

Normalização Espacial de Imagens Frontais de Face em Ambientes Controlados e Não-Controlados

Vagner do Amaral¹, Claudia Fígaro-Garcia², Gilka Jorge Figaro Gattas³ e Carlos Eduardo Thomaz⁴

Resumo:

Este artigo propõe a implementação de um processo de normalização espacial de imagens de face geradas em ambientes controlados e não-controlados. Para isso, desenvolveu-se um arcabouço computacional capaz de reduzir os desvios gerados no processo de captura de imagens. Os resultados obtidos demonstram um aumento de nitidez em pontos específicos das faces médias e o aumento da correlação entre as faces de cada conjunto analisado, justificando a aplicação dos métodos descritos.

Palavras-chave: Normalização Espacial de Imagens; Banco de Faces.

Abstract:

This paper proposes and implements a step-by-step image processing tool to spatially normalize frontal face images. A computational framework has been developed to minimize variations of not only well-framed face images, but also images captured under distinct acquisition protocols. Our experimental results show an increase of sharpness in specific points of the average face images and higher cross-correlation between the normalized face images compared to the original ones, in controlled and non-controlled environments.

Keywords: Spatial Normalization of Images; Face Database.

¹ Vagner do Amaral é Tecnólogo em Análise e Desenvolvimento de Sistemas (Mestrando em Eng. Elétrica) e atua no Centro Universitário da FEI.

² Claudia Figaro-Garcia é Mestre em Psicologia (Doutoranda em Psicologia) e atua na Faculdade de Medicina da Universidade de São Paulo.

³ Gilka Jorge Figaro Gattas é Livre Docente Doutora em Ciências Biológicas e atua na Faculdade de Medicina da Universidade de São Paulo.

⁴ Carlos Eduardo Thomaz é Doutor em Ciência da Computação e atua no Centro Universitário da FEI.

1. Introdução

A normalização espacial de imagens ou, simplesmente, alinhamento de imagens é uma importante etapa de pré-processamento de dados em qualquer estudo de reconhecimento de padrões em imagens.

A ideia básica de tal metodologia é remover efeitos ou sinais da imagem que não sejam relevantes e que retratem, na verdade, particularidades indesejadas inerentes a todo processo imperfeito de aquisição de dados. No contexto desse trabalho, essa metodologia se resume ao mapeamento automático de imagens de faces, capturadas em ambientes controlados e não-controlados, para um sistema comum de coordenadas de tal forma que os pixels descrevam idealmente localidades idênticas em todas imagens.

Normalmente, o alinhamento de imagens é realizado através de um processo de deformação de cada uma dessas imagens com relação a uma imagem referência, ou pontos específicos da imagem utilizados como referência. Tal procedimento de deformação tem essencialmente duas metas: (a) reduzir a variabilidade das imagens devido a diferenças referentes ao tamanho, posição e orientação do padrão a ser analisado; (b) reduzir a variabilidade das imagens devido a diferenças na forma do padrão (RUECKERT *et al.*, 1999).

Neste artigo, para normalizar espacialmente cada imagem de face, algoritmos de deformações globais serão investigados, baseados nos procedimentos propostos em Amaral e Thomaz (2008). A deformação ou transformação global modela basicamente diferenças de rotação, translação e escala, existentes entre a imagem e os pontos de referência pré-definidos. O objetivo deste artigo é validar essas deformações não somente em ambientes de captura controlada, conforme proposto em Amaral e Thomaz (2008), mas também em ambientes onde as características de aquisição das imagens tais como posição, iluminação e distância, são casuais e em alguns casos precárias.

