

Pesquisa sobre Métodos de Interpolação e Solução de Sistemas Lineares

Alunos: Giovanna Paiva, Matheus Sanchez Duda

Curso: Bacharelado em Ciências da Computação

Artigo 1

Título:

Acelerador com reaproveitamento de hardware multiversão de baixo consumo para os filtros de interpolação da estimação de movimento fracionária do codec AV1

Autores:

Daiane Fonseca Freitas, Cláudio Diniz, Mateus Grellert, Guilherme Corrêa

Periódico / Conferência:

Simpósio Brasileiro de Computação e Circuitos Integrados (SBCCI), 2024

Objetivo Principal:

Desenvolver o *Reuse Interpolation Version Core* (RVIC), um acelerador VLSI de baixo consumo energético para filtros de interpolação na etapa de estimação de movimento fracionária (FME/MC) do codec AV1, buscando alta eficiência e baixo custo computacional.

Em outras palavras, o artigo propõe uma arquitetura inovadora que melhora a eficiência energética e acelera a operação dos filtros de interpolação, contribuindo para implementações mais viáveis e eficientes do codec AV1 em sistemas de processamento de vídeo.

Métodos Utilizados:

- **Implementação de Versões Múltiplas dos Filtros de Interpolação:** O acelerador RVIC implementa três versões para cada família de filtros (Regular, Sharp e Smooth) do codec AV1. São usadas uma versão precisa e duas versões aproximadas. Essa estratégia permite reaproveitar o hardware entre diferentes versões, reduzindo a área do circuito e, conseqüentemente, o consumo energético.
- **Abordagem Multiple Constant Multiplication (MCM):** Em vez de utilizar operadores de multiplicação convencionais – que costumam exigir muitos recursos – a arquitetura emprega a técnica MCM. Essa abordagem realiza operações de multiplicação por constantes por meio de somas e deslocamentos, o que simplifica o hardware e diminui os custos computacionais.
- **Técnica de Operand Isolation:** Para otimizar a eficiência energética, o projeto utiliza operand isolation. Essa técnica consiste em isolar circuitos que não estão em uso durante determinados modos de operação, evitando a dissipação desnecessária de potência e contribuindo para a redução do consumo energético global.
- **Propostas de Aproximação dos Coeficientes dos Filtros:** Para cada família de filtros, o artigo apresenta duas versões aproximadas dos coeficientes fixos:
 - *Aproximação 1:* Permite apenas um somador no caminho crítico dos MCMs. Se necessário, utiliza deslocamentos para ajustar o coeficiente, aproximando-o do valor original.
 - *Aproximação 2:* Elimina o uso de somadores no caminho crítico, substituindo-os inteiramente por operações de deslocamento, o que resulta em uma redução adicional no consumo de energia, mesmo que isso leve a uma pequena perda de precisão.
- **Processamento Paralelo e Configuração de Blocos:** A arquitetura é projetada para trabalhar com blocos de 4×4 amostras, utilizando um paralelismo de 12 amostras na entrada. Esse arranjo garante que, além dos pixels essenciais do bloco, os pixels adjacentes (necessários para uma interpolação adequada) também sejam processados, permitindo o escalonamento para blocos maiores e promovendo a operação em tempo real.
- **Síntese e Avaliação de Desempenho:** O acelerador foi descrito em VHDL e sintetizado utilizando a tecnologia de célula padrão TSMC de 40 nm. Durante a avaliação, foram analisados parâmetros como frequência de operação, área (medida em contagem de portas NAND) e dissipação de potência. Essas métricas permitiram comparar os resultados do RVIC com outros trabalhos correlatos, evidenciando ganhos em throughput e eficiência energética.

