

# GRADES COMPUTACIONAIS NA OTIMIZAÇÃO DA RECUPERAÇÃO DE IMAGENS MÉDICAS BASEADA EM CONTEÚDO\*

Marcelo Costa Oliveira<sup>1</sup>, Paulo Mazzoncini de Azevedo-Marques<sup>2</sup>, Walfredo da Costa Cirne Filho<sup>3</sup>

**Resumo** **OBJETIVO:** Utilizar o poder de processamento da tecnologia de grades computacionais para viabilizar a utilização do algoritmo de medida de similaridade na recuperação de imagens baseada em conteúdo. **MATERIAIS E MÉTODOS:** A técnica de recuperação de imagens baseada em conteúdo é composta de duas etapas sequenciais: análise de textura e algoritmo de medida de similaridade. Estas são aplicadas em imagens de joelho e cabeça, nas quais se avaliaram a eficiência em recuperar imagens do mesmo plano e a sequência de aquisição em um banco de 2.400 imagens médicas para testar a capacidade de recuperação de imagens baseada em conteúdo. A análise de textura foi utilizada inicialmente para pré-selecionar as 1.000 imagens mais semelhantes a uma imagem de referência escolhida por um clínico. Essas 1.000 imagens foram processadas utilizando-se o algoritmo de medida de similaridade na grade computacional. **RESULTADOS:** A precisão encontrada na classificação por análise de textura foi de 0,54 para imagens sagitais de joelho e de 0,40 para imagens axiais de cabeça. A análise de textura foi útil como filtragem, pré-selecionando imagens a serem avaliadas pelo algoritmo de medida de similaridade. A recuperação de imagens baseada em conteúdo utilizando o algoritmo de medida de similaridade aplicado nas imagens pré-selecionadas por análise de textura resultou em precisão de 0,95 para as imagens sagitais de joelho e de 0,92 para as imagens axiais de cabeça. O alto custo computacional do algoritmo de medida de similaridade foi amortizado pela grade computacional. **CONCLUSÃO:** A utilização da abordagem mista das técnicas de análise de textura e algoritmo de medida de similaridade no processo de recuperação de imagens baseada em conteúdo resultou em eficiência acima de 90%. A grade computacional é indispensável para utilização do algoritmo de medida de similaridade na recuperação de imagens baseada em conteúdo, que de outra forma seria limitado a supercomputadores. *Unitermos:* Recuperação de imagens baseada em conteúdo; Análise de textura; Registro de imagens; Grades computacionais.

**Abstract** *Grid computing in the optimization of content-based medical images retrieval.*

**OBJECTIVE:** To utilize the grid computing technology to enable the utilization of a similarity measurement algorithm for content-based medical image retrieval. **MATERIALS AND METHODS:** The content-based images retrieval technique is comprised of two sequential steps: texture analysis and similarity measurement algorithm. These steps have been adopted for head and knee images for evaluation of accuracy in the retrieval of images of a single plane and acquisition sequence in a databank with 2,400 medical images. Initially, texture analysis was utilized as a pre-selection resource to obtain a set of the 1,000 most similar images as compared with a reference image selected by a clinician. Then, these 1,000 images were processed utilizing a similarity measurement algorithm on a computational grid. **RESULTS:** The texture analysis has demonstrated low accuracy for sagittal knee images (0.54) and axial head images (0.40). Nevertheless, this technique has shown effectiveness as a filter, pre-selecting images to be evaluated by the similarity measurement algorithm. Content-based images retrieval with similarity measurement algorithm applied on these pre-selected images has demonstrated satisfactory accuracy — 0.95 for sagittal knee images, and 0.92 for axial head images. The high computational cost of the similarity measurement algorithm was balanced by the utilization of grid computing. **CONCLUSION:** The approach combining texture analysis and similarity measurement algorithm for content-based images retrieval resulted in an accuracy of > 90%. Grid computing has shown to be essential for the utilization of similarity measurement algorithm in the content-based images retrieval that otherwise would be limited to supercomputers.

*Keywords:* Content-based image retrieval; Texture analysis; Image registration; Grid computing.

\* Trabalho realizado no Laboratório de Sistemas Distribuídos do Departamento de Sistemas e Computação da Universidade Federal de Campina Grande (UFCG), Campina Grande, PB, Brasil.

1. Doutor em Medicina na área de Informática Médica, Mestre em Física Aplicada à Medicina, Bacharel em Ciência da Computação, Docente da Universidade Federal de Alagoas, Campus Arapiraca, Arapiraca, AL, Brasil.

2. Engenheiro Eletrônico, Doutor em Física Aplicada, Docente do Centro de Ciências das Imagens e Física Médica da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (CCIFM/FMRP-USP), Ribeirão Preto, SP, Brasil.

3. Doutor em Ciência da Computação, Docente do Departamento de Sistemas e Computação, Coordenador do Laboratório de Sistemas Distribuídos da Universidade Federal de Campina Grande (UFCG), Campina Grande, PB, Brasil.

Endereço para correspondência: Dr. Marcelo Costa Oliveira, Universidade Federal de Alagoas, Campus Arapiraca, Coordenação de Ciência da Computação, Rodovia AL 115, km 6,5, Sementeira, Arapiraca, AL, Brasil. E-mail: oliveiramc@gmail.com

Recebido para publicação em 21/11/2005. Aceito, após revisão, em 23/10/2006.