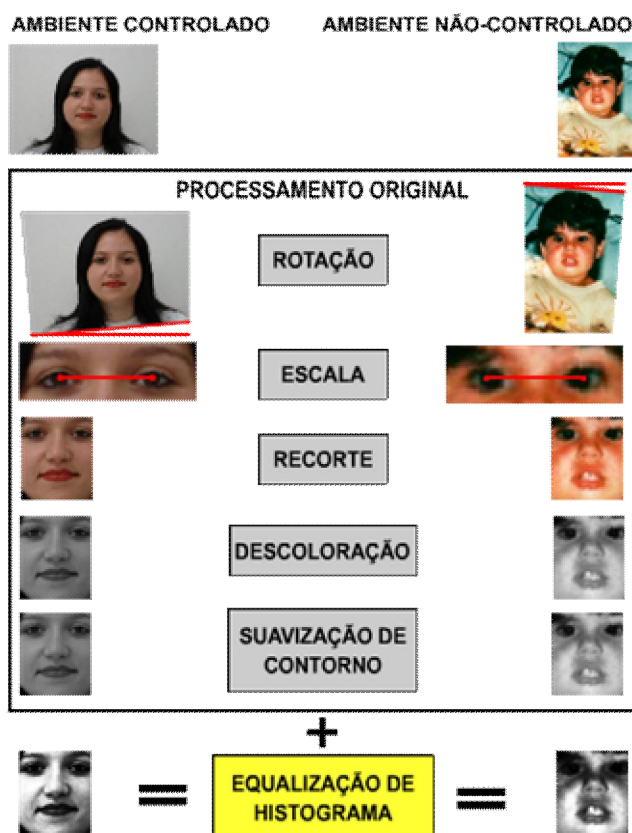
2. Metodologia

Neste trabalho utilizou-se a metodologia apresentada por Amaral e Thomaz (2008) para o alinhamento das imagens, que fundamenta-se na proposta de interação mínima do

usuário, requerendo apenas a definição das dimensões da imagem (em unidades de pixel) e a identificação manual da posição das pupilas como pontos de referência.

A metodologia de normalização espacial de imagens de face sugerida está subdividida em etapas de processamento dedicado, que tratam de desvios específicos de faces frontais e obedecem à seguinte sequência pré-definida: rotação, escala, recorte, descoloração e suavização de contorno. Os procedimentos originais propostos em Amaral e Thomaz (2008) foram preservados, entretanto, as dimensões de escala e recorte foram alteradas e ao final do processamento foi aplicada uma equalização de histograma (GONZALES; WOODS, 2000, p. 124). As subseções seguintes descrevem de forma sucinta os objetivos e benefícios das etapas de processamento e apresentam em detalhes as alterações propostas. O fluxograma dessas etapas de processamento é ilustrado na Figura 1.

Figura 1: Etapas de processamento para normalização espacial de imagens de face.



2.1. Rotação

A etapa de rotação inicializa a padronização da posição da face dentro da imagem e define os parâmetros para execução das etapas seguintes. O objetivo desta etapa é fundamentalmente alinhar a posição das pupilas com relação ao eixo horizontal, utilizando as medidas de seno e cosseno do ângulo θ que descreve a inclinação da reta imaginária que passa através dos pontos de referência (AMARAL; THOMAZ, 2008).

Pode-se utilizar para este processo um método de cálculo do ponto de destino (x_d, y_d) de um pixel através de seu ponto de origem (x_o, y_o) , similar a abordagem apresentada em (CONCI; AZEVEDO; LETA, 2008, p. 120), através da seguinte equação:

$$\begin{aligned}x_d &= x_o \cos\theta + y_o \sin\theta \\y_d &= -x_o \sin\theta + y_o \cos\theta\end{aligned}\tag{1}$$

Entretanto, essa abordagem gera na imagem destino um ruído conhecido como impulsivo (também denominado de “sal e pimenta”) (CONCI; AZEVEDO; LETA, 2008, p.173), pois o endereço (x_d, y_d) dos pixels é formado por números inteiros e o arredondamento de um valor decimal através da equação (1) pode gerar sobreposição de pixels na imagem destino e deixando-a, assim, com alguns pontos sem preenchimento (AMARAL; THOMAZ, 2008). Este tipo de ruído pode ser eliminado de forma eficaz através de um filtro de mediana (CONCI; AZEVEDO; LETA, 2008, p.173), porém, isso acarretaria em mais uma etapa de processamento. Portanto, utilizou-se uma abordagem inversa descrita em Amaral e Thomaz (2008), onde cada ponto na imagem destino identifica seu ponto correspondente na imagem origem, evitando a sobreposição de pixels e ruídos.

