



DETECÇÃO AUTOMÁTICA DE DEDOS E FALANGES EM RADIOGRAFIAS CARPAIS

Bianca Bertoldo de Oliveira*¹ e Milena Bueno Pereira Carneiro²

¹FEELT - Universidade Federal de Uberlândia

Resumo - Nesse trabalho foi desenvolvida uma metodologia para realizar o pré-processamento e identificação de regiões em imagens radiográficas da mão, que possa ser utilizada como primeira etapa de um sistema automatizado de estimação de idade óssea. Apresenta-se os resultados obtidos com algoritmos de realce e binarização, segmentação da mão e dos ossos assim como a detecção automática desses elementos. Com essas informações, é possível implementar diferentes métodos de avaliação da maturidade óssea, seja examinando suas dimensões ou realizando comparações entre imagens do mesmo segmento. É objetivo principal dessa pesquisa identificar a quantidade e localização de dedos na mão e os 14 ossos existentes na região das falanges.

Palavras-Chave- Detecção; Idade óssea; Processamento de imagens; Radiografia; Segmentação.

AUTOMATIC FINGER DETECTION IN X-RAY IMAGES

Abstract - In this work, a methodology was developed to perform the pre-processing and identification of regions in radiographic images of the hand, which can be used as the first step of an automated bone age estimation system. The results obtained with enhancement and binarization algorithms, hand and bone segmentation, as well as the automatic detection of these elements are presented. With this information, it is possible to implement different methods for assessing bone maturity, either examining its dimensions or performing comparisons between images of the same segment. The main objective of this research is to identify the number and location of fingers in the hand and the 14 existing bones in the region of the phalanges.

Keywords - Bone age; Detection; Image processing; Radiography; Segmentation.

I. INTRODUÇÃO

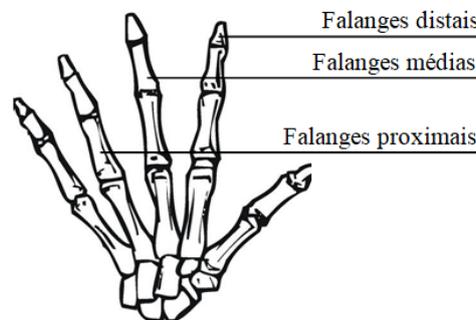
O Processamento Digital de Imagens tem sido muito aplicado na área médica, especialmente no contexto de diagnóstico utilizando resultados de exames de raios-x, ultrassom, tomografia computadorizada e ressonância magnética, possibi-

litando o desenvolvimento de técnicas para análise de forma automática e rápida.

A idade óssea é um método diagnóstico auxiliar na avaliação dos distúrbios do crescimento e da puberdade, uma vez que nem sempre o desenvolvimento do corpo acompanha a real idade cronológica, podendo ser influenciado por fatores genéticos, ambientais e endócrinos.

Assim, o exame de radiografia das mãos é o mais solicitado pelos profissionais, e apesar de teoricamente a estimativa da idade óssea poder ser obtida pela análise de qualquer osso do corpo humano, a mão trata-se de uma região de fácil acesso, não contém órgãos vitais, e pode ser submetida a doses de radiação com maior frequência [1]. A Figura 1 apresenta alguns elementos da estrutura óssea da mão.

Figura 1: Localização das falanges nos ossos pertencentes à mão.



As imagens de radiografias carpais (radiografias de mãos e punhos) são representadas por tons distintos de cinza. Uma imagem pode ser definida como uma função bidimensional $f(x,y)$, onde x e y são coordenadas espaciais representadas por um plano e suas amplitudes em qualquer ponto são chamadas de intensidade ou nível de cinza [2]. Quando as coordenadas e os valores de intensidade são valores finitos e discretos temos uma imagem digital. Para que um computador possa entender uma imagem, ela deve ser transformada em uma matriz de elementos (números), que possuem localização e valor específicos e são conhecidos mais comumente como pixels.

