

CENTRO BRASILEIRO DE PESQUISAS FÍSICAS

COORDENAÇÃO DE ATIVIDADES TÉCNICAS - CAT
LABORATÓRIO DE PROCESSAMENTO DIGITAL DE SINAIS E IMAGENS - LPDSI

**Localização Online do Vértice Primário de
Colisão Utilizando a Transformada de Hough
Adaptativa com Condição de Parada Dinâmica**

Gabriel Mayrink da Rocha Hospodar
Aline da Rocha Gesualdi
José Manoel Seixas
Márcio Portes de Albuquerque

Rio de Janeiro
Junho/2006

Resumo

Este relatório aborda o trabalho realizado durante o período de bolsa de iniciação científica na Coordenação de Atividades Técnicas (CAT) do CBPF. O trabalho consistiu na construção de um sistema em tempo real para o processamento de imagens de colisões de partículas. O desenvolvimento foi feito no âmbito da colaboração entre o CBPF e o CERN (Centro Europeu de Física de Partículas), localizado na Suíça. O objetivo do projeto consistiu em obter o vértice primário de colisão de acordo com as restrições de implementação e de latência estabelecidas pela colaboração. O objetivo do trabalho foi alcançado com a eficiência do método de localização de vértices de $\approx 88\%$.

Palavras-Chave

instrumentação eletrônica, CERN, física de partículas, processamento de imagens, transformada de Hough

Conteúdo

1	Introdução	1
1.1	Centro Europeu de Física de Partículas - CERN	1
1.2	Objetivo	2
1.3	O Documento	2
2	Revisão de Literatura	2
2.1	Localização por Ajuste Linear de Retas (<i>ZFinder</i>)	2
2.2	Localização por Filtros de <i>Anneling</i> Determinísticos	3
3	Fundamentos Teóricos	3
3.1	Transformada de Hough Adaptativa (THA)	4
3.2	Convolução Bidimensional	6
3.3	Filtragem 2D	6
4	Materiais, Técnicas e Métodos	6
4.1	O Conjunto de Dados	6
4.2	Condição de Parada da Transformada de Hough Adaptativa	6
4.3	Implementação do Algoritmo	8
5	Resultados	8
6	Discussão	11
7	Conclusão	11

Lista de Figuras

1	O detector ATLAS e seus quatro sistemas de detecção.	1
2	Ilustração das coordenadas do <i>ZFinder</i>	3
3	Ilustração do processo de recursão determinística para a aglomeração de um conjunto de pontos.	4
4	Fluxograma do algoritmo da transformada de Hough adaptativa.	5
5	Exemplo de um evento de elétrons individuais com alta luminosidade: (a) sem empilhamento e (b) com empilhamento.	7
6	Histograma da diferença $z_{true} - z_{vtx}$ do algoritmo THA.	8
7	Histograma da diferença $z_{true} - z_{vtx}$ do algoritmo <i>ZFinder</i>	9
8	Tempo de execução do algoritmo THA.	9
9	Tempo de execução do algoritmo <i>ZFinder</i>	10
10	Número de iterações realizadas pelo algoritmo THA.	10

1 Introdução

Durante os últimos anos a Engenharia Eletrônica vem se desenvolvendo num ritmo bastante acelerado e tem se tornando cada vez mais presente no cotidiano da população. Esta área permite a realização das mais variadas aplicações, tendo seu papel fundamental para o bem-estar da sociedade e para o desenvolvimento do País.

Uma das áreas mais relacionadas com a Engenharia Eletrônica é a Computação, que influi diretamente no seu desenvolvimento. Outro exemplo de integração está na Engenharia com a Física. Esta interação vem possibilitando a construção de grandes aparatos, como os sofisticados detectores de partículas do CERN (Centro Europeu de Física de Partículas). O papel dos engenheiros nesses projetos se encontra tanto na elaboração dos instrumentos para as medidas de propriedades físicas, quanto na transformação de descobertas científicas em inovações tecnológicas. O presente trabalho foi realizado nesse ambiente de colaboração entre Engenharia e Física.

1.1 Centro Europeu de Física de Partículas - CERN

O CERN é o maior centro de pesquisa de Física de Partículas do mundo e está localizado na Suíça[1]. Nele, encontram-se diversas aplicações de Engenharia Eletrônica que tornam possíveis as realizações dos experimentos físicos. Lá ocorrem experimentos audaciosos que envolvem instituições de pesquisa do mundo todo.