Principais Resultados:

- **Desempenho em Frequência e Throughput:** O acelerador RVIC alcançou uma frequência máxima de 833,3 MHz. Essa performance permitiu o processamento de vídeos com resolução 4320p a 50 quadros por segundo na etapa de Motion Compensation (MC), correspondendo a um throughput de aproximadamente 2.656,14 milhões de amostras por segundo.
- **Eficiência Energética:** Ao testar nove modos de funcionamento (devido à configuração com filtros precisos e aproximados), os resultados demonstraram uma dissipação de potência variando de 21,25 mW na configuração mais econômica até 40,06 mW na configuração mais dispendiosa. Em comparações específicas, certas versões aproximadas (por exemplo, aproximada 2 para as famílias Regular e Smooth) mostraram reduções de potência significativas, de até 20,69% quando comparadas a outros modelos correlatos.
- **Comparação com Trabalhos Relacionados:**
 - Em relação a DOMANSKI et al. (2019), o RVIC opera com uma frequência quase três vezes maior, embora alguns casos impliquem em uma área um pouco maior.
 - Comparado a DOMANSKI et al. (2021) e a propostas anteriores de FREITAS et al., o RVIC demonstrou maior throughput, melhor eficiência energética e uma abordagem configurável que permite ajustar o desempenho de acordo com os cenários de codificação.

Tabela 1: Resultados de síntese e comparação com trabalhos relacionados.

Trabalhos Correlatos	DOMANSKI (2019)	DOMANSKI (2021)	FREITAS (2022)	FREITAS (2023)	RVIC
Famílias de Filtros	Todos+Bil. Preciso	Todos+Bil. Aprox.	3 famílias Preciso	3 famílias Prec./Aprox.	3 famílias Prec./Aprox.
Tec. (nm)	40	40	40	40	40
Freq. (MHz)	279,9	686,0	1000 833,3	833,3	833,3
Portas (K)	141,10	72,64	324,79	473,01	362,74
Potência (mW)	81,31	26,79	51,15	41,30	21,25 a 40,06
Resol. MC	4320p @30qps	4320p @30qps	4320p @60qps	4320p @50qps	4320p @50qps
Through. (Mam./s)	-	-	3.187,5	2.656,14	2.656,14

Relevância e Impacto Potencial:

O projeto apresentado neste artigo é altamente relevante no contexto do codec AV1, que apesar de oferecer ganhos expressivos de compressão, impõe elevado custo computacional, especialmente na etapa de estimação de movimento fracionária. A proposta de um acelerador de hardware multiversão para os filtros de interpolação

dessa etapa permite não apenas atender aos requisitos de desempenho exigidos pelo padrão (como processar vídeos 8K a 50 fps), mas também oferece flexibilidade de configuração com nove modos distintos, equilibrando precisão e economia de energia conforme as necessidades do cenário de codificação.

O impacto potencial do trabalho reside na significativa redução da dissipação de potência alcançada por meio de técnicas como **reaproveitamento de hardware**, **operand isolation** e uso de **versões aproximadas dos filtros**. A arquitetura RVIC mostrou-se mais eficiente que propostas anteriores, tanto em termos de throughput quanto de consumo energético, o que a posiciona como uma solução viável para aplicações em dispositivos com restrições de energia, como dispositivos móveis ou embarcado. Além disso, ao apresentar a única proposta configurável focada nos filtros FME/MC do AV1, o trabalho contribui para a evolução das arquiteturas de compressão de vídeo de alto desempenho.

Referência

FREITAS, DAIANE FONSECA et al. ACELERADOR COM REAPROVEITAMENTO DE HARDWARE MULTIVERSÃO DE BAIXO CONSUMO PARA OS FILTROS DE INTERPOLAÇÃO DA ESTIMAÇÃO DE MOVIMENTO FRACIONÁRIA DO CODEC AV1.

Artigo 2

Título:

PREVIS – Uma abordagem combinada de aprendizado de máquina e interpolação visual para engenharia reversa interativa no controle de qualidade de montagens

Autores:

Patrick Ruediger, Felix Claus, Viktor Leonhardt, Hans Hagen, Jan C. Aurich, Christoph Garth

Publicação:

Pré-print no arXiv, janeiro de 2022

Objetivo Principal do Artigo

Desenvolver o **PREVIS**, uma ferramenta interativa de engenharia reversa para controle de qualidade em processos de montagem, que combina **modelos de aprendizado de máquina** com **técnicas de interpolação visual**, permitindo interpretar e analisar localmente os efeitos de regressão preditiva em superfícies deformadas.