Um dos processos mais simples para isolar o tecido da mão e os ossos são operações realizadas em cima das intensidades que os pixels dessas imagens apresentam. Esse procedimento frequentemente é afetado pela distribuição irregular de inten-

*bianca.bertoldo@ufu.br

sidade dos raios X, ocasionando uma iluminação não uniforme resultada do *efeito Heel* [3]. Essas diferenças fazem com que algumas partes do filme sejam pouco sensibilizadas, enquanto outras sejam sensibilizadas em excesso, prejudicando o contraste necessário para a obtenção de uma boa imagem [4].

Utilizando o procedimento desenvolvido neste artigo, profissionais que trabalham com imagens de radiografias da mão poderão contar com uma ferramenta de pré-processamento, obtendo não apenas uma melhora significativa no contraste e visualização das radiografias, como também identificar as regiões dos dedos e das falanges, onde seria desejável adquirir apenas as informações relacionadas aos contornos, suas localizações, e suas devidas separações entre si.

II. METODOLOGIA

A. Base de dados

Foram utilizadas 100 imagens de radiografias carpais, todas com resoluções e rotações diferentes entre si, assim como níveis de iluminações diversos. Tem-se também posição da mão e afastamento entre os dedos distintos entre cada uma.

As imagens foram obtidas na plataforma Kaggle, e podem ser encontradas aqui: <https://www.kaggle.com/kmader/rsna-bone-age?select=boneage-training-dataset>.

B. Software

A linguagem de programação utilizada nesse trabalho foi o *MatLab*. Esse software lê imagens como matrizes de pixels, tornando a manipulação de imagens bastante fácil. Possui bastante flexibilidade e capacidade de expansão, onde seus chamados *Toolboxes* performam variadas análises em diferentes áreas. Muito útil no processamento digital de imagens, temos as funções da *Image Processing Toolbox*. Por padrão, no caso de imagens em tons de cinza cada pixel pode assumir valores entre 0 e 255, denotando uma imagem de 8 bits. Um pixel de valor 0 é associado ao preto, 255 ao branco e 127 ao cinza propriamente dito.

Para se trabalhar com imagens no *MatLab*, estas devem estar armazenadas em arquivos, sendo que a pasta que contenha estes arquivos deve ser referenciada através da caixa *Current Folder*. Desta forma, em imagens com tons de cinza trabalha-se com matrizes $X(i,j)$ onde i identifica a linha e j a coluna. Para imagens de M linhas e N colunas, os índices são tais que: $1 \leq i \leq M$ e $1 \leq j \leq N$.

C. Processamento das Imagens

1) Equalização de histograma

Para imagens digitais, o histograma é um gráfico que indica a distribuição dos níveis de cinza, sendo representado por uma função discreta $h(r_k) = n_k$, onde n_k é o número de pixels na imagem com intensidade r_k , e r_k é o k -ésimo valor de nível de intensidade. A equalização de um histograma tem a função de reduzir das diferenças acentuadas de intensidade e assim evidenciar detalhes que não estavam visíveis anteriormente [5]. A probabilidade de ocorrência

de cada nível de cinza é dada pela seguinte equação:

$$p_r(r_k) = \frac{n_k}{MN} \quad (1)$$

E a transformação de equalização de histograma pode ser definido por:

$$s_k = (L-1) \sum_{j=0}^k p_r(r_j), \text{ para } k = 0, 1, 2, \dots, L-1 \quad (2)$$

Sendo L a quantidade de níveis de intensidade, a imagem equalizada terá cada pixel da imagem de entrada com intensidade r_k substituída por um pixel correspondente com nível s_k . Esta transformação tende a espalhar o histograma da imagem de entrada, de forma que os níveis de intensidade da imagem equalizado cubram um intervalo maior da escala de intensidade. O resultado final é o realce do contraste.

