCb reoptimized detector design and performance: Technical Design Report, CERN, 2003.
[3] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

This work has been supported by CERN and the Hungarian National Research Fund grant No. OTKA F04672.