

# Normalização Espacial de Imagens Frontais de Face

Vagner do Amaral<sup>1</sup> e Carlos Eduardo Thomaz<sup>2</sup>

**Relatório Técnico: 2008/01**

<sup>1</sup> Coordenadoria Geral de Informática  
Centro Universitário da FEI  
São Bernardo do Campo, São Paulo, Brasil

<sup>2</sup> Departamento de Engenharia Elétrica  
Centro Universitário da FEI  
São Bernardo do Campo, São Paulo, Brasil

## Resumo

O objetivo deste relatório é propor uma sequência de processamento de imagens para normalização de fotos frontais de face. Para isso, desenvolveu-se um arcabouço computacional capaz de minimizar os desvios gerados no processo de captura. Os resultados iniciais apresentados neste trabalho demonstram o aumento da nitidez em pontos específicos da face média e o aumento da correlação matemática entre as faces normalizadas, justificando a aplicação dos métodos sugeridos.

**Palavras-chave:** Processamento de Imagens, Normalização, Banco de Faces.

## Abstract

In this work, we propose and implement a sequence of image processing steps to spatially normalize frontal face images. A computational tool has been developed to minimize variations due to the acquisition protocol. Our preliminary results are promising, showing an increase of sharpness in specific points of the average face image and higher cross-correlation between the normalized face images compared to the original ones.

**Keywords:** Image Processing, Normalization, Face Database.

## 1 Introdução

Nas áreas de Visão Computacional e Processamento de Imagens, atividades relacionadas ao reconhecimento e detecção de faces recebem grande destaque,

porém, em muitos casos, os estudos biométricos são prejudicados devido a desvios de inclinação e escala que podem ocorrer durante a captura da imagem.

Pesquisas baseadas na combinação de imagens de faces devem ter um grau preciso de alinhamento, em razão da necessidade de correlação entre informações discriminantes contidas em faces distintas.

Visando eliminar os desvios de inclinação e escala gerados pelo processo de alinhamento manual utilizado em [1], este trabalho apresenta detalhadamente um método de processamento para normalização de imagens frontais de face, baseado no modelo apresentado em [5].

## 2 Metodologia

Fundamentado na proposta de interação mínima do usuário, a execução do procedimento apresentado requer apenas a definição das dimensões de edição em unidades de pixels e a identificação manual da posição das pupilas como pontos de referência para os cálculos utilizados. As Figuras 1(a) e 1(b) ilustram este procedimento.

O método de normalização de imagens de face sugerido foi subdividido em etapas de processamento dedicado, que tratam desvios específicos e obedecem a seguinte seqüência predefinida: rotação, escala, recorte, descoloração e suavização de contorno. As subseções seguintes descrevem em detalhes cada uma das etapas desse processamento proposto.

### 2.1 Rotação

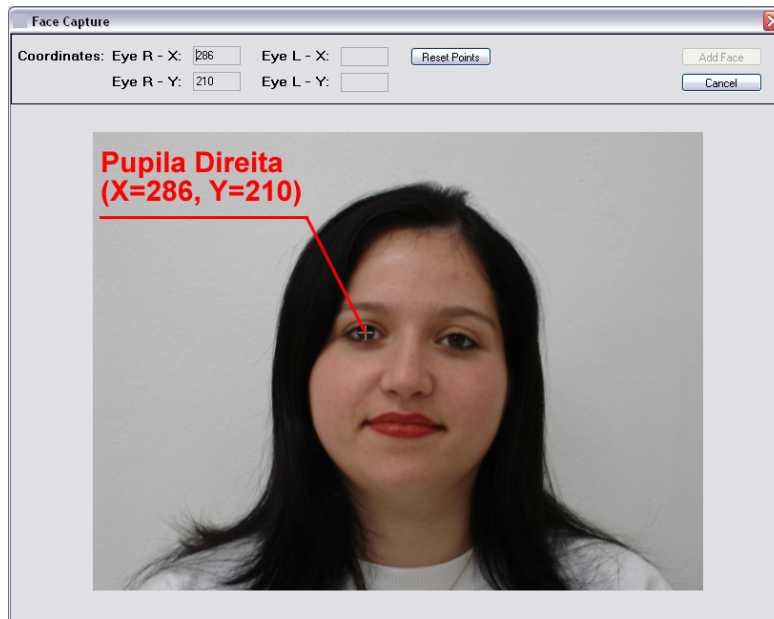
O objetivo desta etapa é alinhar a posição das pupilas no eixo horizontal, baseado nas medidas de seno e cosseno do ângulo  $\theta$ , formado pela inclinação da reta imaginária que passa através dos pontos de referência em relação ao eixo horizontal, conforme ilustrado na Figura 2.