2) Binarização

A binarização é o método mais simples de segmentação de imagens. Como dito anteriormente, uma imagem digitalizada pode ser definida como uma função $f(x,y)$, e a função binarização pode ser então escrita como:

$$g(x,y) = \begin{cases} R_1 & f(x,y) \leq T \\ R_2 & f(x,y) > T \end{cases} \quad (3)$$

Onde R_1 e R_2 são os valores estipulados para os níveis de cinza da imagem binarizada, que nesse caso foi utilizado 0 e 255, T é o limiar, e $g(x,y)$ a imagem resultante segmentada.

Como cada imagem possui iluminações diferentes, não é interessante, e muito menos viável, obter esse limiar manualmente. Por isso, podemos utilizar métodos de binarização automáticos. O primeiro a ser utilizado será o método de Otsu, facilmente aplicável pelo software. Nesse método, encontra-se o valor limite com a variância mínima para a soma do primeiro plano (objeto) e plano de fundo. O valor limite ótimo é calculado minimizando a soma das variâncias ponderadas do grupo [6].

A segunda forma de segmentação é do tipo adaptativa, onde escolhe-se o limite com base na intensidade média local (estatísticas de primeira ordem) na vizinhança de cada pixel. Em complemento à esse método, podemos indicar se o objeto é mais claro ou escuro que o fundo.

3) Transformação de negativo

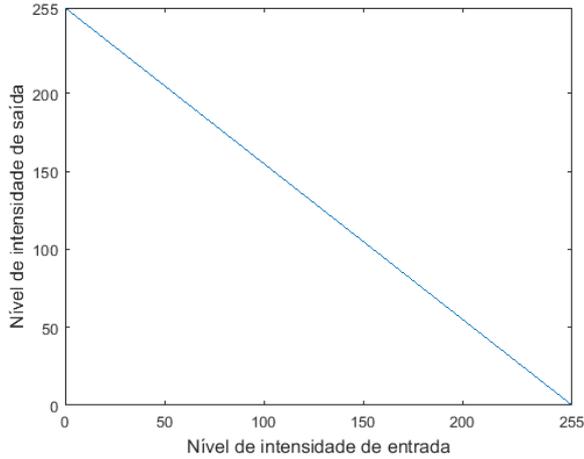
O negativo de uma imagem com 256 níveis de cinza é obtido utilizando a transformação de negativo, nesse caso, pixels com intensidade 0 passam a ter intensidade 255 e vice-versa, resultando em uma função definida por uma reta. Esse tipo de transformação de intensidade se equivale ao negativo fotográfico, e seu uso é útil para realçar detalhes com valores de intensidade maiores quando estão envoltos por uma área escura dominante, ajudando a analisar essas regiões. A expressão para a transformação de negativo pode ser definida como:

$$s = L - 1 - r \quad (4)$$

- s - Nível de intensidade de saída.
- L - Quantidade de níveis de intensidade.
- r - Nível de intensidade de entrada.

A representação gráfica é mostrada a seguir na Figura 2.

Figura 2: Curva da função de transformação de negativo.



4) Detecção das regiões

Sendo um pixel p na coordenada (x, y) com quatro vizinhos horizontais e verticais cujas coordenadas são dadas por:

$$(x + 1, y), (x - 1, y), (x, y + 1), (x, y - 1)$$

e também quatro vizinhos diagonais com coordenadas:

$$(x + 1, y + 1), (x + 1, y - 1), (x - 1, y + 1), (x - 1, y - 1)$$

Esses vizinhos (pontos) juntos são a definição de vizinhança-8 de p , expressos por $N_8(p)$. Tendo um subconjunto S de pixels em uma imagem, dois pixels p e q são conexos em S se existir um caminho entre eles inteiramente de pixels em S , e para qualquer pixel p em S , o conjunto de pixels conectados a ele em S é chamado de um componente conexo de S . Se existir apenas um componente conectado, o conjunto S é chamado de um conjunto conexo. Dessa forma, sendo R um subconjunto de pixels em uma imagem, R será uma região da imagem se for um conjunto conexo [2]. As regiões a serem detectadas nas imagens serão identificadas utilizando a adjacência-8 automaticamente empregando funções do próprio software.