## INTRODUÇÃO

O volume de dados produzidos nos hospitais e centros médicos está em rápida ascensão. A produção anual de dados gerados por exames de imagens nos grandes centros de radiologia é da ordem de 2 terabytes. Isto é devido à obtenção e armaze-

namento de dados dos pacientes, resultantes da crescente importância e utilização dos exames de imagem. É enorme a quantidade de dados que necessitam ser armazenados e indexados de forma inteligente e segura, pois são peças fundamentais no diagnóstico clínico<sup>(1,2)</sup>.

O crescente uso de programas de computador para auxílio diagnóstico (*computer-aided diagnosis* – CAD) está relacionado ao rápido desenvolvimento de algoritmos computacionais aplicados à medicina. O objetivo do CAD é melhorar a acurácia diagnóstica, assim como aprimorar a consistência na interpretação de imagens diagnósticas, mediante o uso da sugestão de resposta diagnóstica fornecida por algum computador<sup>(3)</sup>. Porém, algumas ferramentas CAD que apresentam ótimos resultados ainda não são utilizadas na rotina clínica por apresentarem alto custo computacional, limitando seu uso a centros possuidores de computadores de alta capacidade<sup>(4)</sup>.

As dificuldades em aplicar estes algoritmos CAD na rotina clínica e as limitações ainda existentes para o armazenamento, processamento, busca e recuperação de imagens em grandes bases de dados vem motivando empresas e instituições de pesquisas a encontrarem novas soluções para essas tarefas<sup>(5,6)</sup>.

### Grade computacional

A tecnologia de grade computacional (GC) representa a mais recente e promissora ferramenta na área de computação distribuída. De maneira sucinta, a computação distribuída é uma coleção de computadores independentes que se apresenta ao usuário como um sistema único e consistente. GC permitem a integração de computadores que estão amplamente distribuídos e conectados por meio de redes de longas distâncias. Com esta capacidade de integração de computadores, cria-se um sistema de computação virtual ou cooperativo para solucionar problemas referentes ao armazenamento e acesso de dados em massa e ao processamento de aplicações com alto custo computacional<sup>(6,7)</sup>.

A tecnologia de GC oferece um ambiente único para compartilhar, armazenar e processar dados. Ainda, permite à comunidade médica utilizar uma única base de dados distribuída, capaz de prover o compartilha-

mento de recursos, dados e conhecimento. Todos estes recursos permitem maior interação entre os centros médicos, inclusive oferecendo novas oportunidades a pequenas clínicas e laboratórios de pesquisa com poucos recursos computacionais<sup>(7)</sup>.

Além disso, a GC oferece uma infraestrutura flexível, escalonável, menos sujeita a catástrofes e, portanto, mais confiável, sendo capaz de garantir acesso seguro a qualquer aplicação que a utilize. Diferentemente da computação distribuída e da metodologia de grupos de terminais ou de estações de trabalho ligadas a uma unidade de controle servidora ou grupo de unidades servidoras (*clusters technology*) computacionais, os recursos de GC possuem autonomia administrativa e heterogeneidade de sistema. Estas duas características permitem melhor escalonamento e robustez às aplicações. Entretanto, as mesmas características exigem que os componentes da GC devam seguir padrões visando ao escalonamento e compartilhamento de recursos computacionais de forma aberta<sup>(8)</sup>.

Foster e Kesselman<sup>(9)</sup> apresentam proposta de arquitetura e de componentes de uma GC. A arquitetura formal é composta por quatro camadas (Figura 1). A camada de construção é o nível mais baixo e representa os recursos físicos e dispositivos que os usuários da GC desejam compartilhar e acessar (computadores, rede, sistema de arquivos, catálogos, programas computa-

cionais e instrumentos digitais). Logo acima da camada de construção está a camada de recursos e conectividade, responsável pela comunicação e autenticação necessárias para a troca de recursos, validação do usuário, monitoramento e controle sobre as operações de compartilhamento de recursos. A terceira camada ou camada de cooperação possui os protocolos e realiza os serviços responsáveis pelas transações de recursos (descoberta e alocação de recursos, monitoramento e diagnóstico da funcionalidade dos serviços, replicação de dados e políticas que regulamentam privilégios de usuários no acesso aos recursos da GC). A camada de aplicação do usuário é o mais alto nível e é responsável por invocar todas as outras camadas.

Existe grande número de projetos relacionados à GC descritos na literatura (por exemplo: Globus<sup>(10)</sup>, Legion<sup>(11)</sup>, Condor<sup>(12)</sup> e OurGrid<sup>(13)</sup>) que utilizam diversas tecnologias e são em grande maioria direcionados a determinadas áreas e propósitos, como aplicações de processamento e armazenamento de dados, portais Web e serviços de infra-estrutura para colaboração entre instituições<sup>(14,15)</sup>.

A utilização básica da GC pelo usuário é realizada por meio de uma interface de *software* que permite ao computador do usuário se comunicar com o centro de processamento da GC, conhecido como *broker*. Este é capaz de encontrar os recur-



Figura 1. Arquitetura de grades computacionais.

so necessários à execução das tarefas. Ao término da execução da tarefa, o *broker* retorna o resultado da aplicação ao usuário<sup>(16)</sup>. A Figura 2 ilustra o funcionamento básico do projeto OurGrid.