Inicialmente, utilizou-se um método de identificação 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 [2], baseado na seguinte equação:

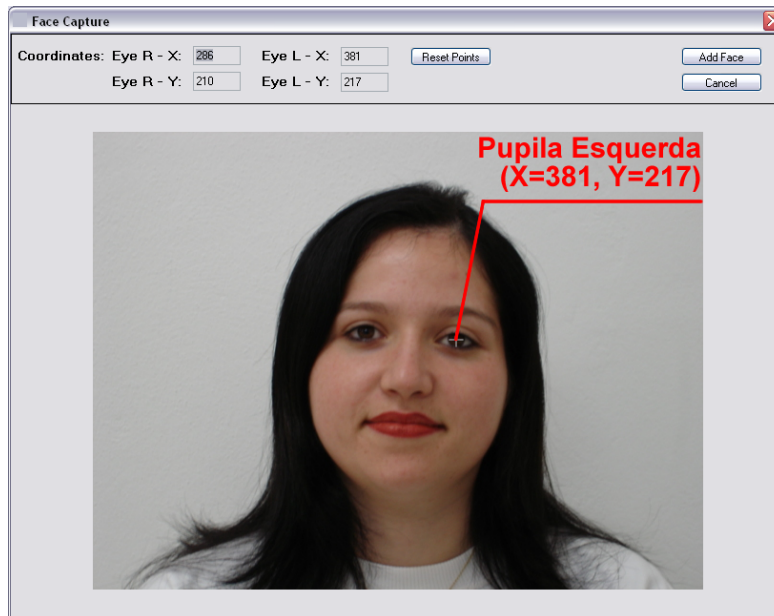
$$\begin{pmatrix} x_d \\ y_d \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \times \begin{pmatrix} x_o \\ y_o \end{pmatrix}. \quad (1)$$

Entretanto, essa abordagem gera na imagem destino um ruído conhecido como impulsivo (também denominado de "sal e pimenta") [4, 2], pois o endereço  $(x_d, y_d)$  dos pixels é formado por números inteiros e o arredondamento de um valor decimal obtido através da equação (1) pode gerar sobreposição de pixels na imagem destino e deixando assim alguns pontos sem preenchimento, conforme ilustrado na Figura 3(a).

Este tipo de ruído pode ser eliminado de forma eficaz através de um filtro de mediana [2], entretanto isso acarretaria em mais uma etapa de processamento, portanto utilizou-se uma abordagem inversa proposta neste trabalho, onde cada



(a)



(b)

Figura 1: (a) Identificação manual das coordenadas da pupila direita; (b) Identificação manual das coordenadas da pupila esquerda.