5) Filtro Gaussiano de suavização

Um filtro de suavização é usado para remover ruído de imagens de uma forma a não perder informações importantes dela. O filtro de suavização Gaussiano consiste em convoluir uma máscara (ou kernel) com os pixels da imagem, onde a máscara é antes espelhada (rotacionada em 180°), começando com o primeiro pixel no canto superior esquerdo e deslizando para o lado até realizar todas as operações de média da soma dos produtos em cada local.

Considerando uma imagem $f(x, y)$ de tamanho $M \times N$, com $w(x, y)$ representando a máscara, tem-se a seguinte equação

de convolução [7]:

$$w(x, y) * f(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b w(s, t) f(x-s, y-t) \quad (5)$$

Onde,

$$a = \frac{M-1}{2} \quad e \quad b = \frac{N-1}{2} \quad (6)$$

Nas máscaras de tamanho par ($2 \times 2, 4 \times 4, \dots$) o resultado da operação em cada ponto é aferido sobre o primeiro pixel, e para tamanho ímpar ($3 \times 3, 5 \times 5, \dots$) o resultado é colocado sobre o pixel de centro. Os coeficientes das máscaras são positivos e a soma deles é sempre igual a 1.

O uso do filtro se mostra muito útil antes de realizar detecção de bordas, pois a redução de ruído suaviza traços não necessários, aperfeiçoando os resultados.

6) Detecção de bordas de Sobel

A detecção de bordas é um processo operacional para detectar o limite entre duas regiões homogêneas da imagem com diferentes níveis de brilho. O operador de Sobel é um método que usa máscaras de tamanho 3×3 para calcular os gradientes horizontais, verticais e diagonais em cada posição de pixel da imagem, e essas derivadas parciais são utilizadas para estimar a intensidade e a direção da borda. Os coeficientes de todas as máscaras somam zero, gerando uma resposta nula nas áreas de intensidade constante. As máscaras de Sobel são definidas como:

$$H = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad V = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

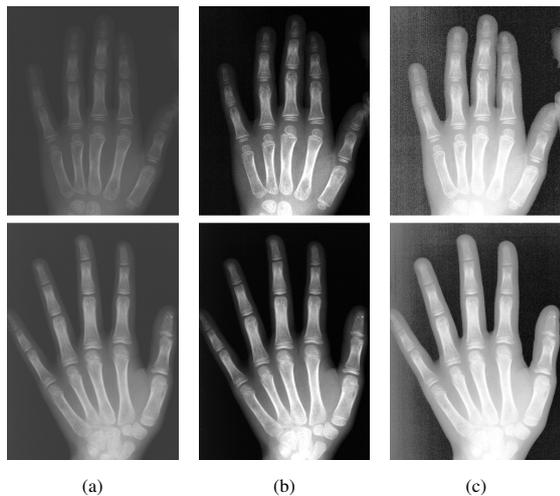
$$45^\circ E = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix} \quad 45^\circ D = \begin{bmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

D. Descrição do Algoritmo

Para esse trabalho, o objetivo é conseguir segmentar dedos da mão e os ossos da região das falanges. Sendo assim, o primeiro passo do programa é realizar um corte na parte inferior da imagem. Esse corte é interessante primeiramente pelo efeito Heel presente em algumas radiografias, o que faz com que os ossos dessa região inferior, ou seja, do pulso, tenha tons muito mais claros do que os ossos dos dedos, dificultando o processo de normalização e realce, e posteriormente dificultando também o processo de binarização adequada. Como todas as imagens possuem dimensões diferentes, para realizar esse corte, foi definido que deve-se manter 80% do número de linhas.