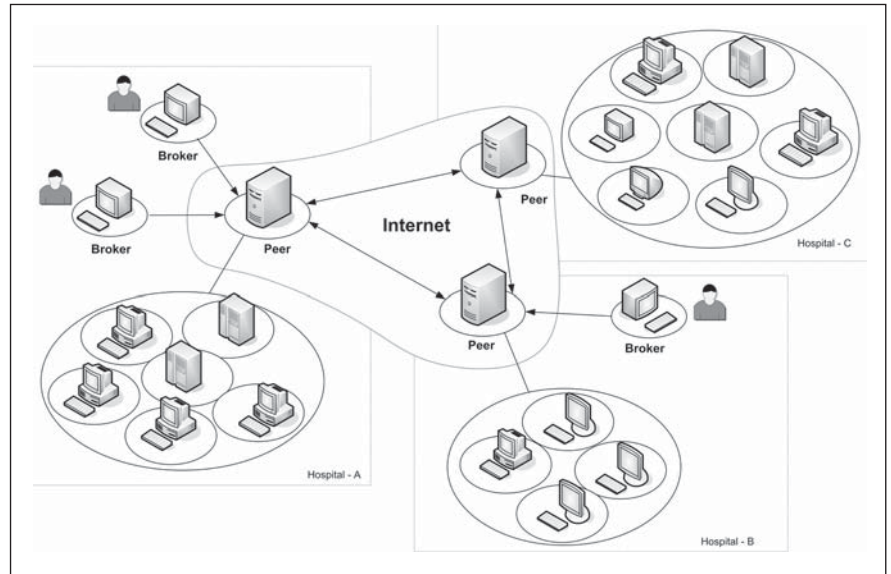
### Recuperação de imagens médicas baseada em conteúdo

Dentre as várias técnicas de CAD, os sistemas de recuperação de imagens médicas baseada em conteúdo (RIBC) são os grandes beneficiados da tecnologia de GC devido às suas características e necessidades: intensidade e complexidade de processamento e grande volume de imagens armazenadas<sup>(17)</sup>.

Por intermédio da RIBC é possível, a partir de uma imagem de referência, encontrar imagens similares contidas em uma ou diversas bases de imagens utilizando atributos inerentes das imagens. No processo de decisões clínicas, a RIBC oferece grandes benefícios, sendo capaz de recuperar imagens de mesma modalidade, região anatômica e alterações estruturais provocadas por determinadas doenças. Assim, a RIBC despertou rápido interesse na comunidade médica, em virtude da sua capacidade de recuperar imagens já diagnosticadas para comparar com uma imagem em estudo, permitindo ao especialista confirmar a sua hipótese diagnóstica<sup>(18)</sup>. Embora parte dessa informação possa estar contida no cabeçalho da imagem médica, esta rotulação textual pode apresentar alta taxa de erro, tendo sido relatados casos de até 16%<sup>(19)</sup>. É amplamente enfatizada nos trabalhos científicos a necessidade de adotar métodos de acesso alternativo aos dados cadastros manualmente inseridos no cabeçalho das imagens médicas<sup>(20-23)</sup>.

Além das técnicas de suporte à decisão clínica, pesquisa e ensino também são grandes beneficiários dos sistemas RIBC. No ensino, a RIBC auxilia os docentes e discentes na utilização de repositórios de imagens educacionais e na análise visual dos resultados encontrados. Além dos estudos baseados no diagnóstico e região anatômica, a análise de casos visualmente similares, porém com diferentes diagnósticos, resulta no aprimoramento da qualidade educacional<sup>(24)</sup>.

RIBC é uma das técnicas de visão computacional mais estudadas nos últimos dez



**Figura 2.** O usuário tem acesso à GC a partir do *broker* instalado em seu computador. O *broker* é capaz de descobrir recursos e requerer computadores locais ou remotos para realizar a execução das tarefas do usuário. O *peer* é responsável pelo gerenciamento das máquinas da rede local e pela negociação de máquinas remotas.

anos e baseia-se nas características visuais da imagem, que são agrupadas em três classes: cor, textura e forma<sup>(25)</sup>. Estes atributos possibilitam o desenvolvimento de ferramentas computacionais robustas e capazes de caracterizar imagens pelo seu próprio conteúdo, adicionando vantagens à identificação de imagens baseadas somente em descritores textuais que formam a classificação tradicional de arquivos de imagens médicas<sup>(23)</sup>.

A distribuição dos níveis de cinza é a característica mais simples de ser caracterizada. Sua caracterização é realizada pela comparação entre histogramas de níveis de cinza utilizando o somatório de diferenças absolutas ou quadráticas sobre o número de elementos da imagem (*pixels*) de cada intensidade de cinza. Por apresentarem ambigüidade (em que diferentes imagens podem gerar o mesmo somatório), não são eficazes para o todo da RIBC, mas por sua simplicidade e baixo custo computacional podem e devem ser utilizadas como filtros primários para outros métodos mais elaborados e de alto custo.

Características baseadas em textura buscam quantificar a variação e a escala de intensidade da imagem. Um dos métodos mais utilizados na literatura para extrair atributos de textura é a matriz de co-ocorrência<sup>(24)</sup>. Haralick et al.<sup>(26)</sup> definiram os

atributos de textura que podem ser obtidos a partir da matriz de co-ocorrência com propósitos de discriminação de textura. Aproximadamente 20 funções estatísticas são propostas pela literatura para obtenção de informação a partir da matriz de co-ocorrência<sup>(27)</sup>. Algumas das funções mais importantes e que produzem satisfatória classificação das texturas são entropia, inércia, energia, matiz, momento da diferença inverso, proeminência, correlação e variância<sup>(27-33)</sup>.

A recuperação de imagens baseada em forma é um dos problemas mais difíceis de serem tratados pelos sistemas de RIBC. Isto é devido à complexidade em realizar segmentação automática de imagens médicas. Após serem segmentadas, as estruturas são descritas pelas características de forma, incluindo informações sobre rotação, translação e escala<sup>(34)</sup>.

