



**Universidade de São Paulo**

**Biblioteca Digital da Produção Intelectual - BDPI**

---

Departamento de Ciências de Computação - ICMC/SCC

Comunicações em Eventos - ICMC/SCC

---

2015-10

# Speeding up the combination of multiple descriptors for different boundary conditions

---

Latin American Computing Conference, XLI, 2015, Arequipa.  
<http://www.producao.usp.br/handle/BDPI/50001>

*Downloaded from: Biblioteca Digital da Produção Intelectual - BDPI, Universidade de São Paulo*

# Speeding up the combination of multiple descriptors for different boundary conditions

Rodrigo Fernandes Barroso\*, Marcelo Ponciano-Silva†, Agma Juci Machado Traina‡, Renato Bueno\*

\*Departamento de Computação  
Universidade Federal de São Carlos, São Carlos-SP, Brasil

† Instituto Federal de Educação, Ciência e Tecnologia do Triângulo Mineiro  
Uberaba-MG, Brasil

‡ Instituto de Ciências Matemáticas e de Computação  
Universidade de São Paulo, São Carlos-SP, Brasil

**Abstract**—Content-based complex data retrieval is becoming increasingly common in many types of applications. The content of these data is represented by intrinsic characteristics, extracted from them which together with a distance function allows similarity queries. Aimed at reducing the “semantic gap”, characterized by the disagreement between the computational representation of the extracted low-level features and how these data are interpreted by the human perception, the use of multiple descriptors has been the subject of several studies. This paper proposes a new method to carry out the combination of multiple descriptors for different boundary conditions in which the balancing is carried out in pairs, starting by the best candidate descriptor. In the experiments, the proposed method achieved computational cost up to 3650 times smaller than the exhaustive search for the best linear combination of descriptors, keeping almost the same average precision, with variations lower than 0.9%.

## I. INTRODUÇÃO

O gerenciamento de dados não tradicionais (como imagens, sons e vídeos) vem se tornando cada vez mais comum em diversos sistemas de computação, exigindo que o gerenciamento e a recuperação desses dados ocorram de forma mais eficiente e efetiva. Esses dados são comumente chamados de dados complexos e em muitos casos não possuem o conceito de dimensão associado.

Em muitos dos sistemas existentes a recuperação desses dados complexos é efetuada sobre a descrição textual referente ao conteúdo. No entanto, esse formato nem sempre retorna bons resultados, sendo dependente da experiência e outros fatores humanos do analista, além de não oferecer recursos satisfatórios para interagir com vários tipos de mídia [1].

Para comparação de dados complexos utiliza-se a similaridade entre pares de elementos, que indica o quão um elemento é parecido com outro. Essa comparação proporciona um grau de similaridade (ou dissimilaridade) entre pares de elementos de um conjunto, possibilitando a realização de consultas por similaridade. Os principais operadores de consulta por similaridade são [2]: consultas por abrangência, em que são retornados os elementos dentro de um raio de abrangência centrado em um elemento de referência; e consultas por k-vizinhos mais próximos, que retorna os k elementos mais similares ao elemento de referência.

Usualmente nas consultas por similaridade não são comparados os elementos de dados complexos em si, e sim características intrínsecas extraídas desses elementos. Consultas por similaridade que utilizam informações extraídas do conteúdo do dado complexo também são chamadas de Consultas Baseadas em Conteúdo. Como exemplo, em imagens pode-se extrair características relativas à textura, forma ou cor, comumente no formato de um vetor de características.

Após a obtenção desses vetores de características, são utilizadas funções para calcular a distância (dissimilaridade) entre pares de elementos. Estas funções resultam em um valor numérico que determina o grau de dissimilaridade entre elementos. A combinação vetor de característica (ou extrator de características) e função de distância é comumente encontrado na literatura como descritor [3] ou binômio [4]. Dados complexos comumente são representados em espaços métricos, considerando apenas as distâncias entre os elementos [2].

Um problema enfrentado pelas consultas por similaridade é o “gap semântico” [5], caracterizado pela divergência entre como os dados são interpretados pela percepção humana realizada a partir de análises de conteúdo de alto nível e a forma como são representados computacionalmente em características de baixo nível.

Visando o combate ao gap semântico em consultas por similaridade, a utilização de combinação de múltiplos descritores tem sido bastante pesquisada e tem demonstrado bons resultados: [6], [7], [8], [9], [10], [11]. Características intrínsecas diferentes tendem a se complementar na representação do dado, bem como é realizada pela percepção humana. Por este motivo a utilização de múltiplos descritores tende a melhorar a capacidade de discriminação dos dados.

Um fator importante a ser levado em consideração é que subconjuntos de dados podem apresentar comportamento diferentes, e a utilização de diferentes descritores para consultas nesses subconjuntos pode melhorar a precisão das consultas: uma característica pode ser mais relevante para evidenciar um subconjunto do restante dos elementos, e esta mesma característica pode não ser tão efetiva para evidenciar elementos de outro subconjunto. Em [12] foi proposta a adição de um parâmetro perceptual nas consultas por similaridade, com o objetivo de delimitar esses subconjuntos.

Em [13] foi proposta a utilização de condições de contorno para delimitar subconjuntos de imagens e então utilizar a melhor combinação de descritores para cada um desses subconjuntos, visando a diminuição no gap semântico. Para encontrar a melhor combinação linear de descritores para cada condição de contorno foram utilizados algoritmos de busca exaustiva.

Neste artigo, é proposto um novo método para realizar a combinação entre múltiplos descritores para diferentes condições de contorno, em que o balanceamento é realizado aos pares, partindo do melhor descritor candidato. No método proposto neste trabalho, a combinação é realizada aos pares partindo do melhor descritor candidato. Nos experimentos realizados o método proposto neste trabalho teve custo computacional até 3.650 vezes menor que a combinação utilizando algoritmos de busca exaustiva pela melhor combinação linear, mantendo praticamente a mesma precisão, com quedas de precisão inferiores a 0,9%.

Inicialmente, na Seção II é detalhado o uso de combinações de múltiplos descritores para diferentes condições de contorno. Em seguida é proposto um novo método para realizar a combinação entre múltiplos descritores, apresentado na Seção III.

## II. REVISITANDO A UTILIZAÇÃO DE MÚLTIPLOS DESCRITORES PARA DIFERENTES CONDIÇÕES DE CONTORNO

Nesta seção é discutido e detalhado o trabalho correlato apresentado em [13], que propôs o uso de condições de contorno para delimitar subconjuntos de imagens, para em seguida determinar o balanceamento de múltiplos descritores de forma a usar combinação ideal para cada uma desses subconjuntos de imagens nas consultas por similaridade.

O método se inicia com a definição das condições de contorno por um especialista no domínio da aplicação. Tomando como exemplo as aplicações de imagens médicas, as condições de contorno podem ser definidas através de dados associados aos exames, como por exemplo a hipótese de diagnóstico ou até mesmo uma análise prévia das imagens pelo radiologista, que pode pré-classificá-las com base em alterações encontradas. Com as imagens pré-classificadas de acordo com as condições de contorno definidas, é realizada a combinação de múltiplos descritores em uma fase de treinamento, visando definir a melhor combinação para cada condição de contorno.

Na Figura 1 são ilustrados os passos realizados na fase de treinamento. No primeiro passo o especialista prepara um conjunto de imagens pré-classificadas para serem utilizadas de acordo com cada uma das condições de contorno definidas. No passo seguinte são aplicados algoritmos de extração de características sobre essas imagens. Já com a base de dados alimentada com as características extraídas, é realizada a combinação de múltiplos descritores visando encontrar o melhor balanceamento entre eles para cada uma das condições de contorno.