Colisionadores de partículas estão sendo contruídos no CERN e um deles é chamado LHC (*Large Hadron Collider*). O túnel do LHC possui uma circunferência de $27km$ e está localizado a $100m$ abaixo do solo. Paralelamente, o detector de partículas ATLAS (*A Toroidal LHC Apparatus*) está sendo contruído numa colaboração de 150 institutos de pesquisa de 34 países, em que o objetivo básico consiste em realizar novas descobertas sobre a natureza fundamental da matéria e sobre as forças básicas existentes no Universo[3].

O detector ATLAS será formado por quatro sistemas de detecção independentes posicionados cilindricamente ao redor de um dos pontos de colisão, como ilustra a Figura 1.

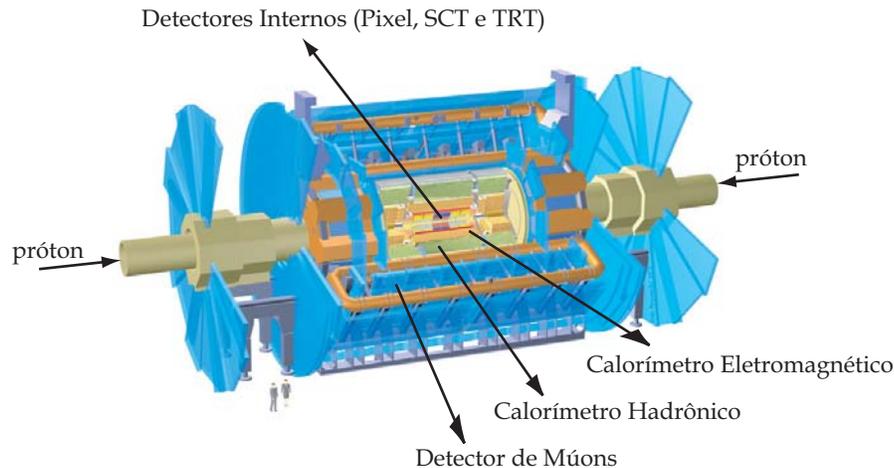


Figura 1: O detector ATLAS e seus quatro sistemas de detecção.

O experimento no ATLAS está previsto para iniciar em 2007. Prótons com $14TeV$ de energia no centro de massa colidirão a cada $25ns$ dentro do LHC com o objetivo principal de comprovar

a existência do Bóson de Higgs[4]. A quantidade de eventos (colisões) gerada será enorme, o que inviabiliza o armazenamento de todos os dados para um posterior estudo *offline*. Conseqüentemente, faz-se necessário um sistema de validação *online* para filtrar os eventos, eliminando ruídos para que apenas os eventos de interesse sejam armazenados.

O sistema de filtragem é composto por três níveis:

- Primeiro nível (LVL1): Eletrônica rápida de detecção de eventos de interesse. Possui granularidade reduzida. Este nível de filtragem não somente reduz a taxa de eventos de $40GHz$ para $100KHz$, como também identifica regiões de interesse ao longo do detector para que a análise dos níveis seguintes de validação seja mais direcionada. Este nível tem que ser bastante rápido já que as colisões ocorrerão a cada $25ns$.
- Segundo nível (LVL2): Utiliza as zonas de interesse detectadas pelo LVL1 com o máximo de granularidade dos detectores, reduzindo a taxa de eventos de $100KHz$ para $1KHz$. Este nível terá mais tempo de processamento, pois ele analisa menos dados.
- Terceiro nível (filtro de eventos): Refina a taxa de eventos para $100Hz$. Este nível decide sobre o armazenamento final de cada evento.

1.2 Objetivo

O trabalho desenvolvido no estágio se aplica ao LVL2 e busca identificar o ponto de colisão através de algoritmos de processamento de imagens. O algoritmo utilizado foi uma variação da técnica de transformada de Hough[5], denominada transformada de Hough adaptativa. O sistema tem restrição de tempo de $1ms$ numa máquina com processamento de $4GHz$.

1.3 O Documento

O documento está dividido em seis seções: Revisão de Literatura; Fundamentos Teóricos e Práticos; Materiais, Técnicas e Métodos; Resultados; Discussão e Conclusão.

A Seção Revisão de Literatura aborda os assuntos e métodos estudados para o desenvolvimento do trabalho. A Seção Fundamentos Teóricos apresenta as técnicas de processamento de imagens utilizadas (transformadas de Hough, convolução bidimensional e filtragem 2D). A Seção Materiais, Técnicas e Métodos aborda o conjunto de dados utilizado, a solução encontrada para o problema de localização de vértices de colisão e a implementação do algoritmo.

Em seguida, tem-se a Seção Resultados e, posteriormente, a Seção Discussão, na qual é feita uma análise crítica do estágio. Por último, tem-se a Seção Conclusão.