Outra técnica de RIBC descrita na literatura é o registro de imagens<sup>(5,17)</sup>. Esta técnica calcula uma transformação rígida bidimensional de coordenadas englobando rotação, translação e escala, buscando o máximo “casamento” (*matching*) entre duas imagens ou entre dois volumes de imagens. A transformação rígida é baseada na minimização do erro quadrático ou soma das diferenças quadradas entre contornos de estruturas utilizando algoritmos

de medida de similaridade entre intensidades de duas imagens<sup>(35,36)</sup>.

Nesse contexto, este trabalho apresenta uma abordagem singular em sistemas de recuperação de imagens médicas baseada em conteúdo, utilizando atributos de textura e o poder computacional da recente tecnologia de GC aplicada ao algoritmo de medida de similaridade baseado na soma das diferenças quadradas.

## MATERIAIS E MÉTODOS

O sistema desenvolvido no presente trabalho utilizou o sistema operacional GNU/Linux Debian e linguagem de programação Java 1.5, sendo o algoritmo de medida de similaridade (AMS) uma implementação do *software* Insight Toolkit<sup>(37)</sup>. Para sua avaliação foi utilizada uma base heterogênea de imagens composta de 2.400 imagens de ressonância magnética de diversas regiões anatômicas, seqüências e planos de aquisição, e níveis de cinza variando entre 4.096 e 65.536.

O sistema apresenta dois módulos de RIBC. O primeiro módulo utiliza análise de textura de segunda ordem (matriz de co-ocorrência) para realizar a classificação das imagens mais semelhantes de acordo com esta técnica. No segundo módulo é aplicado o AMS sobre as imagens selecionadas no primeiro módulo. Devido ao alto custo computacional do AMS, o segundo módulo é processado na GC do OurGrid. Este é uma GC cooperativa, aberta e de livre acesso, em que centros compartilham seus recursos computacionais ociosos e em reciprocidade, quando necessário, recebem recursos ociosos de outros centros. O OurGrid, atualmente, é composto por aproximadamente 500 máquinas.

A interface do usuário com a GC é realizada pelo MyGrid 3.2 (OurGrid; Campina Grande, PB), que é o *broker* do OurGrid. Este é capaz de realizar a escolha dos recursos computacionais que vão ser utilizados na execução da aplicação, além de abstrair toda a complexidade da GC ao usuário; assim, o usuário utiliza a GC como se fosse um único computador<sup>(13)</sup>.

Todas as imagens da base de dados possuem um vetor de característica associado, que foi obtido pela matriz de co-ocorrência de níveis de cinza e seus atributos. A

matriz de co-ocorrência seguiu as orientações em 0°, 45°, 90° e 135° e distância entre os elementos de imagens (*pixels*) igual a 1. Os atributos de textura utilizados na aplicação foram energia, entropia, momento da diferença inversa, penumbra, inércia, *promenace*, correlação e variância. A utilização de oito atributos de textura e quatro orientações angulares propiciou a criação de um vetor de característica de 32 dimensões.

O sistema desenvolvido oferece uma interface gráfica (Figura 3) que permite ao especialista selecionar uma imagem de referência no padrão DICOM (*digital imaging and communication in medicine*), iniciando o primeiro módulo. Ao término do módulo as imagens são classificadas segundo o menor valor da distância euclidiana entre os vetores de características da imagem de referência e das imagens da base de dados.

O segundo módulo utiliza as 1.000 imagens mais similares segundo o primeiro módulo. Este módulo também necessita que o especialista defina em quantas tarefas o processamento do AMS será decomposto e distribuído na GC. Ou seja, qual a “granularidade” da aplicação. A granularidade está relacionada ao volume de imagens que será processado pelo AMS em cada computador da GC. O AMS utiliza transformações afins e interpolação linear buscando o mapeamento dos pontos homólogos entre duas imagens.

## RESULTADOS

Os resultados deste trabalho foram obtidos pela seleção de duas regiões anatômicas — joelho e cabeça — pertencentes à base de imagens. As imagens de joelho são ponderadas em T1 com plano de corte sagital e as imagens de cabeça são ponderadas em T2 e plano de corte axial. A base de dados possui um exame com 20 imagens de joelho no plano sagital e um exame com 40 imagens de cabeça no plano axial. Os experimentos foram repetidos três vezes escolhendo cortes diferentes dos exames descritos. As imagens foram classificadas como corretas quando a aplicação retornava imagens do mesmo plano e seqüência de aquisição das imagens de referência.

O primeiro módulo classificou as imagens mais semelhantes segundo os atributos de textura. A média do tempo de processamento do primeiro módulo foi de 2,3 minutos, obtida pelo cálculo da distância euclidiana entre o vetor de característica de cada uma das 2.400 imagens da base de dados em relação ao vetor de característica da imagem de referência. Os algoritmos processados no computador local utilizaram processador Pentium 4 de 2.8 GHz com 1 Gbyte de memória.

