Development of Pattern Recognition Software for Tracks of Ionizing Radiation In Medipix2-Based (TimePix) Pixel Detector Devices

R Vilalta[1], S Kuchibhotla[1], R Valerio[1], and L Pinsky[2]

[1] Department of Computer Science, University of Houston, 4800 Calhoun Rd., Houston TX 77204-3010, USA

[2] Department of Physics, University of Houston, 4800 Calhoun Rd., Houston TX 77204-3010, USA

E-mail: {vilalta,skuchibhotla,rvalerio}@cs.uh.edu, pinsky@uh.edu

Abstract. The principal aim of our project is to develop an efficient pattern recognition tool for the automated identification and classification of tracks of ionizing radiation as measured by a TimePix version of the hybrid semiconductor Medipix2 pixel detector system. Such a software tool would have a number of applications including dosimeters to assess the risk of human exposure to radiation, and area monitors to characterize the general background radiation environment harmful to humans and electronic equipment. We are particularly interested in the development of the real-time analysis software needed to support an operational dosimeter that can assess the radiation environment during space missions. Our software development project makes use of data taken in beams of heavy ions at HIMAC (Heavy Ion Medical ACcelerator Facility) in Chiba, Japan, including data from several different heavy ions with similar Linear Energy Transfers (LETs) for calibration purposes. We describe two modules of our pattern recognition tool: feature generation and classification. Our first module builds on a segmentation algorithm that identifies tracks from the pixel image assuming an approximately elliptical form that varies in size and degree of elongation based on multiple factors, including the particle species and angle of incidence. Determining the charge and energy of the particles creating each track is a particularly challenging task because different energy and charge incident particles can produce very similar patterns. Our classification module invokes different algorithms such as decision trees, support vector machines, and Bayesian classifiers.

1. Introduction

Recent leading-edge technologies have opened an avenue of research for the design and development of active portable space radiation dosimeters; potential applications exist on space projects, future EVA suits and lunar habitats, predicting conditions that can damage electronic equipment, and in medical imaging (e.g., M.D. Anderson Proton Therapy Center's calibrated ion chambers). Of particular interest is a dosimeter that can assess the radiation environment during Solar Particle Events. One excellent candidate technology stems from the Medipix2 Collaboration, based at CERN (European Organization for Nuclear Research) in Geneva, Switzerland, that has developed a pixel detector technology with a clear applicability in active radiation dosimeters [1, 2].

Medipix2 provides an opportunity to embark on a rapid, low-risk program to provide a major step forward in the equipment available for the monitoring of space flight dosimetry. Using a silicon-based detector version, this technology has an enormous dynamic range in sensitivity (from minimum ionizing muons through very heavy ions). Moreover, the Medipix2 chip itself has been demonstrated to be able to survive and perform for extended periods in strong radiation fields. Nevertheless, the technology currently lacks a pattern recognition tool that can accurately predict the type of source of radiation captured by the pixel detector device. Such tool would bridge the gap between the advanced technology embedded in the detector device and the needs of many commercial and research applications.

In this paper we describe a pattern recognition software tool for the automated classification of sources of ionizing radiation as captured by the hybrid semiconductor pixel detector Medipix2. Our pattern recognition mechanism is designed to predict the type of radiation and intensity encountered in space radiation environments. We report on actual data with beams of heavy ions that are characteristic of space radiation environments. Our data stems from a series of controlled experiments performed at HIMAC (Heavy Ion Medical ACcelerator) in Chiba, Japan, where different heavy ions with similar Linear Energy Transfers (LETs) were used for calibration purposes. The extraction of relevant characteristics for each region of pixels on the 256×256 Medipix2 matrix exploits physical properties of possible sources of radiation.

This paper is organized as follows. Section 2 provides an overview of the Medipix2 chip. Section 3 describes our pattern recognition software. Section 4 shows experimental results. Lastly, Section 5 gives a summary and conclusions.

2. Preliminary Information: The MediPix2 Device

The Medipix2 device is a 256 x 256 pixel readout CMOS-based integrated circuit in which the readout circuitry for each pixel is embedded totally within the footprint pixel area. For use as a charged particle detector, the technology behind MediPix2 enables us to employ silicon (Si) detector layers up to several mm thick[1]. When a penetrating heavy ion traverses the Si and produces a core of charge carriers surrounded by a halo of carriers associated with the track structure, the diffusion of the collected charge by the time it reaches the Medipix2 pixels is generally cone shaped with the charge from the most distant part of the track being disbursed the widest and that from nearest to the pixel contact remaining closest to the track itself. Models of this diffusion predict and measurements confirm that effectively one is seeing an analog-amplified picture (albeit very non-linear) of the track structure cross-section. The interested reader can extract details about the pixel electronics circuitry from several sources [3–5].