Nos experimentos apresentados em [13], as combinações foram definidas utilizando um algoritmo de busca exaustiva pela melhor combinação linear dentre um conjunto de pesos pré-determinados, que variavam entre 1/10 e 10, garantindo assim o balanceamento ideal para a combinação linear para cada

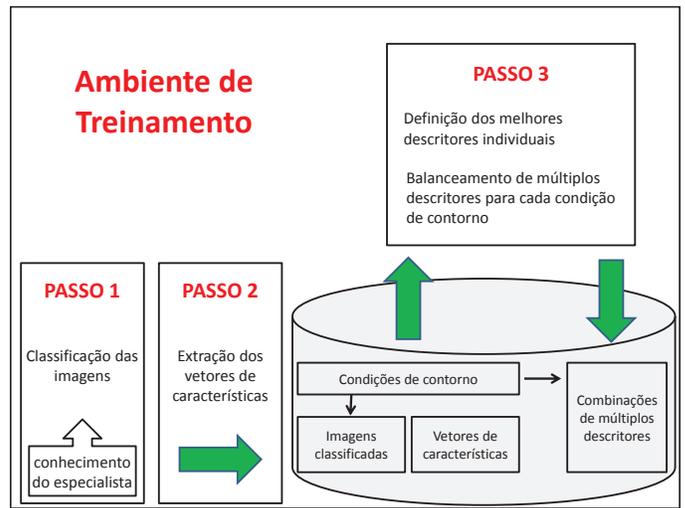


Figura 1. Ambiente de treinamento para combinação de múltiplos descritores para diferentes condições de contorno.

condição de contorno (com o conjunto de pesos definidos), de acordo com a Equação (1).

A fase de treinamento resulta em uma tabela  $T$  que relaciona o balanceamento de múltiplos descritores (CMD) a cada uma das Condições de Contorno (CC) do conjunto, onde temos os elementos  $T = (CC_1, CMD_1), \dots, (CC_i, CMD_i)$ , onde  $i$  representa a quantidade de subconjuntos delimitados por condições de contorno. Com estes resultados gerados, parte-se para o ambiente de utilização, conforme mostrado na Figura 2.

Ao se realizar uma consulta por similaridade, além da imagem de referência (elemento central da consulta) e outros parâmetros relativos ao tipo de consulta (como por exemplo o número de vizinho mais próximos para consultas  $k - NN$  ou raio de abrangência para consultas por abrangência), deve também ser fornecido como parâmetro o identificador da condição de contorno da imagem de referência (passo 1 da Figura

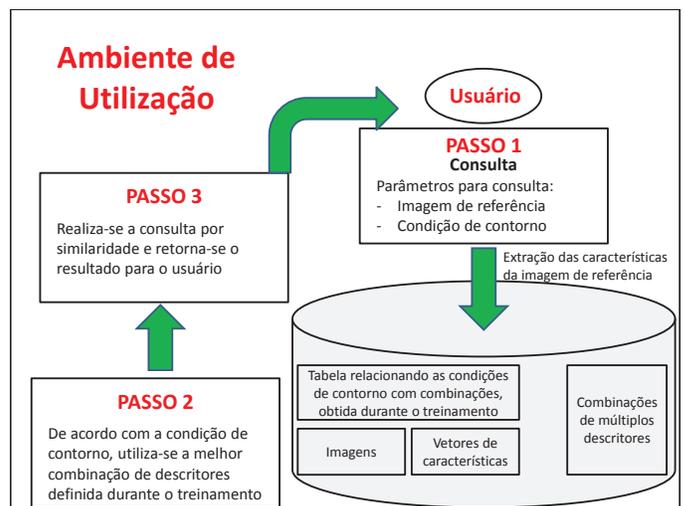


Figura 2. Ambiente operacional: utilização de combinações específicas para diferentes condições de contorno.

2). Esta condição de contorno é utilizada para identificar (ou estimar) a qual subconjunto a imagem de referência pertence e então utilizar a combinação de múltiplos descritores que melhor a evidencie do restante do conjunto, de acordo com a tabela  $T$  resultante da fase de treinamento. A combinação escolhida será utilizada na consulta por similaridade.

Conforme já discutido nesta seção, em [13] foi utilizado um algoritmo de busca exaustiva pela definição da melhor combinação linear para cada condição de contorno. Porém, tal estratégia é proibitiva do ponto de vista do custo computacional, pois são testadas todas as combinações possíveis de pesos para os descritores. Na seção seguinte é apresentada uma nova proposta para realização dessas combinações, com resultados de precisão similares e com custo computacional inferior.

### III. PROPOSTA DE UM NOVO MÉTODO PARA COMBINAÇÃO DE MÚLTIPLOS DESCRITORES PARTINDO DO MELHOR DESCRITOR

Considerando as imagens  $x$  e  $y$ , e as métricas  $\delta_1, \dots, \delta_n$  dos descritores a serem combinados, definidas sobre os domínios dos respectivos vetores de características, a composição da função de distância para combinação dos descritores  $\delta_c$  é dada pela Equação (1):

$$\delta_c(x, y) = \sum_{i=1}^n \omega_i * \frac{\delta_i(x, y)}{dmax_i} \quad (1)$$

onde  $\omega_i$  é o peso atribuído ao  $i$ -ésimo descritor e  $dmax_i$  representa a distância máxima obtida pela comparação entre todos os elementos do conjunto utilizando este mesmo descritor, que normaliza as distâncias entre 0 e 1.

Os algoritmos de busca exaustiva para combinação linear utilizados em [13] testam todas as combinações possíveis (a partir de uma quantidade determinada de possíveis pesos) a fim de encontrar a melhor combinação linear entre os descritores. O extenso espaço de busca torna o processo bastante custoso computacionalmente, com um crescimento exponencial de acordo com o número de descritores utilizados.

Neste artigo propõe-se a realização de combinações lineares dos descritores a partir do descritor que apresente o melhor desempenho individual, avaliado pela sua precisão média na fase de treinamento. Este descritor assume então o peso  $\omega_1 = 1$ .

Em uma fase que precede ao balanceamento são definidos os descritores candidatos que poderão ser combinados. Os descritores podem ser definidos por um especialista no domínio da aplicação. Neste trabalho, esses descritores foram escolhidos experimentalmente: foram testadas combinações de algumas funções de distância com os vetores de características extraídos, calculando a precisão média de consultas realizadas sobre um conjunto de imagens pré-classificadas, utilizando curvas de precisão e revocação [14]. Os descritores candidatos são então ranqueados pela precisão média. Partindo do descritor melhor avaliado, inicia-se a combinação com o segundo descritor melhor avaliado. Considerando as imagens  $x$  e  $y$ , e as métricas dos descritores candidatos ordenados,  $\delta_1, \dots, \delta_n$ , definidas sobre os domínios dos respectivos vetores de características, a composição da função de distância  $\delta_{c_1}$  para

combinação do primeiro descritor com o segundo descritor é dada pela Equação (1):

$$\delta_{c_1}(x, y) = \frac{\delta_1(x, y)}{dmax_1} + \omega_2 * \frac{\delta_2(x, y)}{dmax_2} \quad (2)$$

onde  $\omega_2$  é o peso dado ao segundo descritor e  $dmax_1$  e  $dmax_2$  representam a distância máxima encontrada pela comparação entre todos os elementos do conjunto utilizando o respectivo descritor, utilizadas para normalizar a participação de cada descritor na combinação.

Neste trabalho e em [13], para a definição do peso  $\omega_i$  foram avaliadas as combinações lineares para um mesmo conjunto de pesos que variam entre 1/10 e 10, possibilitando assim a comparação direta entre o método proposto e a utilização de algoritmos de busca exaustiva pela melhor combinação, tanto com relação ao custo computacional como com relação à precisão dos resultados.

Após a definição da melhor combinação entre os dois primeiros descritores, parte-se para a combinação com o terceiro melhor descritor, utilizando a Equação (3):

$$\delta_{c_2}(x, y) = \delta_{c_1}(x, y) + \omega_3 * \frac{\delta_3(x, y)}{dmax_3} \quad (3)$$

onde  $\omega_3$  é o peso dado ao terceiro melhor descritor e  $dmax_3$  representa a distância máxima encontrada pela comparação entre todos os elementos do conjunto utilizando este descritor.

Dessa forma, os descritores são combinados aos pares sequencialmente, sempre testando os pesos do próximo descritor  $i + 1$  na combinação com o melhor balanceamento já definido para os  $i$  descritores anteriores, de acordo com a Equação (4):

$$\delta_{c_i}(x, y) = \delta_{c_{i-1}}(x, y) + \omega_{i+1} * \frac{\delta_{i+1}(x, y)}{dmax_{i+1}} \quad (4)$$

## IV. EXPERIMENTOS

Na sequência são apresentados os resultados dos experimentos realizados com o método proposto para realização das combinações dos descritores. Os resultados são comparados com aqueles obtidos pelos algoritmos de busca exaustiva pela melhor combinação linear, utilizando os mesmos conjuntos de dados utilizados em [13].