Para avaliarmos os resultados obtidos utilizamos *precision* (precisão) e *recall* (revocação), que são parâmetros normalmente utilizados para avaliar sistemas de recuperação de imagens baseada em conteúdo e

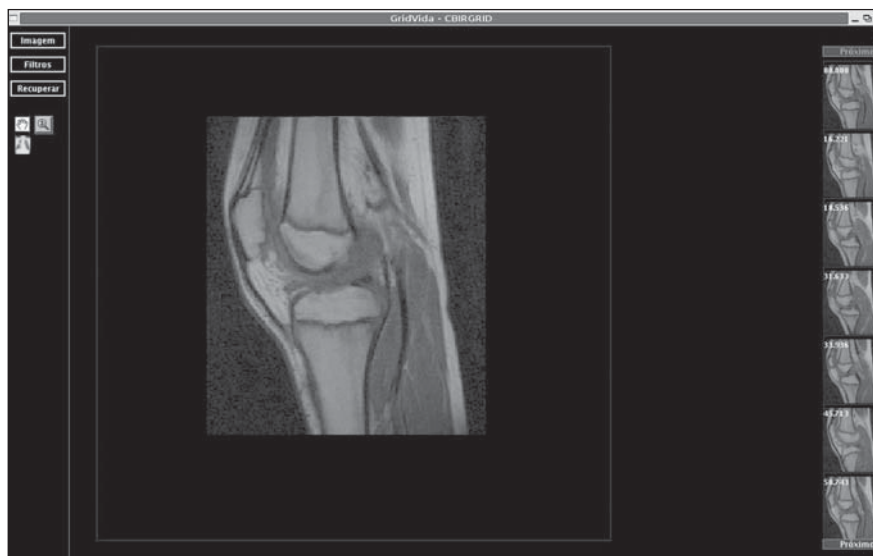


Figura 3. Interface gráfica para interação do usuário com o sistema.



recuperação de informação. Revocação indica a proporção de imagens relevantes na base de dados que foram recuperadas ao responder uma consulta. Precisão é a proporção das imagens recuperadas que são relevantes para a consulta<sup>(38)</sup>.

A Figura 4 ilustra as médias obtidas pela execução do primeiro módulo das curvas de revocação  $\times$  precisão da distância euclidiana entre os vetores de características das imagens de referência em relação às imagens da base de dados. Este resultado permitiu avaliar a eficácia da RIBC utilizando

textura na classificação das imagens mais semelhantes para o segundo módulo. Embora a média da precisão obtida nos experimentos seja de 0,54 (joelho sagital) e 0,40 (cabeça axial), ela é suficiente para filtrar as imagens que serão submetidas ao segundo módulo. No segundo módulo as imagens são processadas pelo AMS utilizando o processamento da GC. A RIBC utilizando o AMS produziu satisfatória precisão em ambos os casos estudados — 0,95 (joelho sagital) e 0,92 (cabeça axial) —, seguindo as médias das curvas de revocação

$\times$  precisão entre as imagens de referência e as classificadas pelo primeiro módulo (Figura 5).

A Figura 6 ilustra a classificação das imagens mais semelhantes após a execução da aplicação. Devido a questões de espaço, somente as nove imagens mais semelhantes são ilustradas.

O alto custo computacional produzido pelo AMS foi amortizado pelo sistema de GC do OurGrid. Na média, o tempo de processamento do AMS aplicado aos experimentos utilizando 50 processadores da

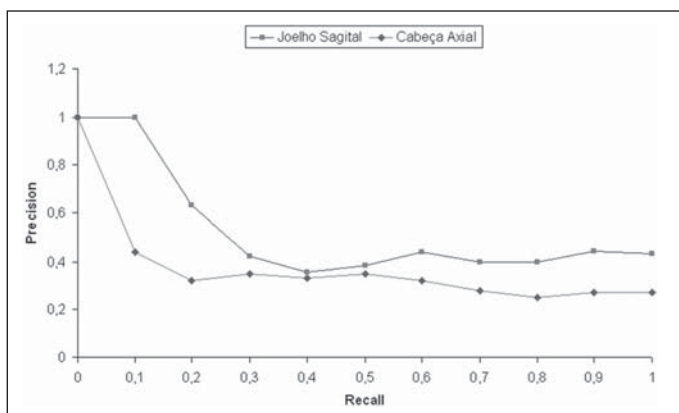


Figura 4. Curvas referentes à execução do primeiro módulo utilizado na filtragem das imagens.

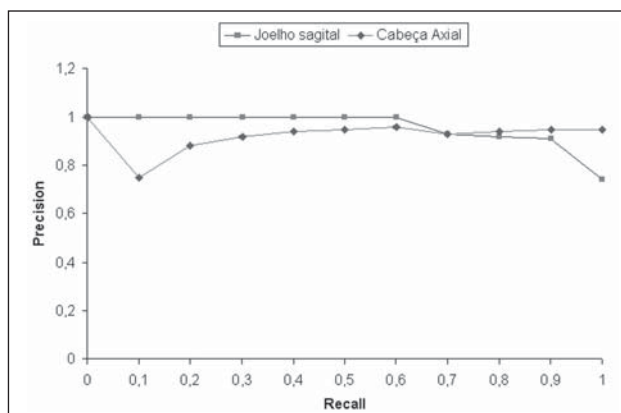


Figura 5. Curvas referentes à eficiência da aplicação em recuperar imagens baseada em conteúdo.