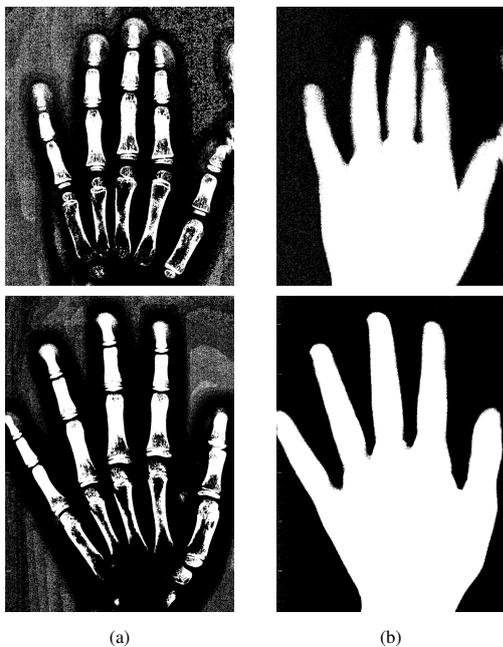
Em seguida é realizado um ajuste automático, mapeando os valores de intensidade e saturando o 1% inferior e o 1% superior de todos os valores de pixel, aumentando o contraste da imagem de saída. Com a imagem ajustada, o próximo passo é equalizar o histograma, o que ajuda a destacar as regiões pertencentes à mão e ossos. A Figura 3 mostra o resultado da aplicação desses três passos: corte, ajuste e equalização do histograma, em duas imagens do banco de dados.

Figura 3: a) Imagens originais cortadas; b) Imagens ajustadas; c) Imagens com histograma equalizado.



Então são definidas duas imagens diferentes, a primeira (b) apenas ajustada, e a segunda (c) ajustada e com histograma equalizado. Essa diferenciação irá propiciar duas binarizações diferentes, uma focada em detectar apenas os dedos, e a outra irá localizar os ossos. Sabendo que imagens diferentes possuem iluminações diferentes, não é interessante, e muito menos viável, obter o limiar de binarização manualmente. Por isso, podemos utilizar os métodos de binarização automáticos: Adaptativo, para as imagens presentes na Figura 3.b, e Otsu, para a Figura 3.c. Na Figura 4 tem-se as imagens com os efeitos desse procedimento.

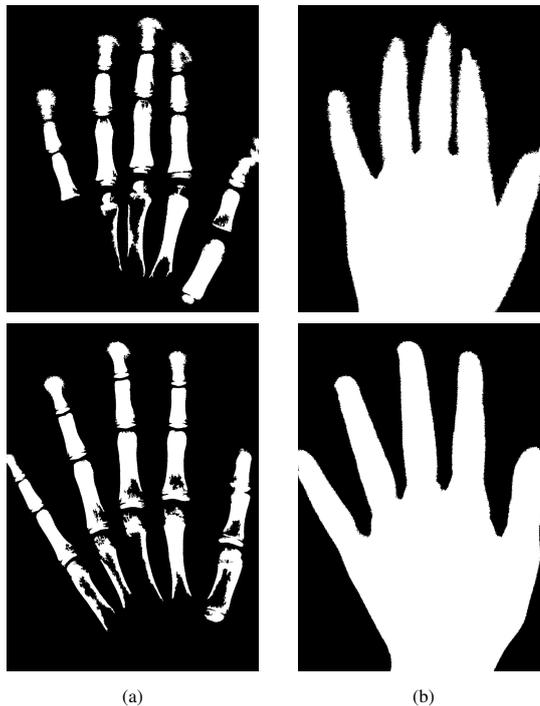
Figura 4: a) Binarização Adaptativa; b) Binarização por Otsu.



Para as duas formas de binarização diferentes, são empregadas funções do software para preencher pequenos buracos (brancos ou pretos) presentes no fundo e no objeto segmentado, e depois, apagar pequenas áreas, realizando uma limpeza na imagem. A Figura 5 apresenta as imagens binarizadas após

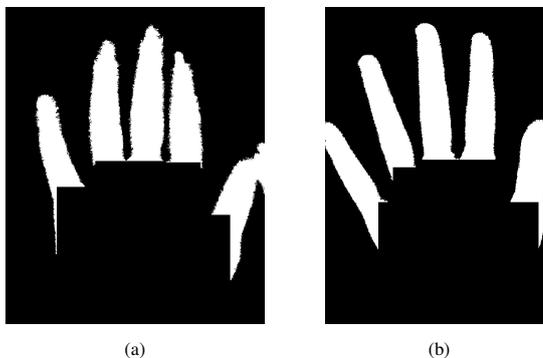
essas operações.