Para facilitar a notação e o entendimento, deste ponto em diante, consideraremos os descritores já normalizados pela maior distância encontrada na comparação entre todos os elementos do respectivo conjunto. Portanto, quando fizermos referência a  $D_i$  nos gráficos e equações que se seguem, estamos considerando a Equação (5):

$$D_i(x, y) = \frac{\delta_i(x, y)}{dmax_i} \quad (5)$$

onde  $\delta_i$  é a métrica utilizada no descritor,  $x$  e  $y$  são duas imagens e  $dmax_i$  é a distância máxima encontrada pela comparação entre todos os elementos do conjunto utilizando  $\delta_i$ .

### A. Experimento 1: Imagens de exames de tomografias de pulmão

Neste experimento foi utilizada uma coleção com 247 imagens de tomografia de pulmão, classificadas de acordo com o achado radiológico em 6 classes: consolidação, opacidade em vidro fosco, espessamento do septo interlobular, faveolamento, enfisema e normal (nenhum achado radiológico). Foi utilizado o parâmetro perceptual detectado pelos médicos como condição de contorno, relacionados com cada achado radiológico, respectivamente: aumento homogêneo e não homogêneo da atenuação, fissura interlobular e baixa atenuação com e sem fissuras.

Foram experimentalmente definidos os seguintes descritores candidatos:

- $D1(x, y)$ : Histograma Baixo-L2;
- $D2(x, y)$ : Histograma Alto-L2;
- $D3(x, y)$ : Histograma Tradicional-Linf;
- $D4(x, y)$ : Textura-Canberra.

1) *Fase de treinamento*: Para exemplificar o emprego do método para definição da combinação de múltiplos descritores para cada condição de contorno partindo do melhor descritor, a seguir são mostrados os passos para a definição da combinação para a primeira condição de contorno, relativa a condição de contorno “Consolidação”. Para essa condição de contorno os descritores candidatos foram ranqueados de acordo com as precisões médias alcançadas em consultas por similaridade realizadas com o conjunto de imagens:

- 1º -  $D1(x, y)$ : precisão média de 46,65%;
- 2º -  $D4(x, y)$ : precisão média de 44,24%;
- 3º -  $D3(x, y)$ : precisão média de 38,26%;
- 4º -  $D2(x, y)$ : precisão média de 34,50%.

De acordo com o método proposto, iniciou-se a fase de balanceamento dos descritores utilizando o melhor descritor,  $D1(x, y)$  (Histograma Baixo-L2), com peso fixo e incluindo o segundo melhor descritor,  $D4(x, y)$  (Textura - Canberra), variando seu peso  $\omega$  (entre 0,1 e 10). Os resultados são apresentados na Figura 3, que mostra como a variação do peso  $\omega$  influencia na precisão da consulta através da combinação entre os dois melhores descritores.

Analisando as informações do gráfico verifica-se que a combinação destes descritores apresenta o melhor resultado com o peso de  $D1(x, y)$  aproximadamente nove vezes maior que o peso de  $D4(x, y)$ . Com essa combinação, a precisão média alcançada foi de 49,53%, apresentando uma melhora de aproximadamente 6% na consulta utilizando apenas o melhor descritor ( $D1(x, y)$ ).

Depois de definida a melhor combinação de pesos entre o melhor descritor e o segundo melhor descritor, fixa-se esta combinação ( $D1(x, y) + 0.11 * D4(x, y)$ ) e testa-se os valores de peso para o próximo descritor,  $D3(x, y)$  (Tradicional-Linf). O resultado é apresentado na Figura 4.

Da mesma forma que no passo anterior, no gráfico é possível observar que a inclusão do terceiro descritor melhorou a

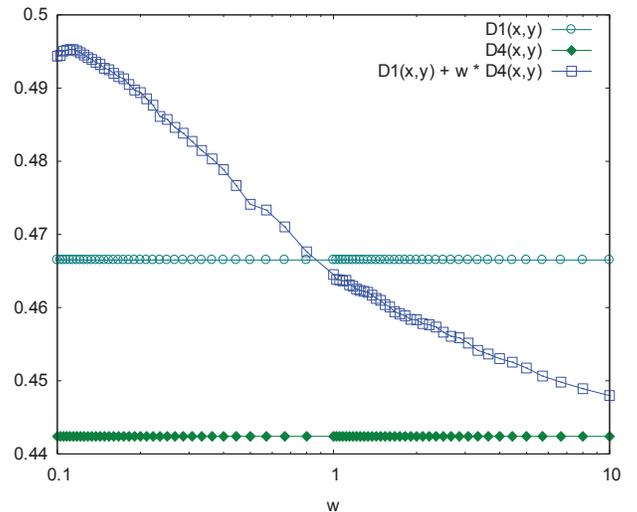


Figura 3. Imagens de tomografias de pulmão, condição de contorno “Consolidação”: Precisão média combinando linearmente os descritores  $D1(x, y)$  (Histograma Baixo-L2) com  $D4(x, y)$  (Textura-Canberra).

precisão média alcançada pela combinação de descritores para alguns valores de  $\omega$ , e o melhor resultado alcançado foi para o peso descritor incluído aproximadamente dez vezes menor que a combinação já definida dos dois primeiros descritores, alcançando 49,91% de precisão média. Portanto, fixa-se o balanceamento da combinação em  $D1(x, y) + 0.11 * D4(x, y) + 0.10 * D3(x, y)$ .

Este processo deve continuar testando a inclusão de novos descritores na combinações para todos os demais descritores candidatos. Neste experimento, como foram testados apenas 4 descritores candidatos, o último passo testou os pesos do descritor  $D2(x, y)$  (Histograma Alto-L2) ao inseri-lo na combinação dos três descritores já definida no passo anterior.

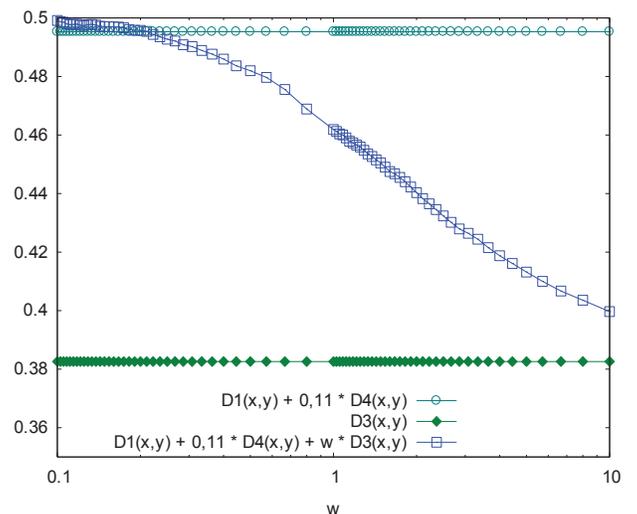


Figura 4. Imagens de tomografias de pulmão, condição de contorno “Consolidação”: Precisão média combinando linearmente  $(D1(x, y) + 0.11 * D4(x, y) + \omega * D3(x, y))$ .

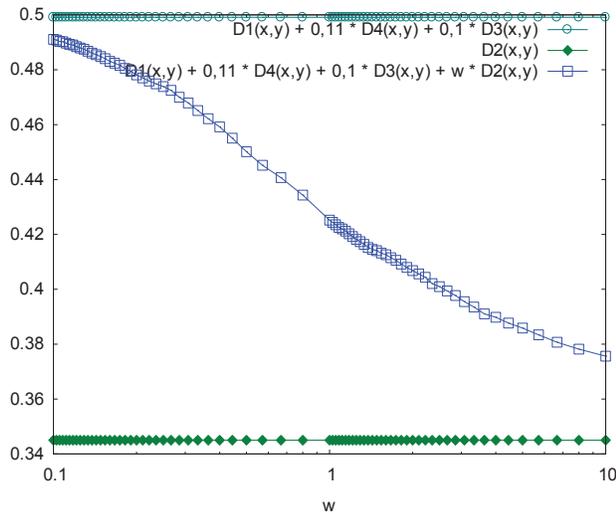


Figura 5. Imagens de tomografias de pulmão, condição de contorno “Consolidação”: Precisão média combinando linearmente  $D1(x, y) + 0.11 * D4(x, y) + 0.1 * D3(x, y) + w * D2(x, y)$ .