Em outras palavras, o artigo propõe um sistema que projeta os erros dos modelos de regressão de volta na geometria espacial da peça analisada, possibilitando uma **visualização explicável e interativa** dos resultados de aprendizado de máquina, com **aplicações diretas na indústria automotiva**.

Métodos Utilizados

Interpolação Visual com Ajuste por Mínimos Quadrados (LLSF):

A base da interpolação visual é uma reconstrução linear baseada em **Análise de Componentes Principais (PCA)**. O modelo interpola novas geometrias por uma combinação linear de componentes principais, ajustando os coeficientes por mínimos quadrados.

Simulações de Elementos Finitos (FEM):

Dados de treinamento são gerados por simulações FEM de deformações estruturais em montagens de peças, representando deformações físicas reais que ocorrem em processos industriais.

Modelos de Aprendizado de Máquina Supervisionado:

- **OLFF (One-Layer Feedforward Network):** Rede neural simples com uma única camada densa, usada para prever os coeficientes de interpolação PCA a partir dos parâmetros de montagem.
- **GCN (Graph Convolutional Network):** Rede neural especializada em estruturas em grafo, utilizada para lidar com malhas tridimensionais irregulares das peças simuladas, promovendo uma melhor generalização topológica.

Integração em Ferramenta Interativa (PREVIS):

A ferramenta permite a **visualização em tempo real** da deformação prevista de uma peça com base nas variáveis de montagem, mapeando os erros do modelo de regressão diretamente sobre a malha da peça tridimensional.

Principais Resultados e Conclusões

Capacidade de Mapeamento Espacial:

A técnica de interpolação permite traduzir diretamente os erros de predição (regressão) para o espaço da geometria da peça, revelando onde estão os maiores desvios nas deformações simuladas.

Comparação entre Modelos de Regressão:

- O modelo **OLFF** apresentou **melhor desempenho médio** em termos de erro quadrático médio (RMSE).
- O modelo **GCN** foi **mais estável** e confiável ao longo de todo o conjunto de dados, especialmente em estruturas topologicamente mais complexas.

Desempenho Computacional:

- A ferramenta PREVIS opera com **tempo de execução inferior a 0,05s por interpolação**, permitindo sua adoção em ambientes industriais onde são exigidas **respostas rápidas em tempo real**.

Efetividade Visual:

- A interpolação espacial ajuda engenheiros a **compreender visualmente** o impacto de erros de predição em deformações físicas, o que melhora a **confiança no modelo de ML** e facilita diagnósticos e ajustes nos processos de montagem.

Relevância e Impacto Potencial

O PREVIS apresenta um avanço relevante na área de **engenharia assistida por computador**, ao unir **interpretação de modelos de aprendizado de máquina com visualização espacial**, focando em **manufatura inteligente** e **controle de qualidade industrial**. Sua capacidade de fornecer feedback visual imediato a partir de dados preditivos o torna útil para:

- **Indústria automotiva**, como demonstrado na montagem de capôs veiculares;
- **Simulações mecânicas e ajustes finos de engenharia reversa**;
- **Educação e treinamento** em ambientes industriais complexos;
- **Análise de confiabilidade e segurança** em componentes montados.

O potencial de **generalização para outras indústrias**, aliado ao suporte a simulações FEM e dados tridimensionais, posiciona o PREVIS como uma ferramenta versátil e

inovadora, integrando **inteligência artificial, visualização científica e engenharia computacional**.

Referência :

RUEDIGER, Patrick; CLAUS, Felix; LEONHARDT, Viktor; HAGEN, Hans; AURICH, Jan C.; GARTH, Christoph. *PREVIS – A Combined Machine Learning and Visual Interpolation Approach for Interactive Reverse Engineering in Assembly Quality Control*. arXiv preprint, 2022. Disponível em: <https://arxiv.org/abs/2201.10257v1>. Acesso em: 17 abr. 2025.

Artigo 3

Título:

Iterative methods in GPU-resident linear solvers for nonlinear constrained optimization

Autores:

Świrydowicz, Kasia; Koukpaizan, Nicholson; Alam, Maksudul; Regev, Shaked; Saunders, Michael; Peleš, Slaven

Periódico:

Parallel Computing, vol. 123, artigo 103123, 2025

Objetivo Principal do Artigo

Desenvolver métodos iterativos eficientes para resolver sistemas lineares mal-condicionados, oriundos de problemas de otimização não linear com restrições, utilizando **solvers residentes em GPU**, a fim de superar as limitações de desempenho dos métodos tradicionais em arquiteturas heterogêneas (CPU + GPU).