2 Revisão de Literatura

Algumas abordagens sobre a localização de pontos de colisão que já foram utilizadas em sistemas de localização de vértices de colisão serão apresentadas a seguir. Elas tratam de métodos de processamento de imagens baseados em diferentes princípios de filtragem e todas têm o objetivo de serem eficientes quanto à precisão e à velocidade de execução.

2.1 Localização por Ajuste Linear de Retas (*ZFinder*)

Este é o algoritmo ora em uso no CERN, sendo a principal referência para comparação com a proposta desse trabalho. Na Localização por Ajuste Linear de Retas, o vértice de colisão é encontrado antes do processo de reconstrução. O algoritmo está baseado na relação aproximadamente

linear entre as coordenadas z e ρ das trajetórias das partículas em campos magnéticos uniformes paralelos ao eixo z .

A trajetória de uma partícula carregada dentro de um campo magnético uniforme é uma hélice. Em coordenadas cilíndricas, com o eixo z paralelo ao campo magnético, a hélice passa pelo vértice com coordenadas $(0,0,z_v)$.

Se (z_1, ρ_1) e (z_2, ρ_2) são coordenadas de dois pontos ao longo da trajetória da partícula (com ρ_1 diferente de ρ_2), z_v pode ser calculado por uma simples extrapolação linear, conforme a Equação 1. A Figura 3 ilustra esse processo apresentando as coordenadas cilíndricas do ATLAS.

$$z_v = \frac{z_2\rho_1 - z_1\rho_2}{\rho_1 - \rho_2} \quad (1)$$

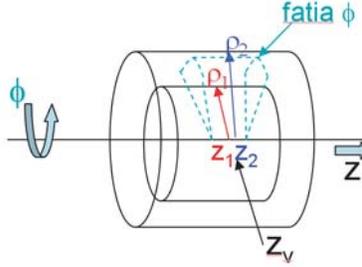


Figura 2: Ilustração das coordenadas do *ZFinder*.

O *ZFinder* consiste no primeiro algoritmo a utilizar as coordenadas cilíndricas para aproximar a hélice observada em (x, y, z) para uma reta em (ϕ, ρ, z) . A eficiência desse algoritmo em termos de precisão é de $\approx 86\%$.

2.2 Localização por Filtros de *Annealing* Determinísticos

O método de Recursão Determinística (*Deterministic Annealing*) é utilizado para encontrar o melhor conjunto de pontos que formam uma trajetória. Esta técnica consiste em outro método de localização de vértices primários adotado pelo colisionador de partículas CMS, também em construção no CERN. Este método oferece uma característica importante para a localização do vértice de colisão, que é a habilidade de localizar os vértices em um ambiente ruidoso sem um conhecimento prévio de quantos vértices devem ser encontrados.

O passo inicial do algoritmo consiste em caracterizar a topologia do sistema. Para isto, é utilizado o método de aglomeração de pontos *Apex* (*Apex Point*) [6], onde cada trajetória é substituída por um ponto representativo de sua densidade. Os resultados de simulações para 4000 eventos de $b\bar{b}$ sem empilhamento mostram uma eficiência de aproximadamente 92,75% para vértices primários.

O estudo deste algoritmo foi interessante para o desenvolvimento do projeto de estágio devido ao método de combinação e aglomeração de pontos utilizados. A Figura 3 ilustra o processo de recursão determinística para a aglomeração de um conjunto de pontos [7].

3 Fundamentos Teóricos

O sistema para localização *online* do vértice primário de colisão é fundamentado no conceito da transformada de Hough. O principal objetivo deste algoritmo é detectar padrões complexos a partir de pontos numa imagem binária. O segundo nível do sistema de validação *online* de eventos possui uma restrição de tempo de execução de $1ms$ numa máquina com processamento de $4GHz$.

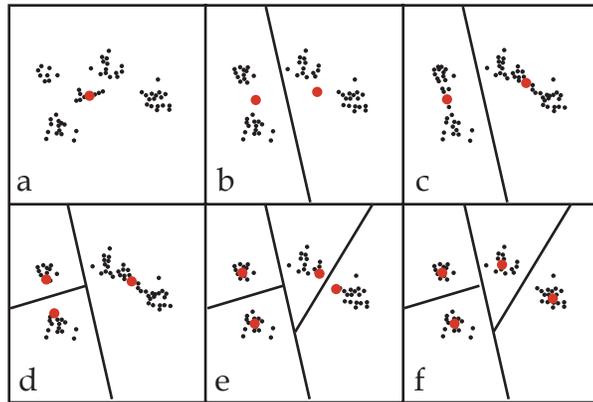


Figura 3: Ilustração do processo de recursão determinística para a aglomeração de um conjunto de pontos.