Como pode-se ver na Figura 5, a inclusão do descritor  $D2(x, y)$  diminuiu a precisão média alcançada apenas com os três primeiros descritores para todos os pesos testados, degradando o poder de discriminação da combinação. Conseqüentemente, este descritor não deve ser inserido na combinação final de descritores para esta condição de contorno.

Neste experimento, através da combinação de múltiplos descritores partindo do melhor descritor, a métrica resultante para a condição de contorno “Consolidação” foi:

$$\delta c(x, y) = D1(x, y) + 0.11 * D4(x, y) + 0,10 * D3(x, y)$$

Esta combinação foi exatamente a mesma obtida pela combinação linear exaustiva, como pode ser visto na Tabela I, que mostra as combinações definidas para todas as condições de contorno pelo algoritmo de busca exaustiva para a melhor combinação linear.

Este mesmo método foi realizado para todas as demais condições de contorno do conjunto de imagens de exames de tomografias de pulmão, resultando nas combinações mostradas na Tabela II.

Condição de Contorno	Combinação de Múltiplos Descritores
Consolidação	$D1(x, y) + D3(x, y) * 0.1 + D4(x, y) * 0.11$
Enfisema	$D2(x, y) + D4(x, y) * 0.57$
Vidro Fosco	$D1(x, y) + D3(x, y) * 0.27 + D4(x, y) * 0.44$
Faveolamento	$D2(x, y) + D4(x, y) * 0.67$
Espessamento	$D2(x,y)*10 + D3(x, y) + D4(x, y) * 1.74$

Tabela I. COMBINAÇÕES OBTIDAS PARA CADA CONDIÇÃO DE CONTOURNO UTILIZANDO A BUSCA EXAUSTIVA PARA DEFINIÇÃO DA COMBINAÇÃO LINEAR

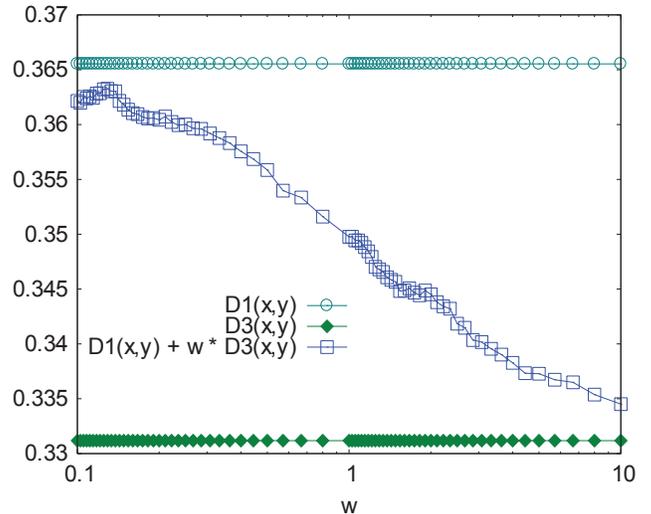


Figura 6. Imagens de tomografias de pulmão, condição de contorno “Faveolamento”: Precisão média combinando linearmente os descritores  $D1(x, y)$  (Histograma Baixo-L2) com  $D3(x, y)$  (Histograma Tradicional-Linf).

Analisando as combinações obtidas pelo algoritmo de busca exaustiva (Tabela I) e pelo método proposto, partindo do melhor descritor (Tabela II) pode-se notar que na maioria das condições de contorno os balanceamentos foram similares.

Porém, algumas combinações obtidas pelo método proposto foram definidas utilizando descritores descartados pela combinação linear exaustiva. Como exemplo pode-se observar a aplicação do método proposto para a condição de contorno “Faveolamento”, exemplificada a seguir. O descritores candidatos foram ranqueados na seguinte ordem:

- 1º -  $D1(x, y)$ : precisão média de 36,55%;
- 2º -  $D3(x, y)$ : precisão média de 33,12%;
- 3º -  $D4(x, y)$ : precisão média de 32,48%;
- 4º -  $D2(x, y)$ : precisão média de 31,73%.

Ao realizar a combinação entre os dois primeiros descritores ranqueados, pode-se observar na Figura 6 que a utilização do descritor  $D1(x, y)$  individualmente apresentou melhores resultados. Portanto, o descritor  $D3(x, y)$  foi descartado.

Na sequência, testou-se a combinação do descritor  $D1(x, y)$  com o terceiro melhor descritor ( $D4(x, y)$ ). Para

Condição de Contorno	Combinação de Múltiplos Descritores
Consolidação	$D1(x, y) + D3(x, y) * 0.10 + D4(x, y) * 0.11$
Enfisema	$D2(x, y) + D3(x, y) * 0.30 + D4(x, y) * 0.60$
Vidro Fosco	$D1(x, y) + D3(x, y) * 0.25 + D4(x, y) * 0.33$
Faveolamento	$D1(x, y) + D2(x, y) * 1.14 + D4(x, y) * 0.67$
Espessamento	$D2(x; y) + D3(x, y) * 0.33 + D4(x, y) * 0.17$

Tabela II. COMBINAÇÕES OBTIDAS PARA CADA CONDIÇÃO DE CONTOURNO UTILIZANDO A COMBINAÇÃO LINEAR PARTINDO DO MELHOR DESCRITOR

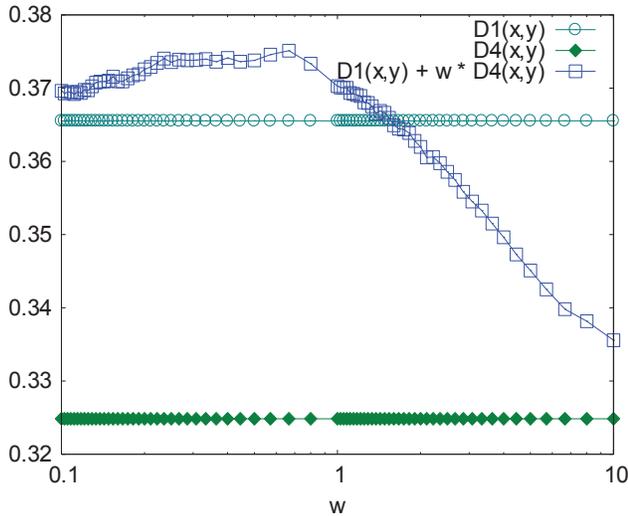


Figura 7. Imagens de tomografias de pulmão, condição de contorno “Faveolamento”: Precisão média combinando linearmente os descritores  $D1(x, y)$  (Histograma Baixo-L2) com  $D4(x, y)$  (Textura-Canberra).

alguns valores de  $\omega$  a combinação obteve melhores resultados do que a utilização individual de ambos os descritores, e foi definido o peso 0,67 para o descritor  $D4(x, y)$ , conforme mostrado na Figura 7.

Finalmente, testou-se a combinação do balanceamento já definido entre os descritores  $D1(x, y)$  e  $D4(x, y)$  com o quarto melhor descritor ( $D2(x, y)$ ). A inserção do descritor  $D2(x, y)$  contribuiu para o aumento da precisão média das consultas, e seu peso foi definido com o valor de 1,14, conforme mostrado na Figura 8.

Comparando as combinações definidas para a condição de contorno “Faveolamento” obtidas por ambos os métodos (apresentadas nas Tabelas I e II), pode-se observar que nenhum dos dois melhores descritores (ranqueados pela precisão média

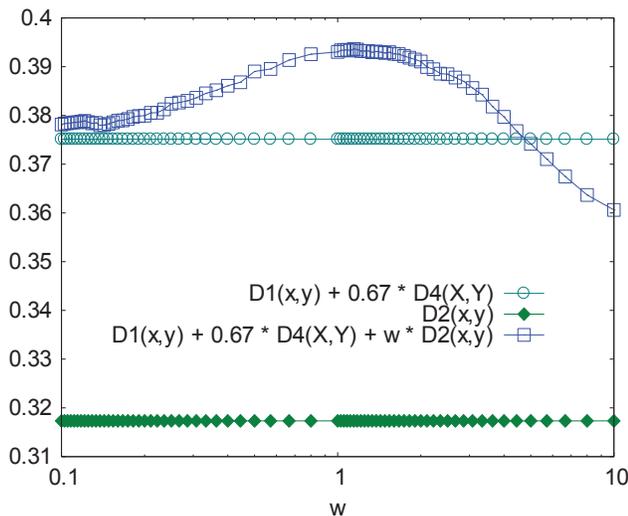


Figura 8. Imagens de tomografias de pulmão, condição de contorno “Faveolamento”: Precisão média combinando linearmente  $(D1(x, y) + 0.67 * D4(x, y) + \omega * D2(x, y))$ .

das consultas quando aplicados individualmente) são utilizados na combinação linear utilizando a busca exaustiva, que apresenta o resultado ótimo para a combinação linear entre todos os descritores candidatos com os pesos possíveis definidos. Porém, na combinação partindo do melhor descritor, apesar de o segundo descritor ranqueado ( $D2(x, y)$ ) ter sido descartado, o descritor  $D1(x, y)$  manteve-se na combinação final.