O método inverso foi desenvolvido a partir da equação (1), utilizando uma técnica conhecida como rotação em duas passadas (CONCI; AZEVEDO; LETA, 2008, p.120), onde o endereço de um dos eixos origem é identificado inicialmente para compor a equação que resultará no endereço do segundo eixo (AMARAL; THOMAZ, 2008). Em outras palavras, utilizando a equação (1), uma das variáveis correspondentes às coordenadas de origem pode ser isolada, ou seja

$$x_d = x_o \cos \theta + y_o \sin \theta \quad \therefore \quad x_o = \frac{x_d - y_o \sin \theta}{\cos \theta}, \quad (2)$$

onde $\cos \theta \neq 0$, e aplicada na equação da segunda variável:

$$y_d = -\left(\frac{x_d - y_o \sin \theta}{\cos \theta}\right) \sin \theta + y_o \cos \theta \quad \therefore \quad y_o = \frac{-\sin \theta \left(\frac{x_d}{\cos \theta}\right) - y_d}{\sin \theta \left(\frac{-\sin \theta}{\cos \theta}\right) - \cos \theta}. \quad (3)$$

Para finalizar a técnica de rotação em duas passadas, o resultado obtido na equação (3) de y_o é aplicado na equação (2) de x_o . Os resultados deste método para (x_o, y_o) correspondem ao endereço original do pixel que será alocado na posição (x_d, y_d) da imagem destino (AMARAL; THOMAZ, 2008). Dessa forma, a rotação é aplicada sem a perda de informações da imagem original, mas pode gerar um efeito de serrilhado na imagem que será tratado na penúltima etapa do processamento, ou seja, na etapa de suavização de contorno descrita na subseção 2.5

2.2. Escala

A alteração de escala nas imagens de face é derivada da padronização da distância entre os pontos de referência (pupilas), que visa corrigir desvios causados durante a captura e aumentar a correspondência dos pixels entre as imagens de face processadas (AMARAL; THOMAZ, 2008).

Esta etapa identifica a coordenada de um pixel na imagem original através da multiplicação de cada pixel da imagem destino por um valor definido pela razão entre a distância interpupilar original e uma distância padrão pré-definida, conforme exemplo apresentado na Figura 2. As dimensões da imagem final são especificadas aplicando a multiplicação pela razão inversa, ou seja, a distância padrão pela distância interpupilar (AMARAL; THOMAZ, 2008).

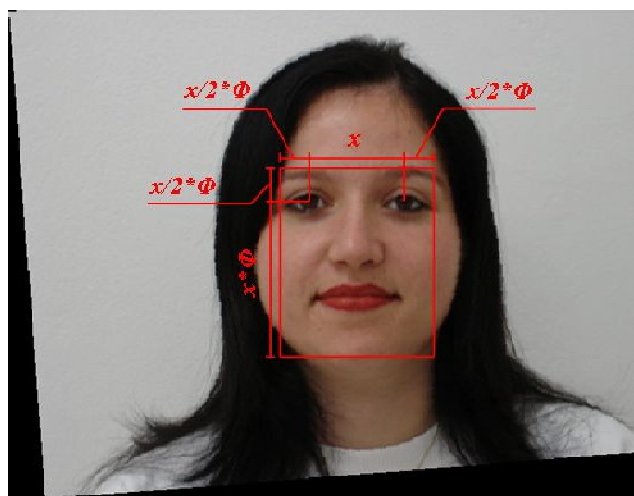
Figura 2: Distância interpupilar original (esq.) e ajustada para 100 pixels (dir.).



2.3. Recorte

Esta etapa é essencial à normalização, pois elimina inúmeros artefatos desnecessários e preserva as informações e características mais relevantes aos estudos sobre biometria de faces frontais. Utilizou-se neste trabalho a razão áurea (ALVES, 2008)(MORIHISA, 2006), também conhecida como “número de ouro” e representada pela letra Φ , onde $\Phi = 1,618$, para estabelecer uma relação entre as medidas nas imagens frontais de faces. A Figura 3 ilustra este processo. Manteve-se para a distância interpupilar o valor de 100 pixels, também utilizada em Amaral e Thomaz (2008), ajustando-se as outras medidas experimentalmente e de acordo com a proporção ilustrada na Figura 3. Aplicou-se, com relação a posição da pupila, a proporção $x/2*\Phi$ para as margens laterais e superior e a proporção $2*\Phi$ para margem inferior, gerando imagens com resolução de 162x193 pixels que representam essencialmente a face das pessoas, excluindo, por exemplo, a testa, o cabelo e o pescoço.

Figura 3: Relação das medidas de recorte, onde $x = 100$ pixels e $\Phi = 1,618$.