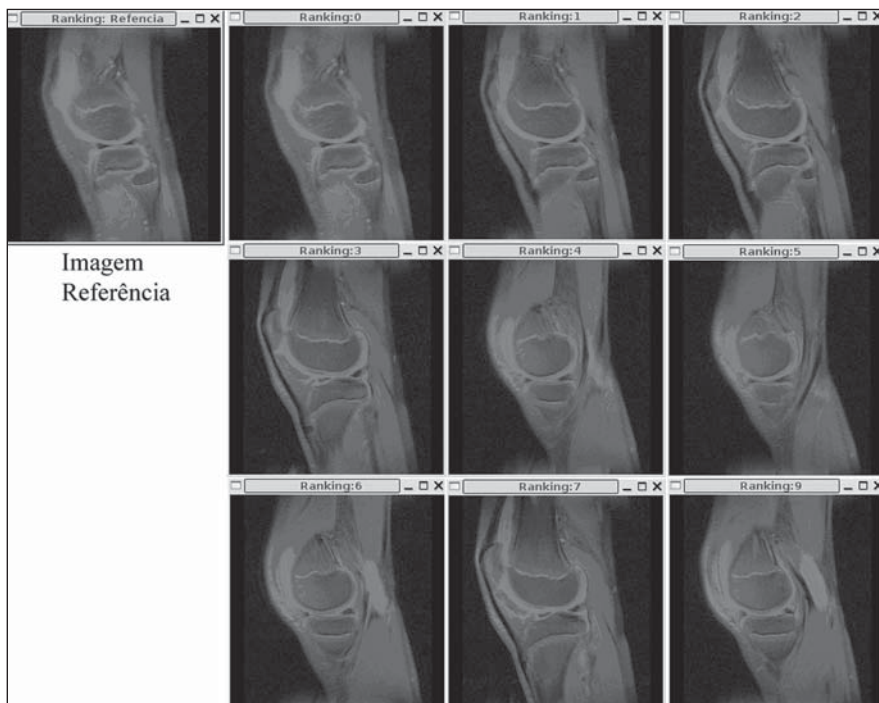


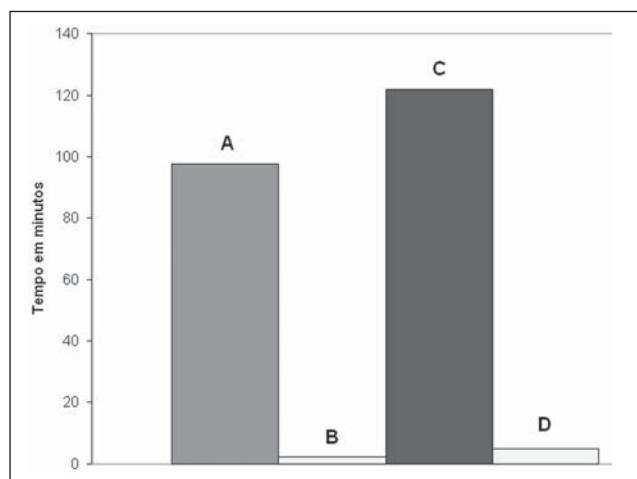
Figura 6. Resultado da aplicação utilizando imagem de joelho no plano sagital e as imagens mais semelhantes obtidas pela aplicação.

GC foi reduzido em 116,97 minutos para imagens de joelho e em 95,15 minutos para as imagens de cabeça, em relação aos tempos de processamento obtidos na máquina local (Figura 7).

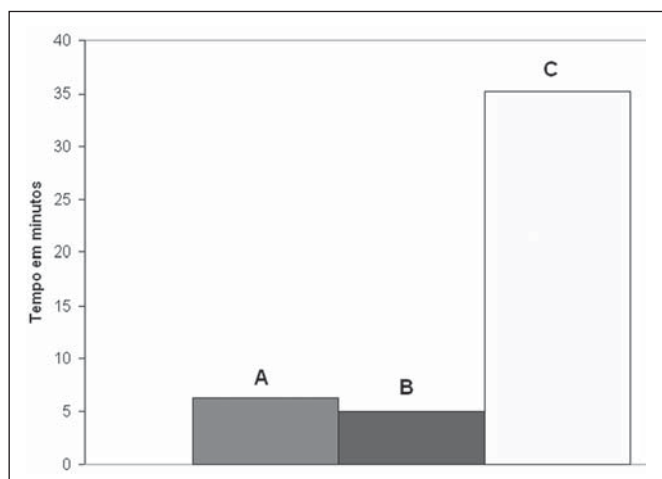
Neste experimento dividimos a aplicação em 20 tarefas composta por 50 imagens cada. As imagens foram compactadas antes de serem enviadas à GC e o tamanho médio dos arquivos com 50 imagens foi de 4 Mbytes. As imagens, ao serem enviadas à GC, possuem um arquivo de identificação único especificando o número da imagem e a qual tarefa ela pertence.

O tempo médio do envio das imagens compactadas para as máquinas da GC foi de 22,2 segundos e o tempo médio de processamento das 50 imagens em cada máquina da GC foi de 11,45 segundos. O baixo tempo do envio das imagens é devido à maioria das tarefas ter sido executada nos computadores contidos na rede local.

O OurGrid permitiu também que as bibliotecas necessárias a execução da aplica-



**Figura 7.** Tempos médios de processamento. **A:** Cabeça axial local. **B:** Cabeça axial grade. **C:** Joelho sagital local. **D:** Joelho sagital grade.



**Figura 8.** Comparação entre os tempos médios utilizando diferentes granularidades no processamento das imagens de joelho sagital, 10 tarefas (A), 20 tarefas (B) e 50 tarefas (C).

ção fossem armazenadas nos computadores remotos, evitando a necessidade de envio de dados.

O tempo médio dos experimentos foi também analisado alterando a granularidade da aplicação entre 10, 20 e 50 imagens por tarefa (Figura 8). A utilização do menor grão, ou seja, 50 imagens por tarefa, implicou uma maior divisão do volume total de imagens por tarefa. Logo, maior número de máquinas da GC foi requisitado, pois houve um aumento da quantidade de tarefas a serem processadas. A necessidade de alocar 50 máquinas da GC para executar a aplicação implicou a distribuição de tarefas para serem processadas fora da rede local. Assim, o tempo total da aplicação foi comprometido pelo tempo de transmissão de dados a computadores remotos.

Contudo, decompor a aplicação em tarefas maiores (dez tarefas no total), ou seja, maior grão, implicou a requisição de menos computadores e na transmissão de grandes arquivos, produzindo maior tempo de processamento executado por máquina. Logo, adotamos neste trabalho um número fixo e intermediário de 20 tarefas.