2) *Fase de utilização*: A precisão das consultas utilizando a combinação realizada pelo método proposto foram muito parecidas àquelas obtidas pela combinação realizada através de algoritmos de busca exaustiva, conforme a Figura 9.

A precisão média alcançada realizando a combinação partindo do melhor descritor foi de 47,75%, enquanto a precisão média alcançada pela combinação linear exaustiva foi de 48,16%. Porém o método partindo do melhor descritor necessitou de apenas 0,0274% de esforço computacional (número de consultas realizadas para definição do balanceamento dos múltiplos descritores), ou seja, teve custo aproximadamente 3650 vezes menor.

Mantendo praticamente a mesma precisão média final das consultas com relação ao método de combinação com algoritmos de busca exaustiva utilizando condições de contorno, com queda inferior a 0,9%, o método de combinação partindo do melhor descritor (utilizando condições de contorno) apresentou ganhos com relação aos outros métodos avaliados em [13]: ganho de 6,27% sobre a precisão média de uma única combinação linear exaustiva para o conjunto todo, 8,52% sobre a tríade [12] e até 33,38% de ganho sobre o uso dos descritores individuais, conforme pode ser visto nas Figuras 10 e 11.

## B. Experimento 2: Imagens de exames Ressonância Magnética

O conjunto de imagens utilizado nesse experimento é composto por 704 imagens de exames de Ressonância Magnética. As imagens foram separadas em 40 classes de acordo com a região do corpo, plano de digitalização e posição do corte. Essas 40 classes foram agrupadas em 8 conjuntos, sendo a condição de contorno definida pelo tipo do exame: Angiograma, Abdômen Axial, Abdômen Coronal, Cabeça Axial, Cabeça Coronal, Cabeça Sagital, Pélvis Axial e Espinha Sagital.

Vale ressaltar que neste experimento as condições de contorno não são diretamente relacionadas às classes de imagens pré-classificadas: a partir de uma mesma condição de contorno, tipo de exame, escolhe-se a melhor combinação de múltiplos descritores para melhor evidenciar os elementos para todas as classes cobertas pela condição.

Foram extraídas as seguintes características: Histograma tradicional de níveis de cinza, para representar as características de cor, 256 primeiros momentos de Zernike para representação de forma [15] e descritores de Haralick para textura [16]. Para realização deste experimento foram realizados testes para a definição dos seguintes descritores candidatos:

- $D1(x, y)$ : Textura-Canberra;
- $D2(x, y)$ : Forma-L2;
- $D3(x, y)$ : Histograma-L1;

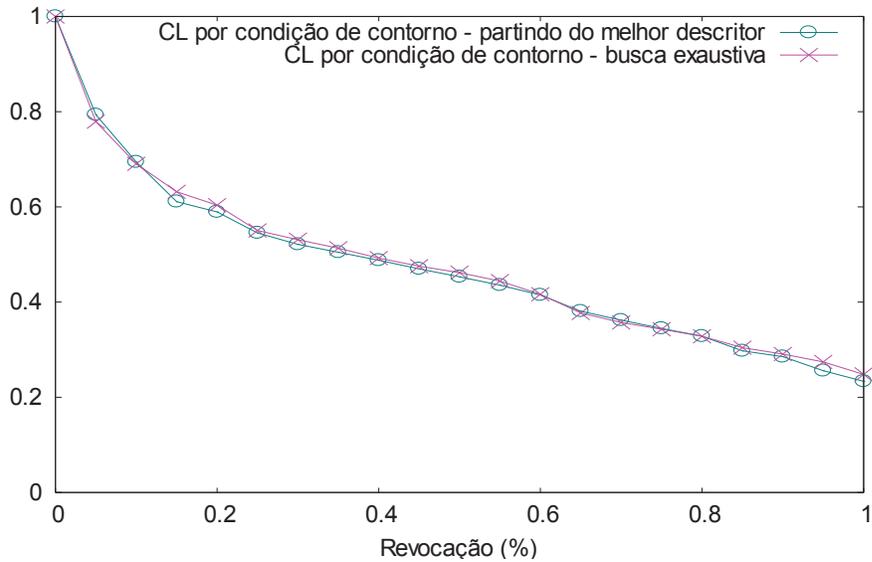


Figura 9. Imagens de tomografias de pulmão: Curva de Precisão x Revocação para consultas utilizando a combinação de descritores definida por algoritmos de busca exaustiva e a combinação partindo do melhor descritor.

1) *Fase de treinamento*: A seguir são mostrados os passos para a definição da combinação de descritores para imagens do tipo “Pélvis Axial”. Foram testados e ranqueados os descritores de acordo com as precisões médias alcançadas para essa condição de contorno:

- 1º -  $D2(x, y)$ : precisão média de 87,07%;
- 2º -  $D1(x, y)$ : precisão média de 83,72%;
- 3º -  $D3(x, y)$ : precisão média de 75,44%;

Iniciou-se o balanceamento dos descritores aos pares, utilizando o melhor descritor  $D2(x, y)$  (Forma-L2) com peso fixo 1 e incluindo o segundo melhor descritor  $D1(x, y)$  (Textura-Canberra) variando seu peso  $\omega$  (entre 1/10 e 10). Na Figura 12 são apresentados os valores de precisão média alcançados na combinação dos dois descritores variando o peso  $\omega$ . A combinação entre os descritores  $D2(x, y)$  e  $D1(x, y)$  apresentou melhores precisões médias do que a utilização dos mesmos descritores individualmente para todos os valores de  $\omega$  testados. O peso dado ao descritor  $D1(x, y)$  que apresentou melhor resultado de precisão média foi 1,33, alcançando 94%

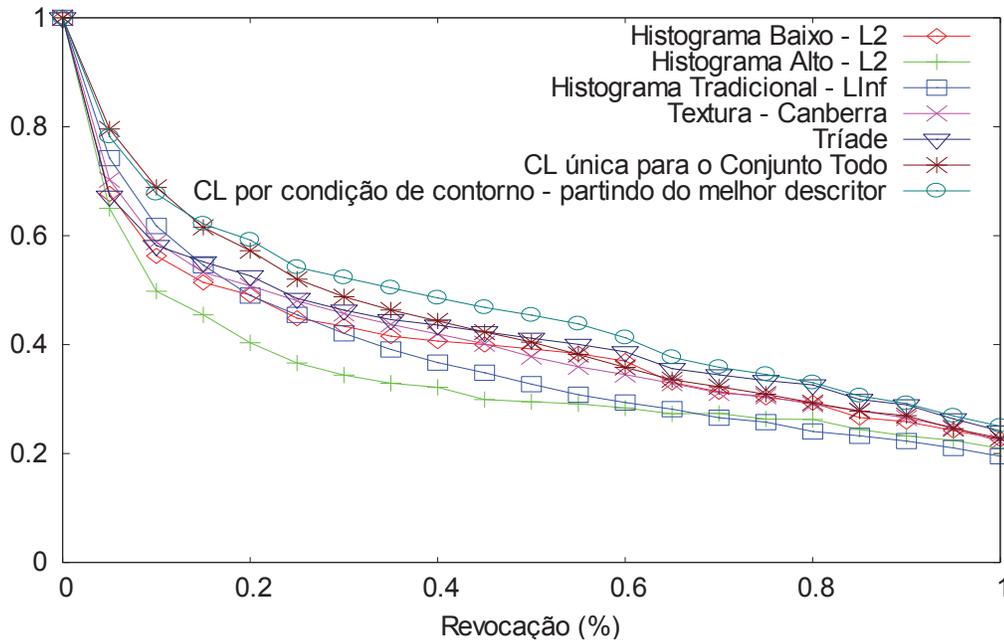


Figura 10. Imagens de tomografias de pulmão: Curvas de Precisão x Revocação comparando os resultados das consultas obtidos pela combinação de descritores usando condições de contorno partindo do melhor descritor com outras combinações e descritores individuais.