A transformada de Hough é um método bastante utilizado em processamento de imagens digitais com elevado ruído de fundo. Ela foi proposta em 1962 por Paul V. C. Hough, em artigo intitulado “*A method and means for recognizing complex patterns*”. Esta transformada foi reconhecida na época como uma técnica quase única na análise de imagens com ruído e dados incompletos. Contudo, sua utilização foi bastante lenta devido principalmente ao elevado custo e à complexidade computacional requeridos para a sua implementação. A descoberta da relação da transformada de Hough com outros métodos de reconstrução e detecção de padrões tem incentivado a pesquisa na área de Física de Partículas, fazendo com que novas aplicações surjam para diversos problemas.

A Transformada de Hough é baseada no mapeamento da imagem para extrair os seus parâmetros e analisá-los. Ou seja, o espaço \mathbb{R}^2 é transformado num espaço de parâmetros. Com isso, o processamento realizado pela transformada converte um problema de detecção global no espaço de dados em um problema mais simples de detecção local no espaço de parâmetros.

Sabendo-se que, segundo a geometria dos detectores, o vértice primário de colisão representa a extrapolação dos pontos no espaço para a região de colisão, o problema é resumido a identificar a reta que possui o maior número de pontos colineares no gráfico de entrada $\rho \times z$. O coeficiente linear desta reta será o vértice primário de colisão.

3.1 Transformada de Hough Adaptativa (THA)

O princípio da THA baseia-se no fato de inicialmente não se saber em que região do espaço de parâmetros o valor máximo será encontrado. Desta forma, constrói-se inicialmente um acumulador com poucas células, o que torna o processamento mais rápido. Em seguida, encontra-se a região de pico deste acumulador e verifica-se se ele satisfaz uma condição pré-determinada de precisão. Caso ele não a satisfaça, cria-se um novo acumulador, recursivamente, limitado pelas delimitações da célula onde o último pico foi encontrado. Caso contrário, os coeficientes da reta procurada são identificados.

Essas recursões sobre as células onde os picos são encontrados aumentam cada vez mais a precisão da busca. A Figura 4 apresenta um fluxograma onde o algoritmo da THA é formalizado.

A condição de parada da THA é apresentada na Seção 4.2.

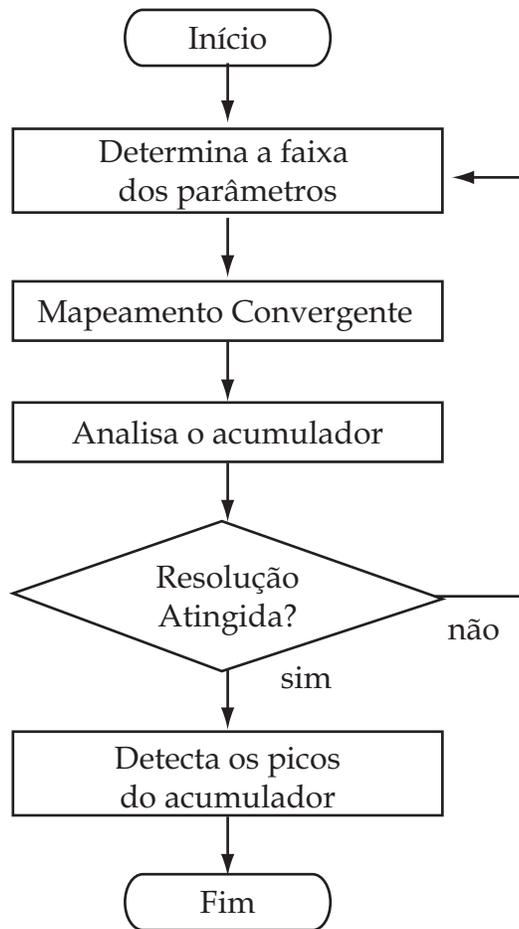


Figura 4: Fluxograma do algoritmo da transformada de Hough adaptativa.

3.2 Convolução Bidimensional

A convolução é uma operação local que representa a função de transformação a ser aplicada a um sinal ou imagem. Ela é definida pela Equação 2.

$$[f(x, y) \otimes g(x, y)](m, n) = \sum_{y=-\infty}^{\infty} \sum_{x=-\infty}^{\infty} f(x)g(m-x, n-y) \quad (2)$$

A convolução bidimensional foi utilizada no algoritmo de filtragem do espaço de parâmetros da transformada de Hough adaptativa.