Figura 5: a) Imagens dos ossos após o processamento; b) Imagens das mãos após o processamento.



Em seguida, para realizar a segmentação dos dedos, é definido um elemento de estruturação morfológica plana em forma de quadrado, com tamanho aproximado da palma da mão. Essa estrutura é necessária para realizar as operações de erosão, onde as regiões com área menor serão apagadas, ou seja, os dedos, e de dilatação, para recuperar partes da estrutura que possam ter sido apagadas. Dessa forma, ao subtrair o resultado dessa operação da imagem binarizada contendo a mão inteira, obteremos uma imagem composta apenas pelos dedos, podendo utilizar a função de remover áreas pequenas para apagar qualquer resquício da palma, como é mostrado na Figura 6.

Figura 6: a) e b) Imagens binarizadas das mãos sem a palma.



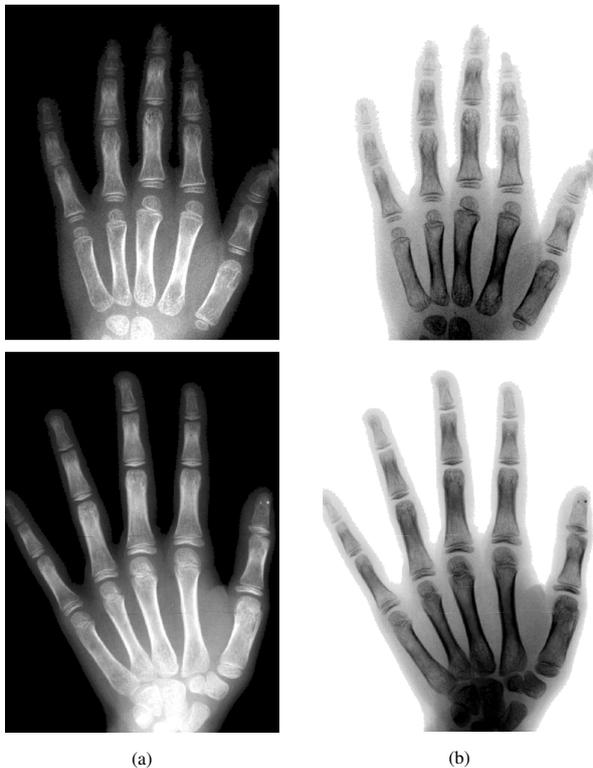
Obtendo então as duas imagens com os dedos e os ossos das falanges segmentados, usa-se uma função de contagem de regiões 8-conectadas, que calcula os valores de área, centróide, caixa delimitadora ao redor desse elemento e os valores de in-

tensidade dos pixels. A quantidade de objetos detectados fica salvo em um vetor, e após terminar de detectar todos, podemos obter o tamanho desse vetor e pedir para que se desenhe as caixas retangulares para cada posição.

Em seguida, com os valores de área de cada região, calcula-se as centróides e é possível posicionar a enumeração de cada região. Com essas enumerações, se desejável, consegue-se extrair cada componente detectado e exibir em uma nova imagem apenas indicando seu índice para o algoritmo.

Agora, considerando a imagem binarizada completa da mão como vimos na Figura 5.b, podemos utilizá-la como uma máscara, e multiplicar seus valores pelos valores da imagem inicial apenas ajustada. A partir disso, teremos em evidência apenas a mão e os ossos, com um fundo completamente preto. Podemos então aplicar a transformação de negativo na imagem resultante, e com os valores dos pixels invertidos, teremos um destaque maior para o objeto em si, demonstrado na Figura 7 a seguir.

Figura 7: a) Imagem ajustada com máscara; b) Imagem com transformação de negativo.