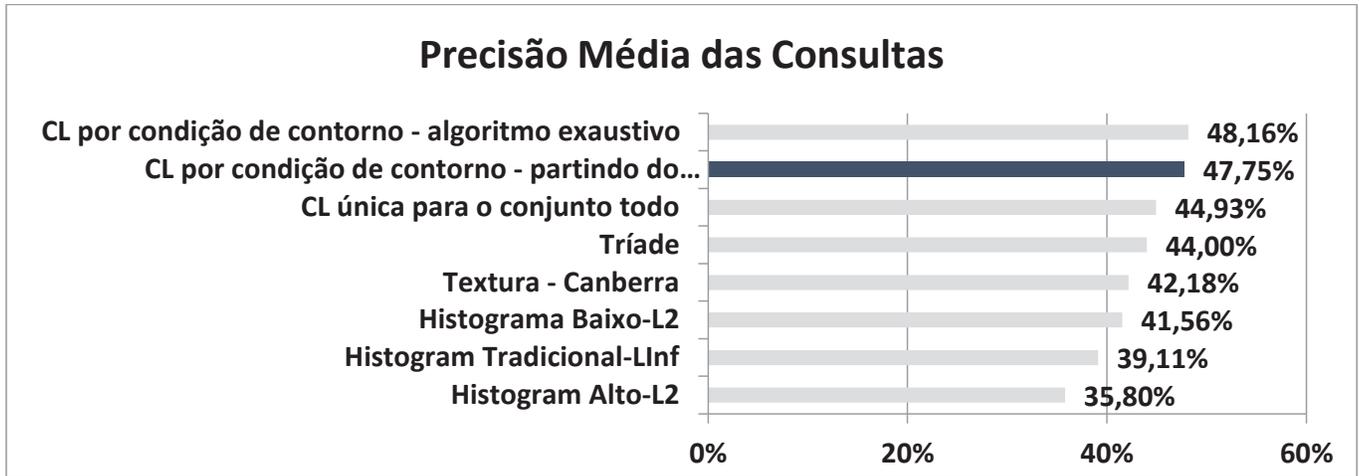


Figura 11. Imagens de tomografias de pulmão: Precisão média das consultas das consultas obtida pela combinação de descritores usando condições de contorno partindo do melhor descritor comparada com outras combinações e descritores individuais.

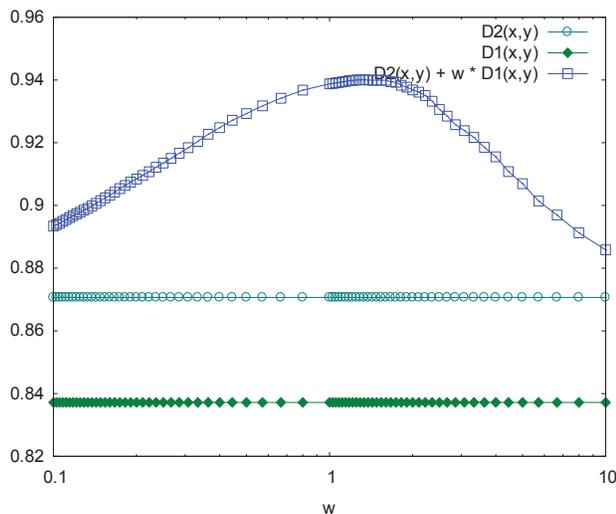


Figura 12. Imagens de ressonância magnética, condição de contorno “Pélvis Axial”: Precisão média combinando linearmente os descritores  $D2(x, y)$  (Forma-L2) com  $D1(x, y)$  (Textura-Canberra).

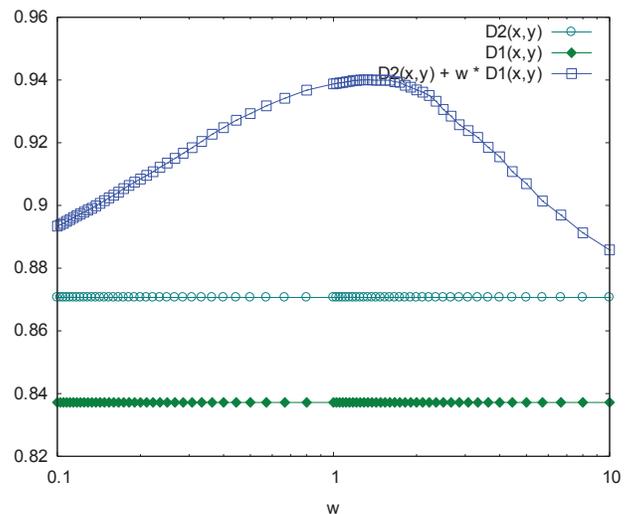


Figura 13. Imagens de ressonância magnética, condição de contorno “Pélvis Axial”: Precisão média combinando o balanceamento definido anteriormente ( $D2(x, y) + 1.33 * D1(x, y)$ ) com  $D3(x, y)$ .

de precisão média. Portanto, fixou-se o valor de 1 para o descritor  $D2(x, y)$  e 1,33 para o descritor  $D1(x, y)$ .

Em seguida foram testados os valores de peso para combinação com o próximo descritor candidato, o descritor  $D3(x, y)$  (Histograma-L1). Os valores de precisão média alcançados para os pesos atribuídos ao descritor  $D3(x, y)$  no balanceamento com a combinação previamente definida são apresentados na Figura 13. A combinação do balanceamento entre os dois melhores descritores (definida na etapa anterior) e o descritor  $D3(x, y)$  também gerou valores maiores de precisão média das consultas para todos os valores de  $\omega$  testados, e o peso para o descritor  $D3(x, y)$  que apresentou melhor resultado foi 0,66.

Portanto, como apenas três descritores candidatos foram testados nesse experimento, a combinação final obtida para a condição de contorno “Pélvis Axial” foi:

$$\delta c(x, y) = D2(x, y) + 1.33 * D1(x, y) + 0,66 * D3(x, y)$$

O mesmo procedimento foi realizado para todas as condições de contorno definidas para o conjunto de imagens de ressonância magnética, calculando-se a melhor combinação dos descritores candidatos sempre partindo do melhor descritor.

Na Tabela IV são apresentadas as combinações definidas pelo algoritmo de combinação linear partindo do melhor descritor para todas as condições de contorno do conjunto de imagens de ressonância magnética, e na Tabela III são apresentadas as combinações definidas pelo algoritmo de busca exaustiva da combinação linear.

Apenas para uma melhor comparação entre as combinações obtidas, os valores dos pesos da Tabela III foram ajustados proporcionalmente para que o descritor  $D1(x, y)$  fosse

Condição de Contorno	Combinação de Múltiplos Descritores
Angiograma	$D1(x, y) + D2(x, y) * 0.19 + D3(x, y) * 0.21$
Abdômen	$D1(x, y) + D2(x, y) * 5.71 + D3(x, y) * 2.67$
Cabeça Axial	$D1(x, y) + D2(x, y) * 2.50 + D3(x, y) * 1.60$
Pélvis Axial	$D1(x, y) + D2(x, y) * 0.50 + D3(x, y) * 0.40$
Abdômen Coronal	$D1(x, y) + D2(x, y) * 3.33 + D3(x, y) * 0.57$
Cabeça Coronal	$D1(x, y) + D2(x, y) * 5.71 + D3(x, y) * 0.50$
Cabeça Sagital	$D1(x, y) + D2(x, y) * 2.86 + D3(x, y) * 0.33$
Espinha Sagital	$D2(x, y)$

Tabela III. COMBINAÇÕES OBTIDAS PARA CADA CONDIÇÃO DE CONTORNO UTILIZANDO A COMBINAÇÃO LINEAR UTILIZANDO ALGORITMO DE BUSCA EXAUSTIVA