A schematic view of the formation of a pixel image of the track structure of an incident traversing charged particle is shown in Figure 1. The *drift cone* of charge is dependent on the Linear Energy Transfer (LET) and the track structure of the particle, and the *footprint* of the image on the pixels is in general distributed in a conic section depending upon the angle of incidence of the particle. Upward moving particles will yield very similar, but slightly modified footprints, depending upon their energy.

Figure 2 shows an image of a frame from a normally incident beam taken at HIMAC, along with a magnification of one of the individual track images. The visible features in the track structures can be used to distinguish the charge and energy of the incident particles. Our goal is to automate the identification and classification of track images to predict the source of the ionizing radiation. This is the task of the pattern recognition software, which we describe next.

[1] The data used in our experiments corresponds to 300 μm thick Si layers.

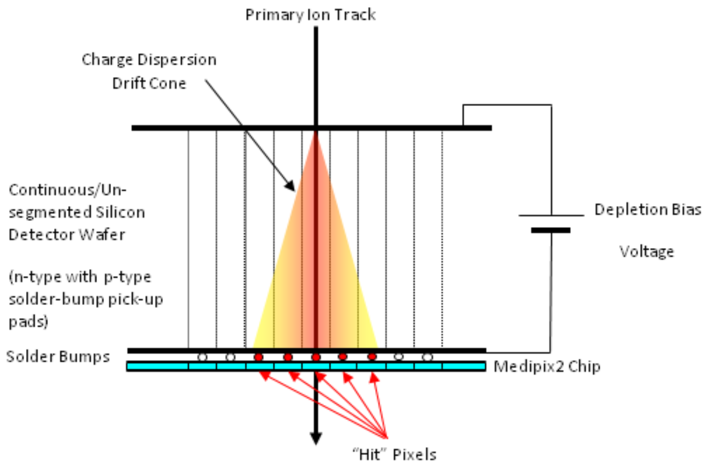

Figure 1. A schematic view of the formation of a pixel image of the track structure of an incident traversing charged particle. The *drift cone* of charge is dependent on the LET and the track structure of the particle, and the footprint of the image on the pixels is in general distributed in a conic section depending upon the angle of incidence of the particle.

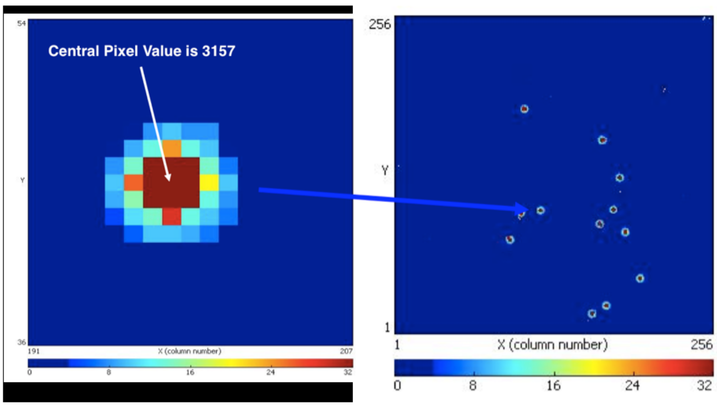

Figure 2. A single frame (right) from a normal incidence beam taken at HIMAC (11B beam with a 1 ms shutter time). The frame on the left is a magnification of one of the track images.

3. Pattern Recognition Software Tool

Our software assumes as input an image generated by a TimePix-based detector after exposure to a source of ionizing radiation. At this stage we are focusing on algorithms to identify and parse heavy ion tracks only; later we will incorporate features to distinguish between single charge particles. The first step is to identify all tracks in the image using a segmentation operator. Our algorithm identifies clusters of pixels assuming an elliptical form that vary in size and degree of elongation based on multiple factors, including the type of radiation source and the angle of incidence. Our algorithm extracts clusters using Thresholding [6] and Connected Component

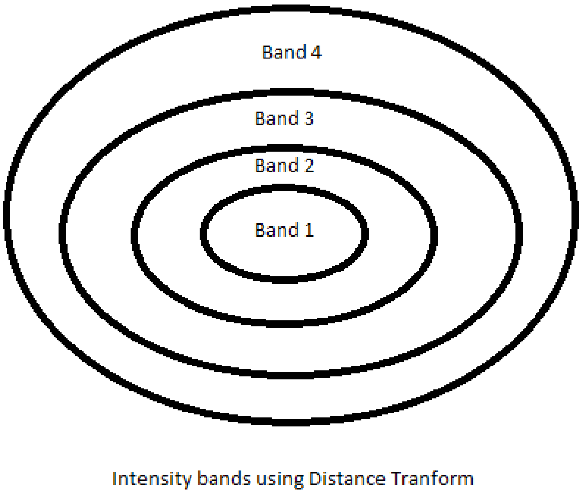

Figure 3. Intensity bands; mean intensity of band 2 includes that of band 1; mean intensity of band 4 implies mean intensity of the entire cluster.