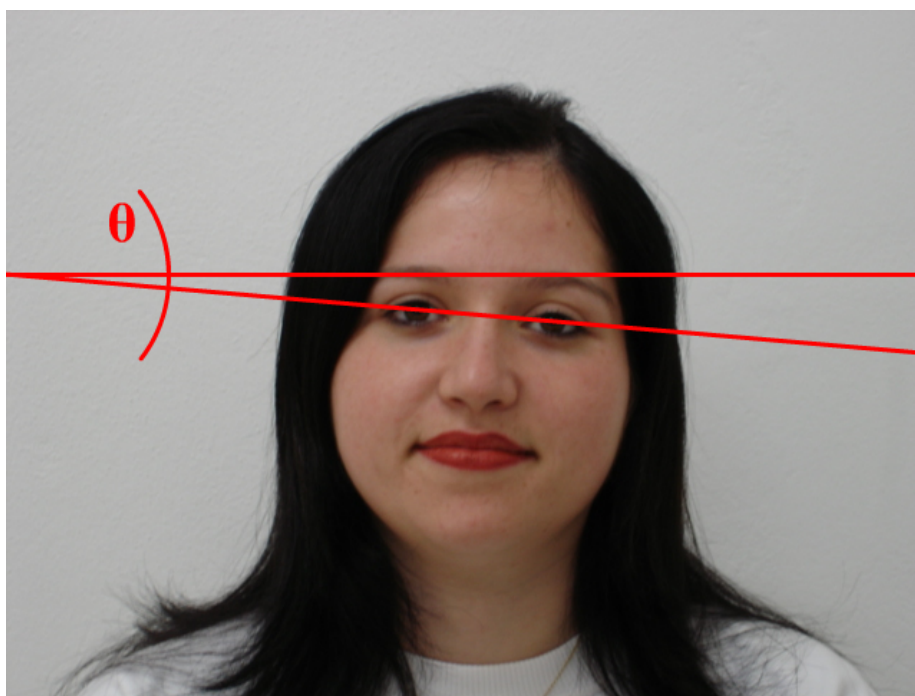


Figura 2: Reta imaginária que define o ângulo  $\theta$  de rotação da imagem.

ponto na imagem destino identifica seu ponto correspondente na imagem origem, evitando a sobreposição de pixels e os ruídos.

O método inverso foi desenvolvido a partir da decomposição da equação (1), utilizando uma técnica conhecida como rotação em duas passadas [2], onde o endereço de um dos eixos origem é identificado primeiramente para compor a equação que resultará no endereço do segundo eixo.

Na etapa inicial foi aplicada a regra distributiva nos elementos da matriz:

$$\begin{aligned}x_d &= \cos \theta \cdot x_o + \sin \theta \cdot y_o \\y_d &= -\sin \theta \cdot x_o + \cos \theta \cdot y_o\end{aligned}\quad (2)$$

Em seguida uma das variáveis correspondentes às coordenadas de origem foi isolada,

$$\begin{aligned}x_d &= \cos \theta \cdot x_o + \sin \theta \cdot y_o \\x_d - \sin \theta \cdot y_o &= \cos \theta \cdot x_o \\x_o &= \frac{x_d - \sin \theta \cdot y_o}{\cos \theta}\end{aligned}\quad (3)$$

e aplicada na equação da segunda variável:

$$y_d = -\sin \theta \left( \frac{x_d - \sin \theta \cdot y_o}{\cos \theta} \right) + \cos \theta \cdot y_o \quad (4a)$$

$$y_d = -\sin \theta \left( \frac{x_d}{\cos \theta} \right) - \sin \theta \left( \frac{-\sin \theta \cdot y_o}{\cos \theta} \right) + \cos \theta \cdot y_o \quad (4b)$$

$$\sin \theta \left( \frac{-\sin \theta \cdot y_o}{\cos \theta} \right) - \cos \theta \cdot y_o = -\sin \theta \left( \frac{x_d}{\cos \theta} \right) - y_d \quad (4c)$$

$$y_o \left( \sin \theta \left( \frac{-\sin \theta}{\cos \theta} \right) - \cos \theta \right) = -\sin \theta \left( \frac{x_d}{\cos \theta} \right) - y_d \quad (4d)$$

$$y_o = \frac{-\sin \theta \left( \frac{x_d}{\cos \theta} \right) - y_d}{\sin \theta \left( \frac{-\sin \theta}{\cos \theta} \right) - \cos \theta} \quad (4e)$$

Para finalizar a técnica de rotação em duas passadas [2], o resultado obtido na equação (4) de  $y_o$  é aplicado na equação (3) 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. Dessa forma a rotação é aplicada sem a perda de informações da imagem original, gerando o resultado apresentado na Figura 3(b).