Condição de Contorno	Combinação de Múltiplos Descritores
Angiograma	$D1(x, y) + D2(x, y) * 0.18 + D3(x, y) * 0.20$
Abdômen	$D1(x, y) + D2(x, y) * 3.75 + D3(x, y) * 1.50$
Cabeça Axial	$D1(x, y) + D2(x, y) * 2.50 + D3(x, y) * 2.00$
Pélvis Axial	$D1(x, y) + D2(x, y) * 0.75 + D3(x, y) * 0.50$
Abdômen Coronal	$D1(x, y) + D2(x, y) * 3.33 + D3(x, y) * 0.57$
Cabeça Coronal	$D1(x, y) + D2(x, y) * 3.00 + D3(x, y) * 0.30$
Cabeça Sagital	$D1(x, y) + D2(x, y) * 2.75 + D3(x, y) * 0.32$
Espinha Sagital	$D2(x, y)$

Tabela IV. COMBINAÇÕES OBTIDAS PARA CADA CONDIÇÃO DE CONTORNO UTILIZANDO A COMBINAÇÃO LINEAR PARTINDO DO MELHOR DESCRITOR

apresentado com peso 1 (como acontece na Tabela IV). Por exemplo, a combinação obtida anteriormente para as imagens de “Pélvis Axial”:

$$D1(x, y) * 1,33 + D2(x, y) + D3(x, y) * 0,66$$

é apresentada como:

$$D1(x, y) + D2(x, y) * 0.75 + D3(x, y) * 0.50.$$

Tal ajuste não tem nenhuma influência nos resultados dos experimentos. Mesmo que fossem utilizados os pesos ajustados, os resultados dos experimentos de precisão média seriam exatamente os mesmos, pois a proporção entre os pesos dos descritores é mantida e os ranqueamentos dos elementos nas consultas por k-vizinhos mais próximos realizadas seriam exatamente os mesmos.

Como pode-se observar comparando-se os dados das Tabelas III e IV, no caso da condição de contorno analisada (Pélvis Axial), existe uma pequena diferença nos pesos definidos pela combinação com o algoritmo de busca exaustiva e pela combinação partindo do melhor descritor (que não testa todas as combinações testadas na combinação por busca exaustiva). Porém, na maioria das condições de contorno, os pesos definidos foram bem próximos ou iguais aos definidos

pelo algoritmo de busca exaustiva.

2) *Fase de utilização*: A grande similaridade entre os balanceamentos obtidos por ambos os métodos se reflete na comparação dos resultados das consultas realizadas, apresentados na Figura 14. Pode-se ver que as curvas de precisão e revocação de ambos os métodos são praticamente idênticas.

A precisão média alcançada utilizando a combinação linear partindo do melhor descritor foi de 91,76%, muito próxima da precisão média alcançada utilizando a combinação linear exaustiva, que foi de 91,81%. Nas Figuras 16 e 15 são comparados os valores de precisão média e curvas de precisão e revocação para combinações e descritores individuais apresentados em [13].

Apesar de alcançar praticamente a mesma precisão média do método de combinação exaustiva, com variação próxima a 0,05%, o custo computacional para a obtenção da combinação partindo do melhor descritor foi aproximadamente 74 vezes menor (apenas 1,35% do esforço computacional), considerando o número de consultas realizadas para definição do balanceamento dos múltiplos descritores.

## V. CONCLUSÃO

A expectativa para os sistemas de recuperação baseada em conteúdo é a de que cada vez mais novos descritores de características, específicos para cada contexto, sejam desenvolvidos para os diversos domínios existentes. Isso porque a literatura mostra que um único descritor tem dificuldades de representar todas as características consideradas na comparação humana das imagens. Com isso, a visão ideal de similaridade alcançada pela percepção humana somente será aproximada computacionalmente quando uma quantidade abrangente de descritores for associada tal como a consideração humana de similaridade. No entanto, quanto maior o número de possíveis descritores para um determinado contexto, exponencialmente maior será o custo computacional exigido para determinar a melhor combinação desses descritores.

O método apresentado nesse trabalho aumenta a perspectiva de utilização de múltiplos descritores, pois mostrou um ganho considerável na otimização computacional no cálculo de melhor combinação de múltiplos descritores. Nos experimentos realizados, o método proposto diminuiu em até 3650 vezes o custo computacional para se definir a melhor combinação de múltiplos descritores para cada condição de contorno apresentada no contexto, mantendo praticamente a mesma precisão das consultas. Esse resultado reforça a visão de que mesmo em cenários onde o número de possíveis descritores para cada contexto possa aumentar, a utilização de múltiplos descritores será uma alternativa computacionalmente viável como possível solução para diminuição do “gap semântico”.

## AGRADECIMENTOS

Agradecimentos à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo suporte financeiro.

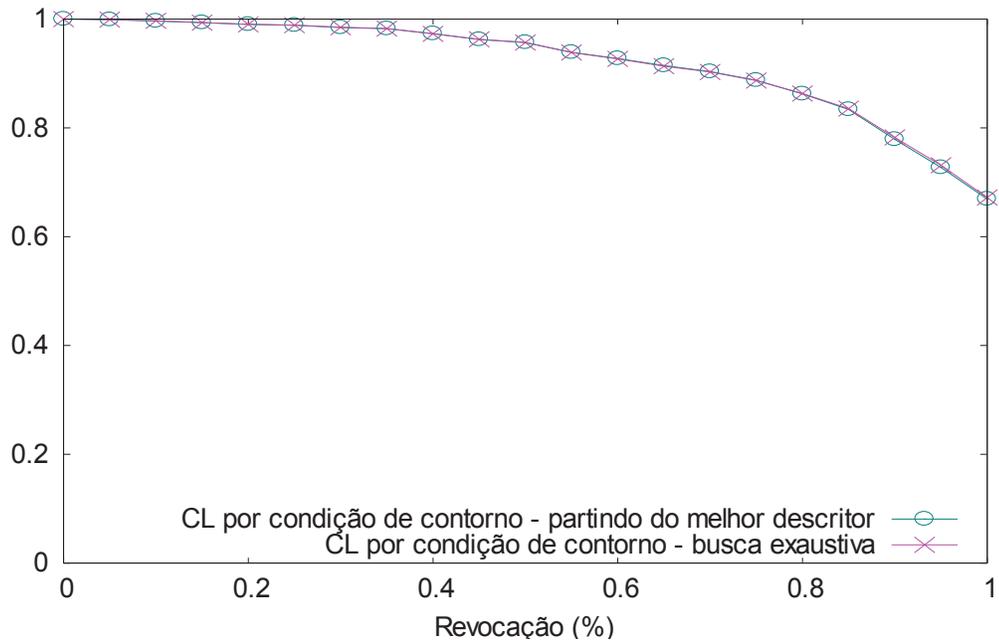


Figura 14. Imagens de ressonância magnética: Curva de Precisão x Revocação para consultas utilizando a combinação de descritores definida por algoritmos de busca exaustiva e a combinação partindo do melhor descritor.

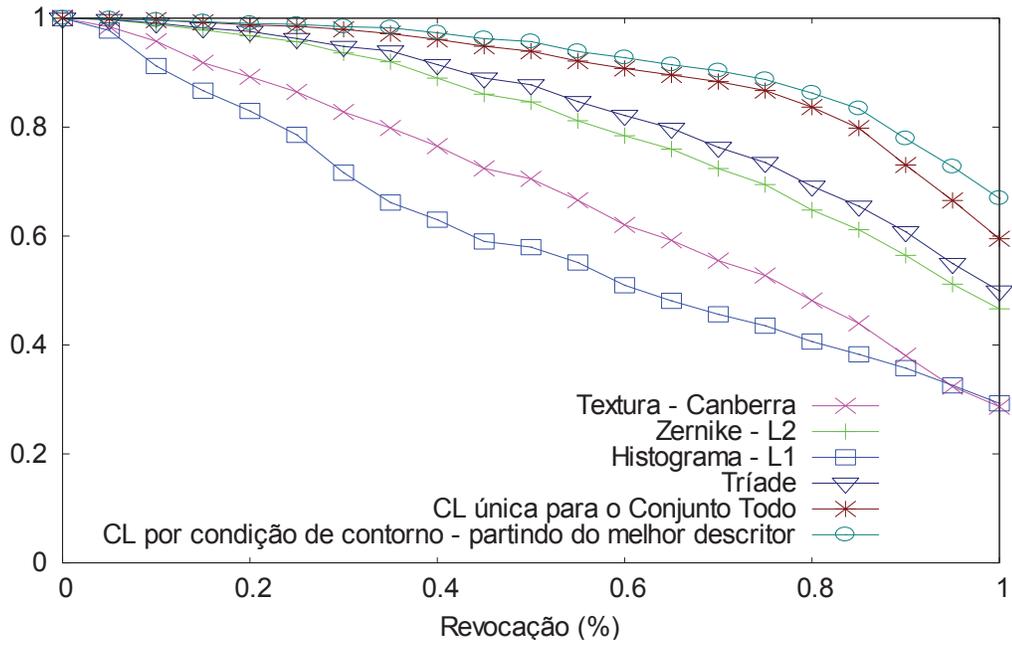


Figura 15. Imagens de ressonância magnética: Curvas de Precisão x Revocação das consultas executadas comparando os resultados das consultas obtidos pela combinação de descritores usando condições de contorno partindo do melhor descritor com outras combinações e descritores individuais.