2.4. Descoloração

A etapa de descoloração foi incluída na normalização visando reduzir o esforço computacional, através da substituição do sistema RGB que trabalha com 3 (três) canais de cores (Red, Green and Blue) para um sistema de uma única camada utilizando a escala de cinza. Este processo é representado pela seguinte equação:

$$(x_d, y_d) = 0.299(x_o, y_o) + 0.587(x_o, y_o) + 0.114(x_o, y_o). \quad (4)$$

As constantes 0.299, 0.587 e 0.114 correspondem ao modelo YIQ (Luminance, In-phase and Quadrature), projetado com base no percentual de sensibilidade da visão humana (GONZALES; WOODS, 2000, p.162). Portanto, esta expressão fornece um resultado mais adequado que a divisão proporcional, com 33% para cada cor (AMARAL; THOMAZ, 2008).

2.5. Suavização de Contorno

A técnica de suavização de contorno é aplicada para eliminar o efeito de serrilhado gerado pela etapa inicial de rotação, tornando a imagem mais similar à original após esse processamento. Utilizou-se nesta etapa um janelamento (ou kernel) de 3x3 pixels, para substituir o tom do pixel central pela mediana dos pixels ao seu redor. Conforme exemplo apresentado em Amaral e Thomaz (2008), dada a seguinte matriz 3x3:

$$\begin{pmatrix} 101 & 110 & 196 \\ 95 & 111 & 115 \\ 89 & 99 & 107 \end{pmatrix},$$

os valores contidos no kernel são ordenados e o valor central do conjunto se refere à mediana, conforme conjunto abaixo:

$$[89; 95; 99; 101; 107; 110; 111; 115; 196].$$

No exemplo dado, o valor da mediana é o quinto elemento do conjunto ordenado e possui valor 107, que deverá substituir o antigo valor central 111 da matriz.

2.6. Equalização de Histograma

Segundo Conci, Azevedo e Leta (2008), a equalização de histograma tem como objetivo melhorar o contraste das imagens através da representação dos níveis de cinza de

maneira uniforme e distribuída. Portanto, ao final do processamento de normalização espacial de imagens foi acrescentada esta técnica visando otimizar imagens de face com baixo contraste, muito comuns em ambientes não-controlados onde a fonte de iluminação é aleatória e muitas vezes precária. Outro benefício que esta técnica proporciona aos estudos biométricos se deve ao realce de rugas e outros detalhes da face, aumentando assim a distinção entre faces de indivíduos diferentes.

Para a aplicação desta técnica, calcula-se inicialmente um histograma, normalizado entre 0 e 1 (CONCI; AZEVEDO; LETA, 2008, p.95), que define a probabilidade original P de distribuição dos níveis de cinza presentes na imagem através da seguinte equação:

$$P(k) = \frac{n_k}{l \times c}, \quad (5)$$

onde n_k é a quantidade de ocorrências do nível de cinza k , e l e c são os números de linhas e colunas da imagem, ou seja, $l \times c$ é igual ao número total n de pixels. Em seguida, para obter o histograma acumulado normalizado utiliza-se a seguinte função de distribuição:

$$T(k) = \sum_{i=0}^k \frac{n_i}{n}. \quad (6)$$

O resultado desta função descrita na equação (6) gera uma escala probabilística uniformemente distribuída entre 0 e 1.

De acordo com Gonzales e Woods (2000), a conversão dos níveis de cinza para valores probabilísticos é realizada através da divisão de cada nível pelo número de valores possíveis. Portanto, a conversão da escala probabilística para níveis de cinza é obtida através da operação inversa e o nível de cinza equalizado de um pixel qualquer u da imagem original é dado por:

$$u_e = T(u_o) * b, \quad (7)$$

onde u_e é o pixel equalizado, u_o é o pixel original, e o parâmetro b é o número de valores possíveis na escala de cinza. Nas imagens utilizadas neste trabalho, $b = 256$. O efeito deste procedimento nos ambientes controlado e não-controlado analisados pode ser verificado na Figura 4.