## DISCUSSÃO

A tecnologia de GC tem-se mostrado ferramenta promissora no processamento e armazenamento de grande volume de dados. Contudo, mais benefícios são esperados desta tecnologia, como o apresentado por Liu et al.<sup>(7)</sup>, que utilizaram a arquitetura

de GC para realizar cópias de segurança de imagens médicas em diversos sistemas PACS (*picture archiving and communication system*) pertencentes à GC.

Este trabalho apresentou uma abordagem mista de técnicas de RIBC para classificar imagens semelhantes de diferentes regiões anatômicas e orientações utilizando o alto poder de processamento de GC. O sistema utilizou técnicas de RIBC baseadas na classificação de textura e medida de similaridade.

A análise de textura apresenta boa aproximação para a percepção visual humana e tem sido utilizada em muitos sistemas de auxílio ao diagnóstico clínico<sup>(39,40)</sup>. O valor médio da precisão da AT foi eficiente atuando como filtro inicial para o segundo módulo, apesar de relativamente baixa — 0,54 para joelho e 0,40 para cabeça. Uma possível solução para aumentar a eficiência dessa filtragem seria o desenvolvimento de métodos para detectar artefatos de movimento, já que informações de textura podem ser perdidas quando rotação, translação e escala são incluídas no espaço de características<sup>(29)</sup>.

A utilização do algoritmo de medida de similaridade da soma das diferenças quadradas aplicado ao segundo módulo apresentou valor médio de precisão bastante satisfatório — 0,95 para joelho e 0,92 para cabeça. O algoritmo foi capaz de recuperar imagens semelhantes de diferentes regiões anatômicas e orientações. A maioria dos artigos da literatura é restrita a deter-

minada região anatômica, modalidade ou procedimento diagnóstico utilizando-se apenas vetores de características<sup>(41)</sup>. No entanto, o custo computacional dos algoritmos de medida de similaridade, quando executados num único computador, é inaplicável ao diagnóstico auxiliado por computador. A GC viabiliza a utilização da técnica de medida por similaridade devido à capacidade de processar tarefas em paralelo nos diversos computadores que compõem a GC.

Apesar de a GC utilizada neste trabalho possuir aproximadamente 500 computadores espalhados em mais de 20 locais, os experimentos com 10 e 20 tarefas foram processados nas máquinas da rede local, não comprometendo o tempo de execução da aplicação. Contudo, os experimentos com 50 tarefas necessitaram ser processados fora da rede local, logo, foram prejudicados pelo custo de transmissão dos dados. Nesses casos devemos avaliar o custo-benefício entre tempo de processamento e envio de dados.

A tecnologia de GC ainda está nos primórdios em aplicações médicas, porém é bastante promissora e espera-se que, por meio de seus recursos, possibilite importantes avanços na área de informática aplicada à saúde.

Para aprimorar os resultados obtidos, dois novos componentes estão em desenvolvimento: a medida de similaridade baseada na correlação cruzada e a segmentação automática de estruturas cerebrais. O

algoritmo da correlação cruzada permitirá a busca em modalidades diferentes, o que é uma limitação do algoritmo de soma das diferenças quadradas. Outra limitação da soma das diferenças quadradas é a alta sensibilidade a pequenas quantidades de *pixels* que possuem grande diferença de intensidade entre duas imagens, como nos casos de injeção de contraste<sup>(35)</sup>. O algoritmo de segmentação automática restringirá a recuperação de imagens a determinadas estruturas, permitindo buscas mais específicas do que as realizadas ao compararmos a imagem completa. A utilização de métodos distintos de forma integrada deverá proporcionar uma melhor discriminação entre as imagens<sup>(42)</sup>.

### Agradecimentos

Ao Projeto GridVida, ao Laboratório de Sistemas Distribuídos da Universidade Federal de Campina Grande (UFCG) e ao Centro de Ciências das Imagens e Física Médica (CCIFM) da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (FMRP-USP).