Apesar do algoritmo de detecção exigir que as imagens sejam binárias para extrair cada região, é possível apresentar essas regiões em qualquer imagem de mesmo tamanho, e assim, utiliza-se então a imagem transformada em negativo.

Além de obter as posições das regiões, uma possibilidade a mais pode ser obter o contorno dessas regiões, e essa operação pode ser feita aplicando o método de detecção de bordas de Sobel. Primeiro, para delimitar os ossos, é possível multiplicar a imagem original pela imagem apenas dos ossos binarizados. Depois, binarizamos novamente esse resultado e aplicamos uma máscara de suavização Gaussiana obtida na

literatura [2] definida como:

$$\text{máscara} = \begin{bmatrix} 0.0625 & 0.125 & 0.0625 \\ 0.125 & 0.25 & 0.125 \\ 0.0625 & 0.125 & 0.0625 \end{bmatrix}$$

Posteriormente, adota-se o operador de Sobel para obter as bordas. Os resultados obtidos com a detecção serão exibidas e discutidas na próxima seção.

III. RESULTADOS

Como explicado anteriormente, os dedos e falanges foram identificadas utilizando função de identificação de regiões 8-conectadas. Para evidenciar as regiões detectadas pelo algoritmo, foram exibidas duas imagens diferentes, uma apenas com a detecção dos dedos, e outra com a detecção das falanges. Dessa forma, para as imagens exposta nesse artigo, tem-se os resultados a seguir nas Figuras 8 e 9:

Figura 8: a) e b) Detecção dos dedos.

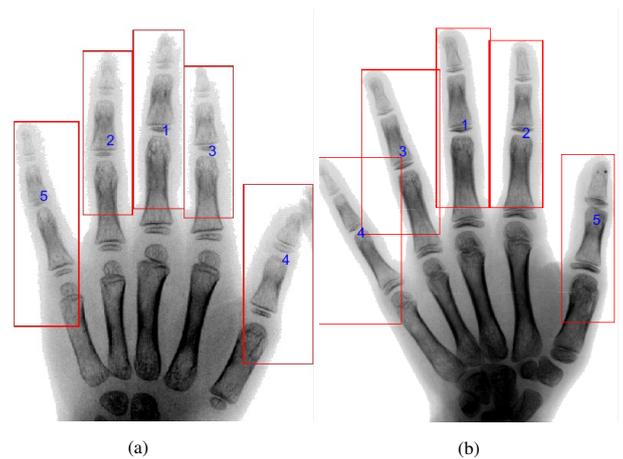
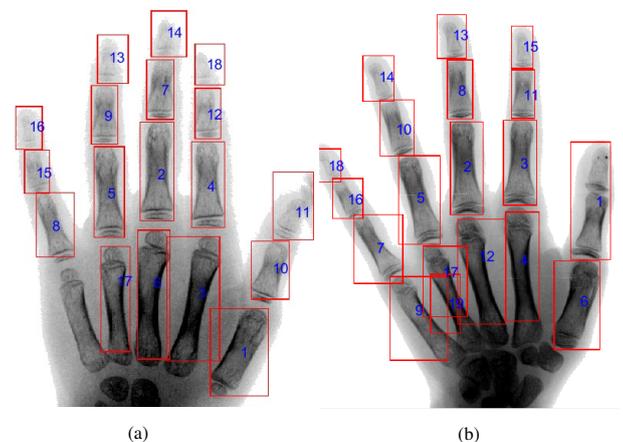
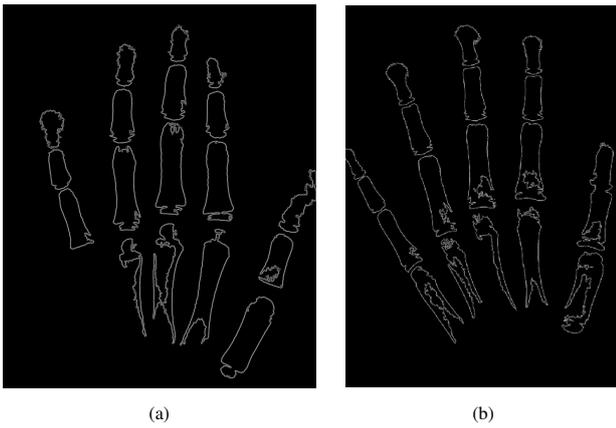


Figura 9: a) e b) Detecção das falanges.