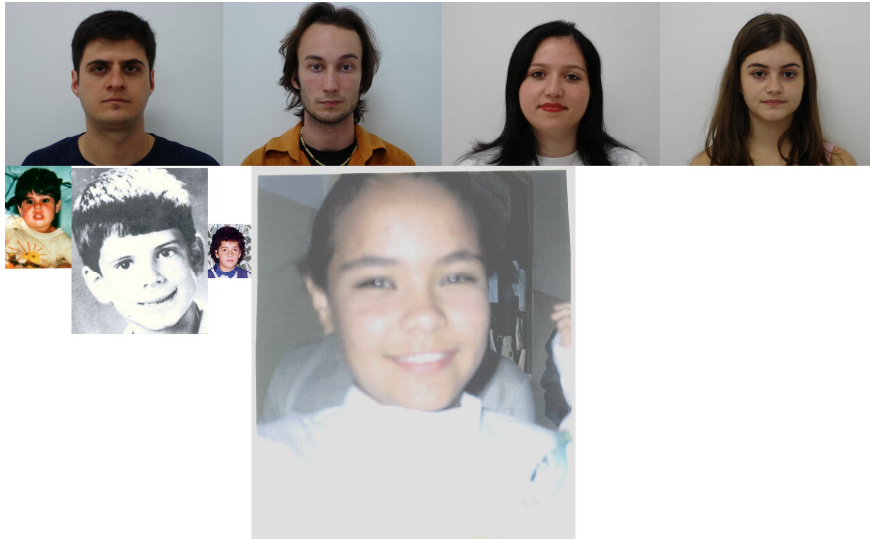
Figura 4: Imagem com histograma original e equalizada para exemplos de faces capturadas em ambientes controlado (esquerda) e não-controlado (direita).



3. Bancos de Imagens de Face

Neste trabalho foram empregadas imagens de duas fontes distintas, uma em ambiente controlado e outra em ambientes não-controlados. As imagens geradas em ambiente controlado compõem um banco de imagens frontais de faces, criado e mantido pelo Departamento de Engenharia Elétrica do Centro Universitário da FEI e composto de 2800 imagens com resolução de 640x480 pixels, capturadas de acordo com o procedimento descrito no relatório desenvolvido por Oliveira Junior e Thomaz (2008). Entretanto, neste trabalho utilizaram-se somente as imagens frontais neutras, totalizando 200 imagens. As imagens geradas em ambientes não-controlados são procedentes de cadastros de crianças e adolescentes desaparecidos, disponíveis publicamente no site da Redesap (Rede Nacional de Identificação e Localização de Crianças e Adolescentes Desaparecidos), em diversas resoluções, poses e níveis de exposição à luz. Para este artigo utilizaram-se 125 faces provenientes de ambientes não-controlados. A Figura 5 ilustra alguns exemplos dessas duas fontes distintas de imagens usadas.

Figura 5: Exemplos de imagens utilizadas neste trabalho: ambiente controlado (em cima); ambiente não-controlado (embaixo).



4. Implementação

Os processos apresentados foram implementados em um aplicativo desenvolvido na linguagem C#, através da ferramenta Microsoft Visual C# 2008 Express Edition, utilizando apenas os recursos disponíveis na .NET Framework até a versão 3.5. O programa foi compilado e executado em um PC Intel Dual-Core 1.86 GHz, 2 GB de RAM e 160 GB de HD, rodando sistema operacional Windows XP Service Pack 3 (AMARAL; THOMAZ, 2008).

5. Resultados Obtidos

Os resultados de normalização espacial descritos neste artigo foram quantificados através do cálculo da correlação média entre os conjuntos de faces analisados (AMARAL; THOMAZ, 2008). Utilizou-se para esta comparação o módulo de correlação em pares para todas as faces de cada conjunto, segundo a equação abaixo:

$$\bar{\rho} = \frac{2}{N(N-1)} \sum_{i=2}^N \sum_{j=1}^{i-1} |\rho_{ij}|, \quad (8)$$

onde N é o número de imagens analisadas por conjunto, ou seja, para faces no ambiente controlado $N = 200$ e para faces no ambiente não-controlado $N = 125$. O valor da correlação ρ_{ij} foi obtido com base no coeficiente de Pearson (BRITO; COELHO; 2002), através da seguinte equação:

$$\rho_{ij} = \frac{\sum_{k=1}^n (i_k - \bar{i})(j_k - \bar{j})}{\sqrt{\sum_{k=1}^n (i_k - \bar{i})^2} \sqrt{\sum_{k=1}^n (j_k - \bar{j})^2}}, \quad (9)$$

onde i_k e j_k correspondem, respectivamente, ao pixel k das imagens i e j , e \bar{i} e \bar{j} representam os valores médios de todos os pixels das imagens i e j . Neste trabalho, $n = 31.266$ pois a resolução final das imagens após a normalização é igual a 162×193 . Devido à necessidade de padronização dimensional para o cálculo de correlação entre as faces de cada conjunto, a escala das imagens capturadas em ambientes não-controlados foi ajustada à escala padrão das faces normalizadas, com resolução também de 162×193 pixels.