### REFERÊNCIAS

- Montagnat J, Breton V, Magnin IE. Using technologies to face medical image analysis challenges. Proceedings of the IEEE CCGrid03 2003, Tokyo, Japan.
- Montagnat J, Breton V, Magnin IE. Partitioning medical image databases for content-based queries on a Grid. *Methods Inform Med* 2005;44:154–160.
- Azevedo-Marques PM. Diagnóstico auxiliado por computador na radiologia. *Radiol Bras* 2001;34:285–293.
- HealthGrid, HealthGrid White Paper. [Acessado em: 10/10/2006]. Disponível em: <http://www.healthgrid.org>
- Montagnat J, Bellet F, Benoit-Catin H, et al. Medical images simulation, storage, and processing on the European DataGrid testbed. *J Grid Comput* 2004;2:387–400.
- Breton V, Blanchet C, Legré Y, Maigne L, Montagnat J. Grid technology for biomedical applications. *Lecture Notes in Computed Science* 2005;204–218.
- Liu BJ, Zhou MZ, Documet J. Utilizing data Grid architecture for the backup and recovery of clinical image data. *Comput Med Imaging Graph* 2005;29:95–102.
- Foster I, Kesselman C, Tuecke S. The anatomy of the Grid: enabling scalable virtual organizations. *International Journal of High Performance Computing Applications* 2001;15:200–222.
- Foster I, Kesselman C. *The Grid 2: blueprint for a new computing infrastructure*. San Francisco, CA: Morgan Kaufmann Publishers, 2004.
- Foster I, Kesselman C. Globus: a metacomputing infrastructure toolkit. *International Journal of Supercomputing Applications* 1997;11:115–128.
- Grimshaw AS, Wulf WA. The legion vision of a worldwide virtual computer. *Communications of the ACM* 1997;40:39–45.
- Condor. [Acessado em: 13/9/2006]. Disponível em: <http://www.cs.wisc.edu/condor>
- Cirne W, Brasileiro F, Andrade N, et al. Labs of the World, Unite!!! UFCG/DSC Technical Report 07/2005;1–12.
- de Roure D, Baker M, Jennings NR, Shadbolt N. The evolution of the Grid. In: Berman F, Fox G, Hey AJG, editors. *Grid computing – making the global infrastructure a reality*. New York, NY: Wiley, 2003;65–100.
- Foster I. The Grid: computing without bounds. *Scientific American* April 2003;228:80–85.
- Grid Café. [Acessado em: 29/8/2006]. Disponível em: <http://gridcafe.web.cern.ch/gridcafe>
- Montagnat J, Duque H, Pierson JM, Breton V, Brunie L, Magnin IE. Medical image content-based queries using the Grid. Proceedings of the First European HealthGrid Conference 2004, Lyon, France.
- Rahman M, Wang T, Desai B. Medical image retrieval and registration: towards computer assisted diagnostic approach. In: IDEAS Workshop on Medical Information Systems: The Digital Hospital (IDEAS-DH'04), 2004. Washington, DC: IEEE Computer Society, 2004;78–89.
- Güld MO, Kohlen M, Keysers D. Quality of DICOM header information for image categorization. Proceedings of the International Symposium on Medical Imaging 2002, San Diego, CA.
- Tagare HD, Jaffe C, Duncan J. Medical image databases: a content-based retrieval approach. *J Am Med Inform Assoc* 1997;4:184–198.
- Traina Júnior C, Traina AJM, Santos RR, Senzako EJ. A support system for content-based medical image retrieval in object oriented databases. *J Med Syst* 1997;21:339–352.
- Rosset A, Ratib O, Valle J. Integration of a multimedia teaching and reference database in a PACS environment. *RadioGraphics* 2002;22:1567–1577.
- Petrakis EGM. Content-based retrieval of medical images. *Int J Comput Res* 2002;11:171–182.
- Müller H, Michoux N, Bandon D, Geissbuhler A. A review of content-based image retrieval systems in medical applications – clinical benefits and future directions. *Int J Med Inform* 2004;73:1–23.
- Azevedo-Marques P, Honda MH, Rodrigues JAH, et al. Recuperação de imagem baseada em conteúdo: uso de atributos de textura para caracterização de microcalcificações mamográficas. *Radiol Bras* 2002;35:93–98.
- Haralick RM, Shanmuga K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern* 1973;SMC3:610–621.
- Walker RF, Jackway P, Longstaff ID. Improving co-occurrence matrix feature discrimination. In: Proc DICTA '95, 3rd Conference on Digital Image Computing: Techniques and Application 1995;643–648.
- McLean GF. Vector quantization for texture classification. *IEEE Trans Syst Cybern* 1993;23:637–644.
- Freeborough PA, Fox NC. MR image texture analysis applied to the diagnosis and tracking of Alzheimer's disease. *IEEE Trans Med Imaging* 1998;17:475–479.
- Mathias JM, Tofts PS, Losseff NA. Texture analysis of spinal cord pathology in multiple sclerosis. *Magn Reson Med* 1999;42:929–935.
- Materka A, Strzelecki M. Texture analysis methods – a review. In: COST B11 Report. Lodz, Poland: Technical University of Lodz, Institute of Electronics, 1998.
- Konak ES. A content-based image retrieval system for texture and color queries. (M.Sc. degree thesis). Ankara, Turkey: Department of Computer Engineering and Institute of Engineering and Science, Bilkent University, 2002.
- Sharma M, Singh S. Evaluation of texture methods for image analysis. 7th Australian and New Zealand Intelligent Information System Conference. Perth, Australia, 2001;117–121.
- Veltkamp RC, Hagedoorn M. State-of-the-art in shape matching. In: Lew M, editor. *Principles of visual information retrieval*. London: Springer-Verlag, 2000;87–119.
- Hajnal JV, Hill DLG, Hawkes DJ. Medical image registration. In: Neuman MR, editor. *Biomedical engineering*. Boca Raton, FL: CRC Press, 2001.
- Yoo TS. Insight into images: principles and practice for segmentation, registration, and image analysis. Wellesley, MA: AK Peters, 2004.
- InsightToolkit. [Acessado em: 8/10/2006]. Disponível em: <http://www.itk.org>
- Bueno JM. Suporte à recuperação de imagens médicas baseada em conteúdo através de histogramas métricos. São Carlos, SP: Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2002.
- Shyu CR, Bradley CE, Kak AC, Kosaka A, Aisen AM, Broderick LS. ASSERT: a physician-in-the-loop content-based retrieval system for HRCT image databases. *Computer Vision and Image Understanding* 1999;75:111–132.
- Kuo WJ, Chang RF, Lee CC, Moon WK, Chen DR. Retrieval technique for the diagnosis of solid breast tumors on sonogram. *Ultrasound Med Biol* 2002;28:903–909.
- Lehmann TM, Güld MO, Thies O, et al. Content-based image retrieval in medical applications. *Methods Inform Med* 2004;43:354–361.
- Traina AJM, Traina C, Bueno JM, Chino FJT, Azevedo-Marques P. Efficient content-based image retrieval through metric histograms. *World Wide Web J* 2003;6:157–185.