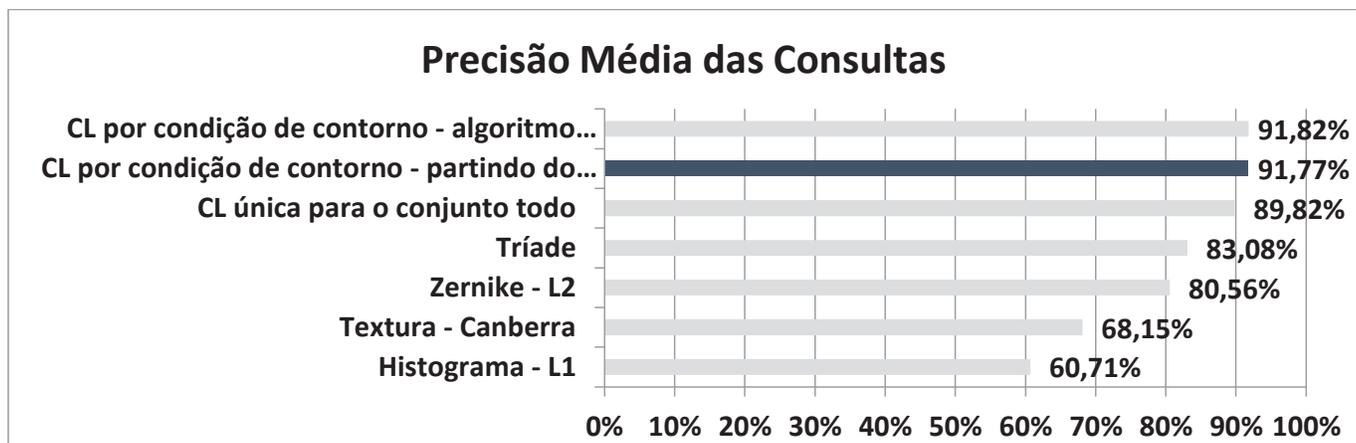


Figura 16. Imagens de ressonância magnética: comparação da precisão média das consultas utilizando a combinação de descritores por condições de contorno partindo do melhor descritor com outras combinações e descritores individuais.

#### REFERÊNCIAS

- [1] C. Morikawa and G. C. de Silva, "User interaction techniques for multimedia retrieval," in *The Joint International Conference on Human-Centered Computer Environments, HCCE '12, Aizu-Wakamatsu, Japan - March 08 - 13, 2012*, 2012, pp. 68–75. [Online]. Available: <http://doi.acm.org/10.1145/2160749.2160765>
- [2] E. Chávez, G. Navarro, R. A. Baeza-Yates, and J. L. Marroquín, "Searching in metric spaces," *ACM Comput. Surv.*, vol. 33, no. 3, pp. 273–321, 2001. [Online]. Available: <http://doi.acm.org/10.1145/502807.502808>
- [3] R. da Silva Torres and A. X. Falcão, "Content-based image retrieval: Theory and applications," *RITA*, vol. 13, no. 2, pp. 161–185, 2006.
- [4] P. H. Bugatti, A. J. M. Traina, and C. T. Jr., "Assessing the best integration between distance-function and image-feature to answer similarity queries," in *Proceedings of the 2008 ACM Symposium on Applied Computing (SAC), Fortaleza, Ceara, Brazil, March 16-20, 2008*, 2008, pp. 1225–1230. [Online]. Available: <http://doi.acm.org/10.1145/1363686.1363969>
- [5] T. M. Deserno, S. Antani, and L. R. Long, "Ontology of gaps in content-based image retrieval," *J. Digital Imaging*, vol. 22, no. 2, pp. 202–215, 2009. [Online]. Available: <http://dx.doi.org/10.1007/s10278-007-9092-x>
- [6] R. Bueno, D. S. Kaster, H. L. Razente, M. C. N. Barioni, A. J. M. Traina, and C. T. Jr., "Using visual analysis to weight multiple signatures to discriminate complex data," in *15th International Conference on Information Visualisation, IV 2011, London, United Kingdom, July 13-15, 2011*, 2011, pp. 282–287. [Online]. Available: <http://doi.ieeecomputersociety.org/10.1109/IV.2011.59>
- [7] J. M. Barrios and B. Bustos, "Automatic weight selection for multi-metric distances," in *Fourth International Conference on Similarity Search and Applications, SISAP 2011, Lipari Island, Italy, June 30 - July 01, 2011*, 2011, pp. 61–68. [Online]. Available: <http://doi.acm.org/10.1145/1995412.1995425>
- [8] B. Bustos, D. A. Keim, D. Saupe, T. Schreck, and D. V. Vranic, "Using entropy impurity for improved 3d object similarity search," in *Proceedings of the 2004 IEEE International Conference on Multimedia and Expo, ICME 2004, 27-30 June 2004, Taipei, Taiwan, 2004*, pp. 1303–1306.
- [9] A. Grigorova, F. G. B. D. Natale, C. K. Dagli, and T. S. Huang, "Content-based image retrieval by feature adaptation and relevance feedback," *IEEE Transactions on Multimedia*, vol. 9, no. 6, pp. 1183–1192, 2007. [Online]. Available: <http://dx.doi.org/10.1109/TMM.2007.902828>
- [10] C. D. Ferreira, J. A. dos Santos, R. da Silva Torres, M. A. Gonçalves, R. C. Rezende, and W. Fan, "Relevance feedback based on genetic programming for image retrieval," *Pattern Recognition Letters*, vol. 32, no. 1, pp. 27–37, 2011. [Online]. Available: <http://dx.doi.org/10.1016/j.patrec.2010.05.015>
- [11] J. Lokoc, D. Novák, M. Batko, and T. Skopal, "Visual image search: Feature signatures or/and global descriptors," in *Similarity Search and Applications - 5th International Conference, SISAP 2012, Toronto, ON, Canada, August 9-10, 2012. Proceedings, 2012*, pp. 177–191. [Online]. Available: [http://dx.doi.org/10.1007/978-3-642-32153-5\\_13](http://dx.doi.org/10.1007/978-3-642-32153-5_13)
- [12] M. Ponciano-Silva, A. J. M. Traina, P. M. Azevedo-Marques, J. C. Felipe, and C. T. Jr., "Including the perceptual parameter to tune the retrieval ability of pulmonary CBIR systems," in *Proceedings of the Twenty-Second IEEE International Symposium on Computer-Based Medical Systems, August 3-4, 2009, Albuquerque, New Mexico, USA, 2009*, pp. 1–8. [Online]. Available: <http://dx.doi.org/10.1109/CBMS.2009.5255399>
- [13] R. F. Barroso, M. Ponciano-Silva, A. J. M. Traina, and R. Bueno, "Using boundary conditions for combining multiple descriptors in similarity based queries," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications - 18th Iberoamerican Congress, CIARP 2013, Havana, Cuba, November 20-23, 2013, Proceedings, Part I, 2013*, pp. 375–382. [Online]. Available: [http://dx.doi.org/10.1007/978-3-642-41822-8\\_47](http://dx.doi.org/10.1007/978-3-642-41822-8_47)
- [14] R. A. Baeza-Yates and B. A. Ribeiro-Neto, *Modern Information Retrieval*. ACM Press / Addison-Wesley, 1999. [Online]. Available: <http://www.dcc.ufmg.br/irbook/>
- [15] A. Khotanzad and Y. H. Hong, "Invariant image recognition by zernike moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 5, pp. 489–497, 1990. [Online]. Available: <http://doi.ieeecomputersociety.org/10.1109/34.55109>
- [16] R. M. Haralick, K. S. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973. [Online]. Available: <http://dx.doi.org/10.1109/TSMC.1973.4309314>