3.3 Filtragem 2D

A filtragem bidimensional foi o método utilizado para suavizar o espaço de parâmetros da transformada de Hough, reduzindo os ruídos de alta frequência e evidenciando o valor máximo deste espaço. A localização do valor máximo é o objeto da aplicação da transformada de Hough, uma vez que o vértice de colisão corresponde ao coeficiente linear da THA.

O processo de filtragem utilizou a convolução bidimensional de uma máscara 3×3 gaussiana com o espaço de parâmetros da THA. A máscara gaussiana equivale a aplicação de um filtro passa-baixas bidimensional.

4 Materiais, Técnicas e Métodos

Nesta seção serão apresentados o conjunto de dados de entrada e a condição de parada da THA.

4.1 O Conjunto de Dados

O LHC e o ATLAS ainda estão sendo construídos, ou seja, não há dados reais sobre os eventos de colisão. Apesar disso, os colaboradores precisam de dados para testar, projetar e otimizar seus algoritmos de análise. Para isso, há uma equipe de desenvolvimento, validação e armazenamento de simulações dos dados de colisão e das respostas dos sistemas.

Foram utilizados conjuntos de dados simulados de elétrons individuais, pois eles estão presentes na maioria das assinaturas de interesse para a colaboração. Esses dados possuem ruídos provenientes de diversas fontes, como por exemplo raios cósmicos e empilhamento.

O conceito de elétrons individuais com empilhamento se refere às sensibilizações de regiões bem próximas (ou iguais) do detector geradas pelos subprodutos das colisões. Sendo assim, quanto maior a quantidade de prótons acelerada para colidir, maior será o ruído na projeção dos pontos do espaço nas coordenadas $\rho \times z$.

A Figura 5 apresenta o gráfico de um evento de alta luminosidade ($10^{34} \text{cm}^{-2} \text{s}^{-1}$) projetado em $\rho \times z$.

4.2 Condição de Parada da Transformada de Hough Adaptativa

A condição de parada implementada analisa os dois maiores picos do espaço de parâmetros e verifica se a diferença entre eles é de N contagens. O parâmetro N foi escolhido através de diversas tentativas em que foram verificados o tempo de execução e a eficiência de acerto do ponto de colisão. No âmbito da colaboração, a eficiência é dada pela Equação 3, onde z_{true} é o valor do vértice simulado e z_{vtx} é o valor do vértice obtido pela THA.

$$|z_{true} - z_{vtx}| < 1,0 \text{mm} \quad (3)$$

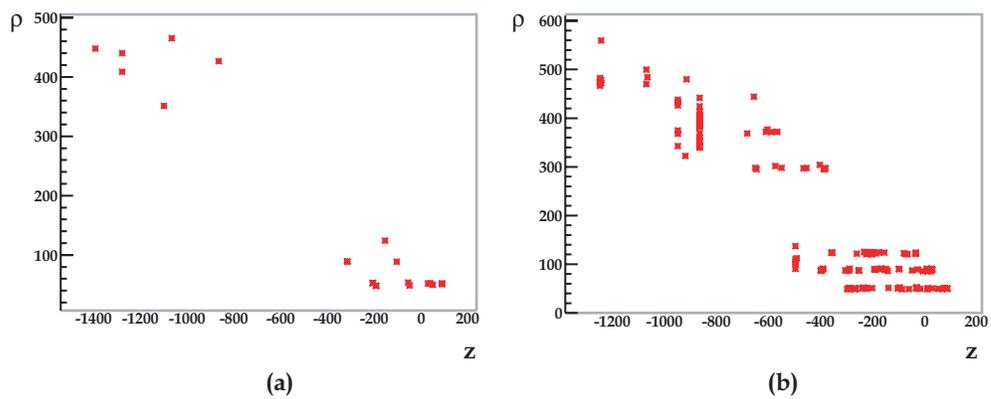


Figura 5: Exemplo de um evento de elétrons individuais com alta luminosidade: (a) sem empilhamento e (b) com empilhamento.

4.3 Implementação do Algoritmo

O desenvolvimento do sistema foi feito no ambiente Athena através das linguagens de programação C e C++. O sistema Athena simula os produtos das colisões, a resposta dos detectores e a resposta do LVL1, permitindo antever os funcionamentos dos algoritmo de análise de dados que serão executados no experimento. Isto permite que os colaboradores testem e aperfeiçoem seus algoritmos.

A visualização dos eventos e a criação dos histogramas de tempo, de eficiência e do espaço de parâmetros foram feitas no programa Root. O Root é um software de análise em C++ bastante utilizado no ambiente do ATLAS, sendo referência para as comparações entre os códigos.