Entretanto, como pode-se observar na Figura 3(b), a rotação dos pontos da matriz gera um efeito de serrilhado na imagem, que será tratado na última etapa de processamento, ou seja, na etapa de suavização de contorno descrita na subseção 2.5.



(a)



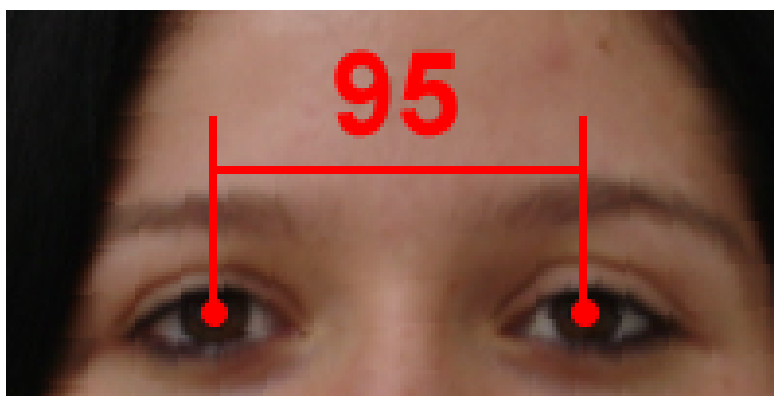
(b)

Figura 3: (a) Imagem rotacionada apresentando ruído impulsivo; (b) Imagem rotacionada pela abordagem proposta, sem a presença de ruído impulsivo.

## 2.2 Escala

A alteração de escala nas imagens de face é derivada da padronização na distância entre os pontos de identificação das pupilas, que visa corrigir desvios causados durante a captura e aumentar o grau de correlação entre imagens de face de pessoas distintas.

Esta etapa identifica a coordenada de um pixel na imagem original através da multiplicação de cada pixel da imagem destino por um escalar definido pela razão entre a distância interpupilar original e uma distância padrão predefinida [5], conforme exemplo apresentado na Figura 4. As dimensões da imagem final são especificadas aplicando a multiplicação pela razão inversa, ou seja, a distância padrão pela distância interpupilar. Para os cálculos realizados neste trabalho foi utilizada uma distância interpupilar padrão de 100 pixels [5].



(a)



(b)

Figura 4: (a) Dimensão na imagem para a distância interpupilar original; (b) Dimensão predefinida para a distância interpupilar final.

### 2.3 Recorte

Este processo consiste em isolar a região da face, com o objetivo de otimizar a área de processamento evitando interferências do fundo. A região extraída é calculada através de medidas predefinidas [5], que utilizam como referência os pontos de identificação das pupilas para estabelecer os limites das margens laterais, superior e inferior da imagem final. A Figura 5 apresenta um exemplo dessa etapa de processamento das imagens de face.

### 2.4 Descoloração

A fim de reduzir o esforço computacional, nesta etapa cada pixel da imagem original é convertido para escala de cinza, eliminando os canais RGB através da seguinte equação:

$$(x_d, y_d) = 0.299(x_o, y_o) + 0.587(x_o, y_o) + 0.114(x_o, y_o). \quad (5)$$

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 [4]. Portanto, esta expressão fornece um resultado mais adequado que a divisão proporcional, com 33% para cada cor.

### 2.5 Suavização de Contorno

Devido ao efeito de serrilhado (também denominado *aliasing*) [2] gerado pela rotação da imagem, foi implementado ao final do processo de normalização uma técnica de *anti-aliasing* para amenizar o serrilhado presente nos contornos.