A Tabela 1 apresenta os resultados da correlação dos conjuntos analisados. Pode-se observar pela Tabela 1, que os valores gerados possuem a mesma tendência dos resultados apresentados por Amaral e Thomaz (2008), onde a média da correlação entre imagens normalizadas é, pelo menos, 20% maior que a média da correlação entre as imagens originais. Entretanto, as adaptações propostas neste artigo proporcionaram um aumento da correlação entre as imagens analisadas por Amaral e Thomaz (2008), que pertencem ao conjunto de imagens oriundas de um ambiente de captura controlada exclusivamente. Este aumento da correlação média se deve principalmente a eliminação de artefatos através do novo procedimento de recorte proposto neste artigo e descrito em detalhes no item 2.3.

Tabela 1: Médias das correlações em pares das imagens analisadas.

Ambiente	Originais	Normalizadas
Controlado	0,562454506	0,712162609
Não-Controlado	0,262074469	0,351251183

A Figura 6 ilustra através das faces médias o resultado visual gerado pelo processo de normalização dos conjuntos de imagens capturadas em ambientes controlados e não-controlados.

Figura 6: Imagem média das faces frontais antes e após a normalização espacial: ambiente controlado (esquerda); ambiente não-controlado (direita).



6. Conclusão

Este artigo propôs um procedimento de normalização espacial de imagens frontais de face em ambientes cujos processos de captura de imagens são controlados e não-controlados. Verificou-se que tal procedimento é funcional tanto em termos visuais como quantitativos, em ambas as situações. Os resultados dos métodos aplicados são mais perceptíveis em imagens oriundas de ambientes não-controlados, onde a presença de artefatos e desvios é muito comum. Constatou-se também que as alterações sugeridas neste trabalho proporcionaram um resultado superior ao apresentado por Amaral e Thomaz (2008), que analisou exclusivamente imagens em ambientes controlados sem as adaptações de recorte implementadas neste trabalho.

Referências Bibliográficas:

ALVES, Walison José Barberá. **Identificação de pessoas através de algoritmos genéticos aplicado em medidas das proporções áureas da face humana.** 2008. 454 f. Dissertação (Mestrado em Engenharia Elétrica) – Escola de Engenharia de São Carlos da Universidade de São Paulo, São Carlos. 2008.

AMARAL, Vagner; THOMAZ, Carlos E. **Normalização espacial de imagens frontais de face**. Relatório Técnico 2008/01, São Bernardo do Campo: Centro Universitário da FEI, 2008. 13 p. (disponível em <http://www.fei.edu.br/~cet/publications.html>).

BRITO, Jorge; COELHO, Luiz. **Fotogrametria Digital**. 1 ed. Instituto Militar de Engenharia, Rio de Janeiro, 2002.

CONCI, Aura; AZEVEDO, Eduardo; LETA, Fabiana. R. **Computação Gráfica**. Rio de Janeiro, Editora Campus/Elsevier, 2008. v. 2. 432 p.

GONZALES, Rafael C.; WOODS, Richard E. **Processamento de imagens digitais**. São Paulo: Edgard Blücher, 2000. 509 p.

MORIHISA, Olivia. **Avaliação comparativa entre agradabilidade facial, proporção áurea e padrão facial**. 2006. 352 f. Dissertação (Mestrado em Odontologia) – Faculdade de Odontologia da Universidade Metodista de São Paulo, São Bernardo do Campo. 2006.

OLIVEIRA JUNIOR, Leo L.; THOMAZ, Carlos E. **Captura e Alinhamento de Imagens: Um Banco de Faces Brasileiro**. Relatório, São Bernardo do Campo: Centro Universitário da FEI, 2005. 10 p. (disponível em <http://www.fei.edu.br/~cet/publications.html>).

RUECKERT, D. et al. **Non-rigid registration using free-form deformations: Application to breast MR images**. IEEE Transactions on Medical Imaging, vol. 18, no. 8, pp. 712-721, 1999.