Em outras palavras, o artigo propõe estratégias híbridas que combinam **refatoração simbólica na CPU, fatoração numérica na GPU e refinamento iterativo com FGMRES**, com o objetivo de melhorar a eficiência computacional e a escalabilidade em aplicações como o **fluxo de potência ótimo AC (ACOPF)** em redes elétricas de grande porte.

Métodos Utilizados

Refinamento Iterativo com FGMRES:

A solução inicial do sistema linear é obtida com LU ou LDL^T e depois refinada usando o método **FGMRES** com pré-condicionamento, que melhora a precisão sem exigir nova fatoração completa. O refinamento é executado diretamente na GPU, usando tolerâncias adaptativas e critérios de parada baseados em normas residuais relativas.

Refatoração Reutilizável (Refactorization):

É realizada apenas uma **fatoração simbólica inicial** na CPU, e depois apenas **fatorações numéricas** são feitas na GPU para os próximos sistemas, aproveitando que a estrutura de esparsidade dos sistemas se mantém. Isso reduz drasticamente o custo computacional para sistemas sucessivos.

Comparação entre Solvers:

O desempenho dos solvers propostos é comparado com:

- **MA57** (solver de referência baseado em CPU);
- **cusolverRf** e **cusolverGLU** (solvers da NVIDIA com e sem refinamento embutido);
- **HyKKT**, um solver híbrido que utiliza **Cholesky + CG** adaptado para GPUs e transforma os sistemas KKT em formas simétricas definidas positivas (SPD).

Implementação em CUDA:

O código é desenvolvido em C++/CUDA, utilizando bibliotecas otimizadas da NVIDIA (cuBLAS, cuSPARSE e cuSOLVER), e avaliado em supercomputadores como o **Summit** do Oak Ridge National Lab.

Principais Resultados e Conclusões

Melhora no Desempenho Computacional:

- O uso de refinamento iterativo com **FGMRES** permitiu reduções de até **40% no tempo total de execução** em relação ao solver MA57 (CPU).
- A estratégia híbrida **cusolverRf + FGMRES** apresentou o **menor tempo de convergência e menor número de passos da otimização**, especialmente em redes grandes como a **Eastern + Western U.S. Grid**.

Qualidade Numérica dos Resultados:

- O refinamento iterativo reduziu o erro normativo (NSR) das soluções em **duas ordens de magnitude** comparado ao uso de refatoração sem refinamento.

- Em grandes redes (ex: ACTIVSg70k com mais de 1,6 milhão de variáveis), o refinamento iterativo garantiu **estabilidade e precisão**, mesmo com sistemas altamente mal-condicionados.

Comparativo com HyKKT:

- O solver HyKKT apresentou desempenho semelhante ao GLU, mas com **custo de configuração 4× menor** e **maior escalabilidade** em sistemas grandes.

Relevância e Impacto Potencial

O artigo apresenta avanços significativos para o campo da **computação científica de alto desempenho (HPC)**, especialmente na área de **otimização de sistemas elétricos**. A proposta de solvers híbridos e refinamento iterativo em GPU permite:

- Reduzir drasticamente os tempos de simulação e planejamento de redes elétricas;
- Tornar viável o uso de **métodos internos (interior-point)** em tempo real em dispositivos embarcados, como subestações inteligentes;
- Aumentar a precisão e a estabilidade de soluções para problemas de grande escala, como simulação de energia, veículos autônomos, aeronaves e sistemas logísticos.

Além disso, os autores destacam que os dados e ferramentas são abertos, promovendo **reprodutibilidade científica** e permitindo que outros pesquisadores validem e aprimorem os métodos apresentados.

Referência:

ŚWIRYDOWICZ, Kasia; KOUKPAIZAN, Nicholson; ALAM, Maksudul; REGEV, Shaked; SAUNDERS, Michael; PELEŠ, Slaven. *Iterative methods in GPU-resident linear solvers for nonlinear constrained optimization*. Parallel Computing, v. 123, p. 103123, 2025. Disponível em: <https://doi.org/10.1016/j.parco.2024.103123>. Acesso em: 17 abr. 2025.