Alguns métodos foram testados e o melhor resultado foi obtido através de um filtro de mediana, utilizando um kernel 3x3 para substituir o tom do pixel central pela mediana dos pixels ao seu redor. Por exemplo, dada a matrix 3x3 abaixo

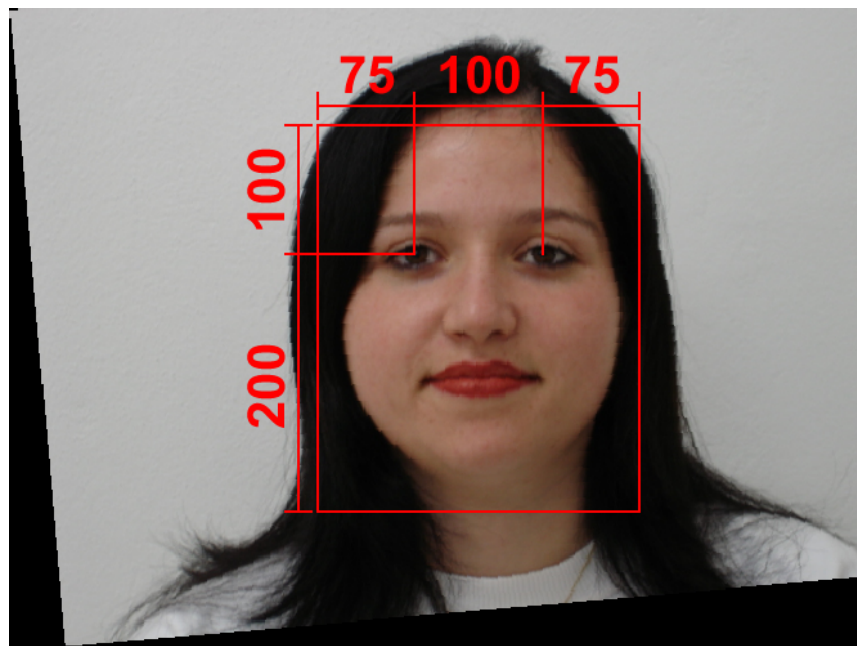
$$\begin{pmatrix} 101 & 110 & 196 \\ 95 & 111 & 115 \\ 89 & 99 & 107 \end{pmatrix}, \quad (6)$$

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

$$[89; 95; 99; 101; 107; 110; 111; 115; 196]. \quad (7)$$

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 do kernel. Através desta abordagem foi obtido um bom nível de suavização sem o ofuscamento de informações relevantes. A Figura 6 ilustra um exemplo da aplicação deste procedimento.





(a)



(b)

Figura 5: (a) Dimensões predefinidas para o recorte da imagem; (b) Imagem recortada.



(a)



(b)

Figura 6: (a) Imagem apresentada com efeito de serrilhado nos contornos; (b) Imagem após o filtro de mediana utilizado para suavizar o efeito de serrilhado.

### 3 Implementação

As imagens empregadas neste trabalho estão organizadas em um banco de faces criado e mantido pelo Departamento de Engenharia Elétrica do Centro Universitário da FEI, composto de 14 imagens de 200 indivíduos em diversos ângulos, totalizando 2800 imagens capturadas de acordo com o procedimento descrito em [1]. Porém, somente as imagens frontais neutras e sorrindo foram utilizadas neste trabalho, totalizando 400 (2x200) imagens.

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 2.0. 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.

### 4 Resultados Obtidos

Os resultados dos procedimentos descritos neste trabalho foram quantificados através da comparação da correlação média dos 2 conjuntos de faces analisados (imagens neutras e sorrindo), antes e após a normalização. Visando eliminar a interferência do fundo na análise, as faces originais foram cortadas nas dimensões predefinidas para normalização.

Utilizou-se para comparação o módulo da correlação em pares para todas as faces de cada conjunto, segundo o modelo matemático 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,  $N = 200$ . O valor da correlação  $\rho_{ij}$  foi obtido com base no coeficiente de Pearson [3], 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,  $n = 75000$  pois a quantidade de pixels (resolução) final das imagens após a normalização é igual a 250x300. A Tabela 1 mostra os resultados da correlação desses conjuntos analisados.

Conjunto de Faces	Originais	Normalizadas
Neutras	0,33268690	0,40004576
Sorrindo	0,32047679	0,38592069

Tabela 1: Médias das correlações em pares dos conjuntos analisados.