5 Resultados

Esta seção apresenta os resultados obtidos nesse trabalho. A Figura 6 apresenta o histograma do resultado da diferença entre z simulado (z_{true}) e z obtido (z_{vtx}). A Figura 7 apresenta o histograma análogo para o algoritmo referência ($ZFinder$).

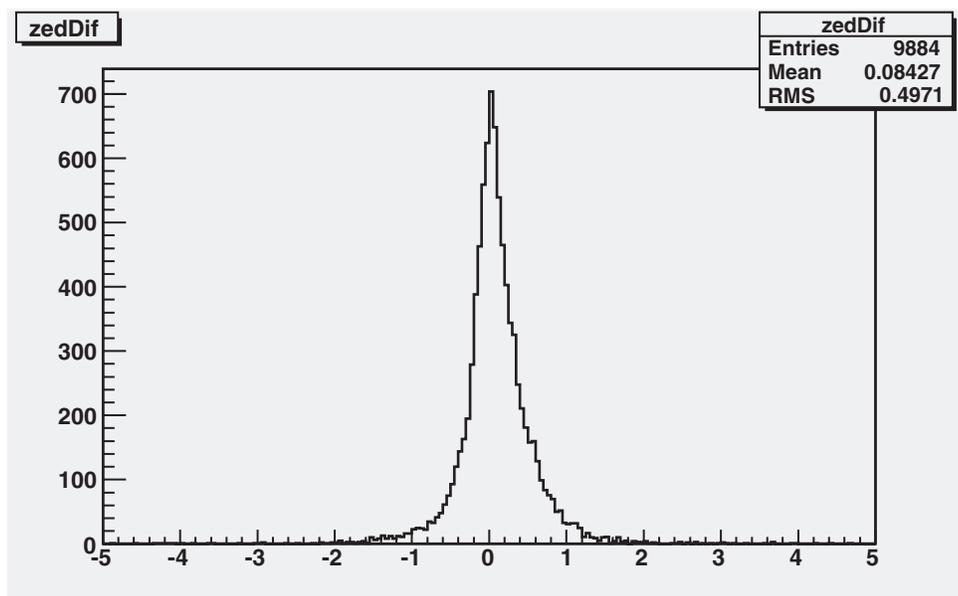


Figura 6: Histograma da diferença $z_{true} - z_{vtx}$ do algoritmo THA.

Os tempos de execução dos algoritmos THA e $ZFinder$ estão representados nas Figuras 8 e 9, respectivamente. O número de recursões realizadas pelo algoritmo THA está ilustrado na Figura 10.

A eficiência do algoritmo THA foi de 87,87% (8394 acertos em 9884 eventos). O algoritmo $ZFinder$, para o mesmo conjunto de dados, possui eficiência de 86,61% (8561 acertos em 9884 eventos). O tempo médio de execução da THA foi de 12,58ms com valor *rms* de 2,886ms em uma máquina com processamento de 500MHz. O algoritmo $ZFinder$ obteve tempo de execução de 23,29ms com valor *rms* de 4,685ms em uma máquina de 500MHz.

Desta forma, o algoritmo implementado durante o período de iniciação científica possui eficiência de localização de vértices compatível com o $ZFinder$ e tempo de execução inferior, sendo viável e interessante para a colaboração.