O resultado da detecção de bordas se encontra na Figura 10, com as falanges bem definidas:

Figura 10: a) e b) Detecção de bordas dos ossos.



Ambas as determinações apresentaram bons resultados, considerando que o algoritmo realiza todas as funções de forma totalmente automática, precisando que o usuário entre apenas com o nome do arquivo. A relação entre a quantidade de elementos que deveriam ser identificados e as detecções obtidas foi calculada utilizando o MAPE (Erro Percentual Absoluto Médio, em português) geral, uma medida de precisão que diz quanto, em média, foi o erro, nesse caso, nas detecções, sem compensar erros negativos com erros positivos [8]. Pode ser definido pela fórmula:

$$MAPE = \frac{\sum |Detectados - Atual|}{\sum Atual} \quad (7)$$

Para a primeira etapa de identificação dos dedos, foi obtido um MAPE geral de 8,13% de erro, e para a segunda etapa foi calculado um erro MAPE de 18,73%. Como interpretação destes resultados, valores menores de 10% são considerados detecções altamente precisas e entre 10% e 20% boas detecções [9].

Melhores resultados seriam obtidos com imagens que possuem iluminação uniforme e adequada, porém, caso seja necessário, o algoritmo pode ser facilmente manipulado para atender outros requisitos. Os valores e percentuais apresentados demonstram a validade dos métodos de processamento de imagens implementados.

IV. CONCLUSÕES

Foram utilizados nesse trabalho diversas técnicas de processamento de imagens diferenciadas com o intuito de obter

as melhores determinações possíveis, sem a necessidade de grande custo computacional. Melhores resultados seriam alcançados com imagens que possuem iluminação uniforme e adequada, porém, levando em consideração as imagens trabalhadas, bons resultados foram obtidos.

O objetivo de implementar um sistema automático de processamento e detecção em imagens de radiografias carpais foi atingido e vale destacar que, caso seja necessário, o algoritmo poderia ser facilmente manipulado para atender outros requisitos.

A manipulação e pré-processamento dessas radiografias são de grande importância para construir sistemas de diagnóstico automatizados, levando em conta que o desenvolvimento ósseo é considerado uma das principais referências para estimar a maturação fisiológica em humanos.

REFERÊNCIAS

- [1] K. L. Bontrager, *Tratado de Técnica Radiológica e Base Anatômica*, Guanabara Koogan, 4ª Edição, Rio de Janeiro, 1999.
- [2] R. C. Gonzalez, R. E. Woods, *Digital Image Processing*, Pearson, 3ª Edição, 2007.
- [3] G. Behiels, et al, "Retrospective Correction of the Heel Effect in Hand Radiographs", in *Medical Image Analysis*, vol. 06, pp. 183-190, 2002.
- [4] P. Gratale, D. L. Wright, L. Daughtry, "Using the anode heel effect for extremity radiography", in *Radiologic Technology*, vol. 61, no. 3, pp. 195-198, 1990.
- [5] W. K. Pratt, *Digital Image Processing*, John Wiley and Sons, 3ª Edição, 2001.
- [6] J. Yousefi, *Image Binarization using Otsu Thresholding Algorithm*, DO 10.13140, RG.2.1.4758.9284, 2015.
- [7] A. Oppenheim, A. S. Willsky, S. H. Nawab, *Signals and Systems*, Prentice Hall, 2ª Edição, 1997.
- [8] Myttenaere, B. Golden, B. Le Grand, F. Rossi, "Mean absolute percentage error for regression models", in *Neurocomputing*, 2016.
- [9] C. D. Lewis, *International and Business Forecasting Methods*, Butterworths, Londres, 1982.