Pode-se observar pela Tabela 1 que a correlação entre as imagens de cada conjunto normalizado (neutras ou sorrindo) aumentou em aproximadamente 20% com relação às imagens de cada conjunto original correspondente. Estes resultados podem ser visualizados através das faces médias geradas após a normalização que apresentam maior nitidez em regiões específicas, principalmente ao redor dos olhos que são os pontos de referência para os cálculos aplicados no projeto. As Figuras 7 e 8 ilustram a comparação visual das imagens médias geradas a partir das faces originais antes e após a normalização.



Figura 7: (a) Imagem média das faces frontais neutras originais; (b) Imagem média das faces frontais neutras normalizadas.

## 5 Conclusão e Trabalhos Futuros

Este trabalho descreveu uma abordagem para o desenvolvimento de bancos de faces padronizados através de etapas de processamento de imagens orientadas pela posição das pupilas, demonstrando a eficiência do processo de normalização proposto para o alinhamento de regiões comuns entre faces frontais distintas. Como trabalhos futuros, pretende-se investigar e implementar um método para a identificação automática das pupilas com o intuito de reduzir a interação com o usuário e o risco de falha humana, otimizando o processo de normalização proposto.

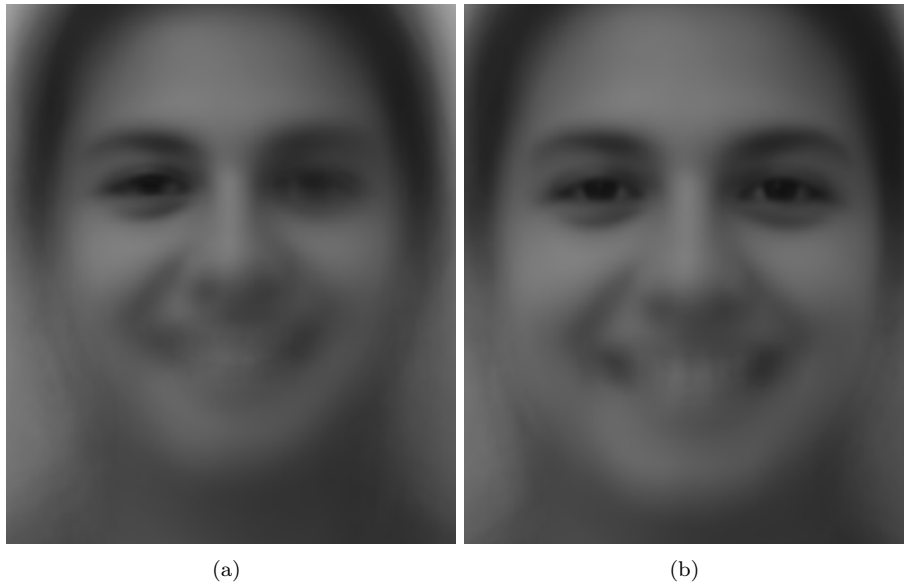


Figura 8: (a) Imagem média das faces frontais sorrindo originais; (b) Imagem média das faces frontais sorrindo normalizadas.

## Agradecimentos

Os autores desse trabalho gostariam de agradecer o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) através do projeto de pesquisa 472386/2007-7.

## Referências

- [1] L. L. Oliveira e C. E. Thomaz. Captura e alinhamento de imagens: Um banco de faces brasileiro. Relatório técnico, Centro Universitário da FEI, Departamento de Engenharia Elétrica, Junho 2006. Bolsa FEI/PBIC.
- [2] A. Conci e E. Azevedo e F. R. Leta. *Computação Gráfica*, volume 2. Editora Campus/Elsevier, Rio de Janeiro, 2008.
- [3] J. Brito e L. Coelho. *Fotogrametria Digital*. Instituto Militar de Engenharia, Rio de Janeiro, Brasil, 2002.
- [4] R. C. Gonzales e R. E. Woods. *Processamento de Imagens Digitais*. Editora Edgard Blücher, 1992.
- [5] G. N. Schroeder. Morphing aplicado ao envelhecimento de imagens faciais. Master's thesis, Universidade Estadual de Campinas (UNICAMP), Departamento de Engenharia Elétrica e de Computação, Abril 2007.