Analysis [7]; each cluster becomes a track amenable to analysis. We employ Matlab as the image analysis tool.

3.1. Feature Generation

Generating relevant features from each track is challenging because different sources of radiation produce very similar tracks. Subtle features carry high discriminatory power such as the rate of change of the intensity cone from inner ellipses to outer ellipses. Our set of feature descriptors is as follows:

- *Area*, defined as the total number of pixels within the track.
- *Volume*, computed as the sum of the energy of each pixel within the track.
- *Eccentricity* of the bounding ellipse of the track.
- *Major and Minor Axis* of the bounding ellipse of the track.
- *Mean energy intensity* of four bands within the bounding ellipse.

All features excluding the last are easily derived once an elliptical boundary is assigned to each track. For the last feature we extracted new information as follows. The idea is to capture the flow of energy within the track; our algorithm creates four bands for each ellipse, as shown in Figure 3, and computes the average energy within each band (band 4 is the average energy for the entire track, whereas band 1 corresponds to the energy of the innermost band only). We employed a technique called distance transform [8] to obtain the bands. Distance transform involves computing the perimeter pixels of the track and then labeling each pixel within the track with the distance to the nearest perimeter pixel; we used Euclidean distance as our distance metric. The result is a new pixel representation within a track where higher values correspond to pixels farther away from the perimeter or border (i.e., closer to the center of the track) and low values correspond to pixels closer to the border. Such distances can then be used to generate different concentric bands.

3.2. Classification

The automatic identification of sources of ionizing gas from tracks can be cast as a classification problem. We assume an input training set $T = \{(\mathbf{x}, y)\}$, where each element $\mathbf{x} = (a_1, a_2, \cdots, a_9)$ in the set is a vector of nine features (area, volume, eccentricity, major axis, minor axis, mean energy within band {1,2,3,4}). Each vector $\mathbf{x} \in \mathcal{X}$ is attached a label $y \in \mathcal{Y}$. Our classes are the elements acting as sources of ionizing radiation (Si, Fe, O, N, and Ne). The outcome of the classifier is a function f mapping a track (characterized by a feature vector) to one element class, $f : \mathcal{X} \rightarrow \mathcal{Y}$. Function f can then be used to predict the class of new unseen tracks.

We invoked several classifiers to train on dataset T, including non parametric techniques (decision trees, support vector machines, and neural networks) and one parametric technique (Bayesian classifier). All of these techniques are able to delineate complex decision boundaries over the feature space, while avoiding models overly complex (i.e., while avoiding over-fitting).

4. Experiments

For our experimental analysis we processed images obtained from beams of heavy ions at HIMAC. In our first experiment we had five classes corresponding to tracks of Si, Fe, O, N, and Ne. For each class we generated 4000 samples for training, and 2000 extra samples for testing. We used the training set (20,000 samples) to build a model (e.g., decision tree, neural network, etc.) and validated the model on the testing set (10,000 samples). Results are shown in Table 1. The 1st column shows the classification algorithms used for our experiments; the 2nd column shows accuracy without any distinction for the angle of incidence, while columns 3-4 show equivalent results but under fixed incidence angles. When looking at results for all incidence angles, accuracy reaches a maximum of around 80%. Varying the incidence angle does produce different results, but differences with the top performance algorithm (decision trees) are about the same. Overall an angle of 60 degrees produces the best setting to discriminate among classes.

An analysis of the confusion matrix for the first experiment shows that three of the five element classes, namely Si, Ne, and O, produce very similar tracks. For our second experiment we grouped these elements into a single class, and repeated the analysis under the new class representation. Table 1 (right) shows our results, where no distinction exists among incidence angles. We noticed a significant improvement in predictive performance, which supports our belief that the classes grouped together are the most difficult for classification.

In addition we tested the relevance of our features by measuring the amount of mutual information between each feature and the target class, using a metric known as information gain [9]. Table 2 shows our results (features are ordered from left to right based on information gain). The high rank of bands 4 and 3 point to the outer part of every track as the region where there is notable change of energy that can be used to differentiate among different sources of ionizing radiation. This information is relevant to understand efficient approaches to represent class signatures, and complements previous studies along this research direction [10, 11].

Table 1. Average predictive accuracy with beams coming at different incident angles. Numbers enclosed in parentheses represent standard deviations.