Artigo 4

Título:

Abordagem de aprendizado profundo para extração de quadros significativos em volumes de tomografia computadorizada

Autor:

Lucas Almeida da Silva

Instituição / Tipo de Publicação:

Universidade Federal do Amazonas (UFAM) – Dissertação de Mestrado
Programa de Pós-Graduação em Informática (PPGI), 2023

Objetivo Principal do Artigo

Propor um método automatizado e não supervisionado, baseado em **Grad-CAM**, para **selecionar quadros (slices) significativos em volumes de tomografia computadorizada (TC)**. O objetivo é reduzir o custo computacional e melhorar a eficiência de modelos de aprendizado profundo ao selecionar apenas os quadros mais relevantes para análise clínica.

Em outras palavras, o trabalho busca explorar como o conhecimento adquirido por modelos convolucionais 3D pode ser interpretado e utilizado para identificar regiões mais informativas nos volumes de imagem médica, com foco em aplicações no diagnóstico de sequelas causadas por Covid-19.

Métodos Utilizados

1. Grad-CAM Slice Selection (GSS):

A técnica proposta utiliza mapas de ativação da última camada convolucional de uma **rede neural 3D convolucional (CNN 3D)** para estimar a importância de cada quadro do volume. Através da ativação média por slice, o método ranqueia os quadros com maior contribuição na decisão do modelo.

2. Rede Neural Profunda CNN3D Customizada:

Uma rede neural tridimensional foi projetada para classificar volumes de tomografia de pulmão quanto à presença de sequelas por Covid-19. A arquitetura foi adaptada para trabalhar com volumes reduzidos em profundidade, possibilitando simulações realistas mesmo com restrições de hardware.

3. Estudo de Caso com Base MosMed:

O método foi aplicado à base **MosMed**, que contém tomografias rotuladas de pacientes com e sem sequelas pulmonares causadas pela Covid-19. Essa base permitiu a validação clínica e experimental da abordagem.

4. Comparação com Técnicas Clássicas de Seleção de Slices:

A GSS foi comparada a métodos tradicionais como **Even Slice Selection (ESS)**, **Subset Slice Selection (SSS)** e técnicas de interpolação (como **Spline Interpolated Zoom - SIZ**), tanto em termos de desempenho computacional quanto de acurácia do modelo.

Principais Resultados e Conclusões

- O método **GSS superou todos os demais** em termos de **AUC (área sob a curva ROC)** e **F1 Score**, garantindo melhor desempenho classificatório com um número significativamente reduzido de quadros.
- A redução da profundidade de entrada do volume permitiu **economia computacional** sem perda de desempenho — fator essencial para aplicações clínicas em ambientes com recursos limitados.
- As ativações geradas pelo Grad-CAM possibilitaram **explicabilidade** nos modelos de classificação, mostrando em quais regiões pulmonares o modelo concentrou suas atenções durante a inferência.
- O GSS é **independente de supervisão humana**, eliminando a necessidade de pré-seleção manual de imagens por especialistas.

Relevância e Impacto Potencial

Este trabalho oferece uma solução inovadora e escalável para problemas reais de **diagnóstico por imagem médica**, especialmente em situações em que o volume de dados dificulta a análise convencional, como no caso da pandemia de Covid-19.

A contribuição mais notável reside no uso da técnica Grad-CAM não apenas como interpretabilidade, mas como mecanismo **ativo de seleção de dados**. Isso pode revolucionar o pré-processamento em pipelines de aprendizado profundo, otimizando o uso de recursos computacionais e melhorando a precisão clínica.

O método é aplicável a **outras modalidades médicas tridimensionais** (ex: ressonância magnética) e tem potencial de impacto em áreas como **radiologia**, **pneumologia**, **medicina de precisão** e **sistemas de apoio à decisão médica**.

Referência:

SILVA, Lucas Almeida da. Abordagem de aprendizado profundo para extração de quadros significativos em volumes de tomografia computadorizada. 2023.