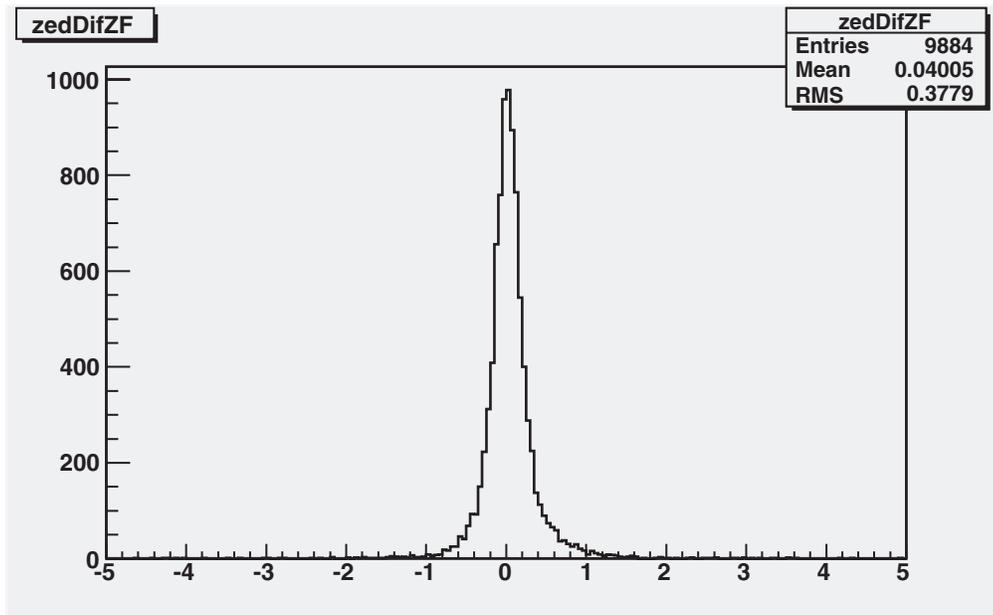


Figura 7: Histograma da diferença $z_{true} - z_{vtx}$ do algoritmo *ZFinder*.

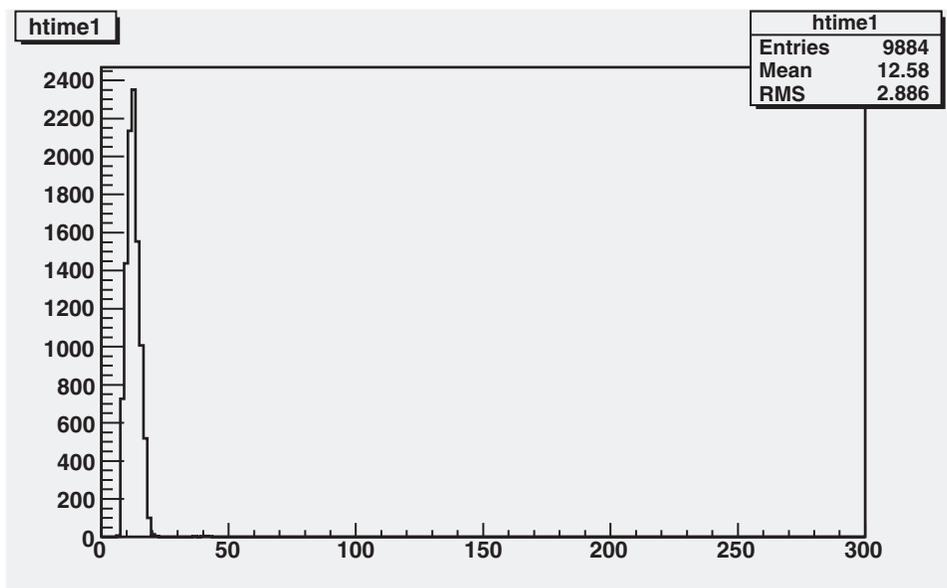


Figura 8: Tempo de execução do algoritmo THA.

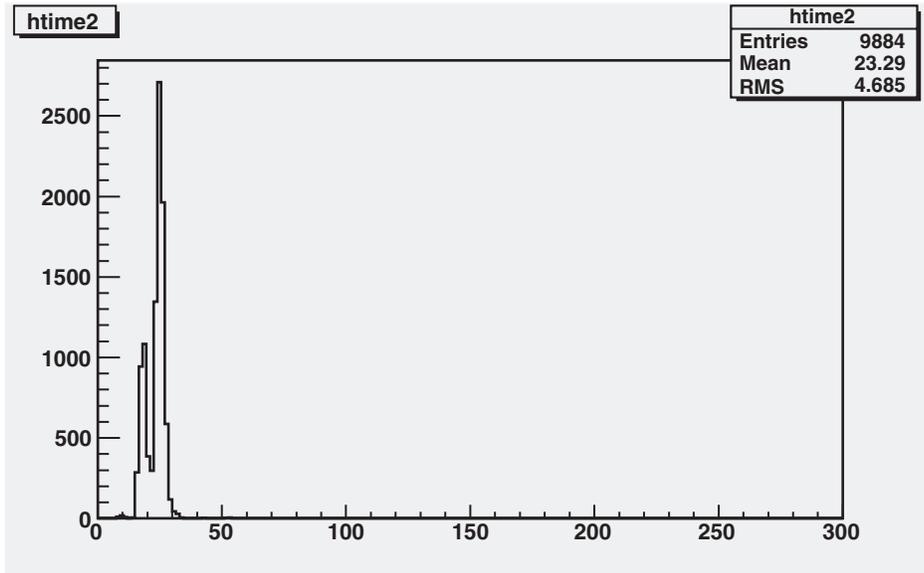


Figura 9: Tempo de execução do algoritmo *ZFinder*.