Learning Technique	Predictive Accuracy (5 classes)			Predictive Accuracy (3 classes)
	All angles	0 degrees	60 degrees	All angles
Decision Trees	80.44 (0.50)	79.41 (0.44)	81.10 (1.47)	99.23 (0.26)
Support Vector Machines	71.13 (0.72)	71.24 (1.33)	76.38 (0.78)	97.08 (0.33)
Neural Networks	77.20 (1.21)	79.17 (1.19)	81.16 (1.52)	99.22 (0.24)
Bayesian Classifier	65.34 (0.79)	67.56 (0.96)	72.58 (1.10)	93.39 (0.46)

Table 2. Features ranked based on information gain.

	Features								
	Volume	Band 4	Band 3	Major Axis	Area	Minor Axis	Band 2	Eccent.	Band 1
Info. Gain	1.33	1.25	0.94	0.92	0.87	0.87	0.63	0.45	0.40

5. Summary and Conclusions

This paper describes a pattern recognition system designed to work together with the MediPix2 device. Our goal is to produce an accurate classifier to predict the type of ionizing radiation corresponding to each track recorded by MediPix2. As a direct application, such system can be employed as a local alarm capability that can be set based on logical combinations of relatively sophisticated dose-related factors (e.g., during a mission on the International Space Station).

Our experiments show accuracy values around 80% when classifying tracks produced by five possible elements. Grouping three of these classes together yields values around 98%. Some avenues for future work are the following: 1) we can vary the shutter time for Medipix2 to change the net pixel occupancy per frame; 2) we can look for additional features to enhance our track representation, such as long range δ-ray artifacts surrounding the main footprint images; and 3) we can look for a more granular representation of intensity bands (Figure 3) in search for a set of features with high discriminant power.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grants no. IIS-0812372 and IIS-448542.

References

[1] Llopart, X.; Ballabriga, R.; Campbell, M.; Tlustos, L; and Wong, W.: "Timepix, a 65k Programmable Pixel Readout Chip for Arrival Time, Energy and/or Photon Counting Measurements". *Nuclear Instruments and Methods in Physics Research A*, Vol. 581, Issues 1-2, pages 485-594, October 2007.

[2] Llopart, X.; Campbell, M.; Dingploi, R.; Sansegundo, D.; and Pernigotti, E.: "Medipix2: A 64-k Pixel Readout Chip with 55-μm Square Elements Working in Single Photon Counting Mode". *IEEE Transactions on Nuclear Science*, Vol. 49, Issue No. 5, pp. 2279-2283, 2002.

[3] Fiederle, M.; Greiffenberg, D.; Idarraga, J. et. al.: "Energy Calibration Measurements of MediPix2". *Nuclear Instruments and Methods in Physics Research A*, Vol. 591, pp. 75-79, 2008.

[4] Kroupa, M.; Jakubek, J.; Krejci, F.: "Charge Collection Characterization with Semiconductor Pixel Detector Timepix" *Proc. of the IEEE Nuclear Science Symposium Conference Record* pp. 259-262, 2008.

[5] Jakubek, J.; Cejnarova, A.; Holy, T.; Pospisil, S.; Uher, J.; Vykydal, Z.: "Pixel Detectors for Imaging with Heavy Charged Particles". *Nuclear Instruments and Methods in Physics Research A*, Vol. 591, No. 1, pp. 155-158, 2008.

[6] Otsu, N.: "A Threshold Selection Method from Gray-Level Histograms". *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 9, No. 1, pp. 62-66, 1979.

[7] Haralick, Robert M.; Shapiro, Linda G. *Computer and Robot Vision, Addison-Wesley*, Vol. 1, pp. 28-48, 1992.

[8] Breu, H.; Gil, J.; Kirkpatrick, D.; Werman, M.: "Linear Time Euclidean Distance Transform Algorithms". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, No. 5, pp. 529-533, 1995.

[9] Rokach, L.; Maimon, O. *Data Mining with Decision Trees: Theory and Applications*, World Scientific Publishing, 2008.

[10] Holy, T.; Heijne, E.; Jakubek, J.; Pospisil, S.; Uher, J.; Vykydal, Z.: "Pattern Recognition of Tracks Induced by Individual Quanta of Ionizing Radiation in Medipix2 Silicon Detector" . *Nuclear Instruments and Methods in Physics Research A*, Vol. 591, No. 1, pp. 287-290, 2008.

[11] Bouchami, J.; Gutierrez, A.; Holy, T.; Houdayer, A.; Jakubek, J.; Lebel, C.; Leroy, C.; Macana, J.; Martin, J.; Pospsil, S.; Prak, S.; Sabella, P.; Teyssier, C.: "Measurement of Pattern Recognition Efficiency of Tracks Generated by Ionizing Radiation in a Medipix2 Device". *Nuclear Instruments and Methods in Physics Research A*, 2010.