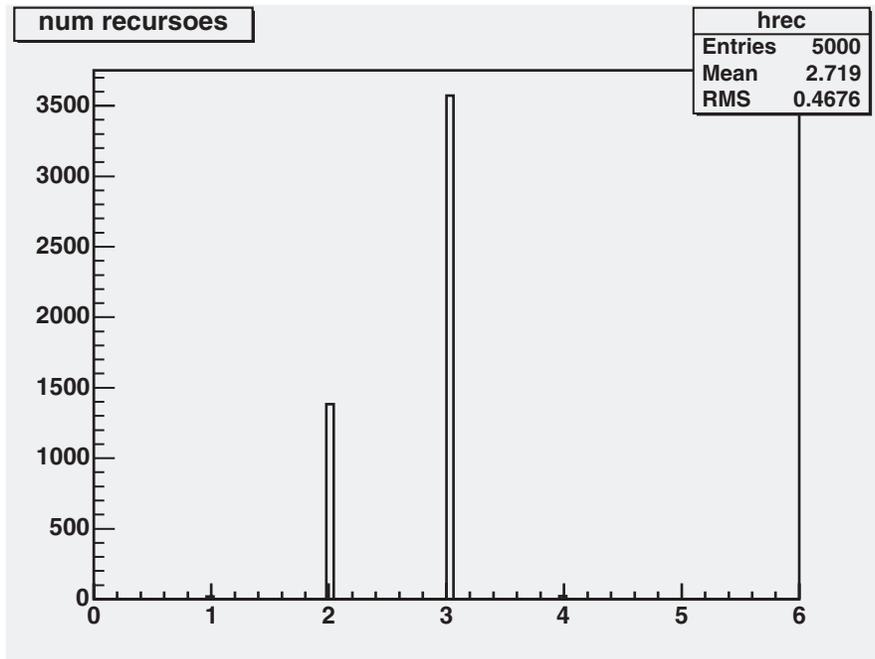


Figura 10: Número de iterações realizadas pelo algoritmo THA.

6 Discussão

Na primeira etapa do estágio supervisionado ocorreu a adaptação ao ambiente do projeto de colaboração CBPF/CERN. Nesta etapa, o CERN e o contexto do experimento ATLAS foram apresentados.

Em seguida, a ferramenta Root foi estudada com a utilização de dados simulados para a visualização de eventos e a criação de histogramas, o que deu início à parte prática do desenvolvimento.

O passo seguinte foi o estudo da transformada de Hough. A primeira abordagem desenvolvida foi a transformada de Hough padrão, seguida da transformada de Hough local, que foi implementada em C++ utilizando-se funções pertencentes ao Root.

Com o conceito de transformada de Hough solidificado, iniciou-se o estudo da transformada de Hough adaptativa em paralelo com discussões sobre qual seria a sua condição de parada. Após esta definição, implementou-se a THA em C++.

Para localizar o pico no espaço de parâmetros foi utilizado um filtro passa-baixas bidimensional, que é a convolução de uma máscara 3×3 gaussiana com o histograma do espaço de parâmetros. Todos os códigos desenvolvidos foram testados no Athena e comparados com os resultados obtidos pelo *ZFinder*.

7 Conclusão

O algoritmo desenvolvido durante o período de iniciação científica (THA), mostrou-se mais rápido do que o algoritmo ora utilizado no CERN, o *ZFinder*. Quanto à eficiência na busca do vértice de colisão, o algoritmo THA alcançou $\approx 88\%$, enquanto que o *ZFinder* alcançou $\approx 86\%$.

Conclui-se que o algoritmo desenvolvido no estágio é superior ao utilizado atualmente no CERN.

Referências

- [1] CERN – Centro Europeu para a Pesquisa Nuclear
URL: <http://www.cern.ch/>
- [2] LHC – The Large Hadron Collider
URL: <http://lhc.web.cern.ch/>
- [3] ATLAS – A Toroidal LHC ApparatuS
URL: <http://atlas.web.cern.ch/>
- [4] Inner Detector Technical Design Report, *ATLAS Inner Detector Community*, ISBN = 92-9083-102-2, ATLAS CERN, Abril, 1997.
- [5] J. Illingworth and J. Kittler, *A survey of the Hough transform*, Computer Vision, Graphics and Image Processing, v. 44, pp. 87-116, 1988.
- [6] R. Frühwirth and K. Prokofiev et al., *New developments in vertex reconstruction for CMS*, Nucl. Instrum. Methods Phys. Res., A 502, 699-701, 2003.
- [7] E. Chabanat and N. Estre, *Deterministic Annealing for Vertex Finding at CMS*, Computing in High Energy Physics and Nuclear Physics (CHEP 2004), Interlaken, Switzerland , pages 287, October 2004.