

## 4 Metodologia

Neste capítulo, se detalham as características dos bancos de imagens estáticos e de vídeo que foram usadas no desenvolvimento desta dissertação; posteriormente, são apresentadas as ferramentas digitais que permitiram o desenvolvimento desta pesquisa, com especial interesse no procedimento adotado na criação do detector de faces frontais, o que foi inserido no protótipo de Vídeo Monitoramento do LVC, que permitiu a captura de imagens faciais normalizadas a partir de sequências de vídeo; e na biblioteca facereclib com a que se compararam e avaliaram os diferentes algoritmos de reconhecimento facial apresentados no Capítulo 3.

Finalmente, se expõe o processo experimental empregado na execução dos testes.

### 4.1. Bancos de Imagens Estáticas e Vídeo

No desenvolvimento deste estudo foram usados bancos de imagens estáticas e imagens extraídas de vídeos, ambos de acesso público, mas que em alguns casos requerem a permissão explícita dos autores ou das entidades responsáveis.

- **ORL Face Database** (Samaria, et al., 1994), contém imagens faciais de 40 participantes (10 imagens por participante), em alguns casos as fotos foram tiradas variando a iluminação, expressões, acessórios faciais (óculos) e pequenas rotações de perfil, ver Figura 5.
- **FEI Face Database** (De Oliveira, et al., 2006), é um banco de imagens faciais brasileiro composto de 14 imagens de cada um dos 200 indivíduos (100 homens e 100 mulheres) que compõem o banco; todas as imagens são coloridas contra um fundo branco uniforme. Além disso, o banco apresenta imagens com variação de iluminação, expressões faciais (sorriso) e rotação de perfil distribuídas da seguinte forma:

- 1 imagem de face frontal e expressão neutra.
- 1 imagem de face frontal e expressão de sorriso.
- 8 imagens com rotação de perfil (22,5° para cada posição).
- 2 imagens de  $\pm 10^\circ$  de rotação no perfil com respeito à frontal.
- 1 imagem de face frontal com iluminação não uniforme.
- 1 imagem com variações em iluminação, expressão e rotação de perfil.

Na Figura 6, se apresentam as imagens pertencentes a um indivíduo do banco de dados.

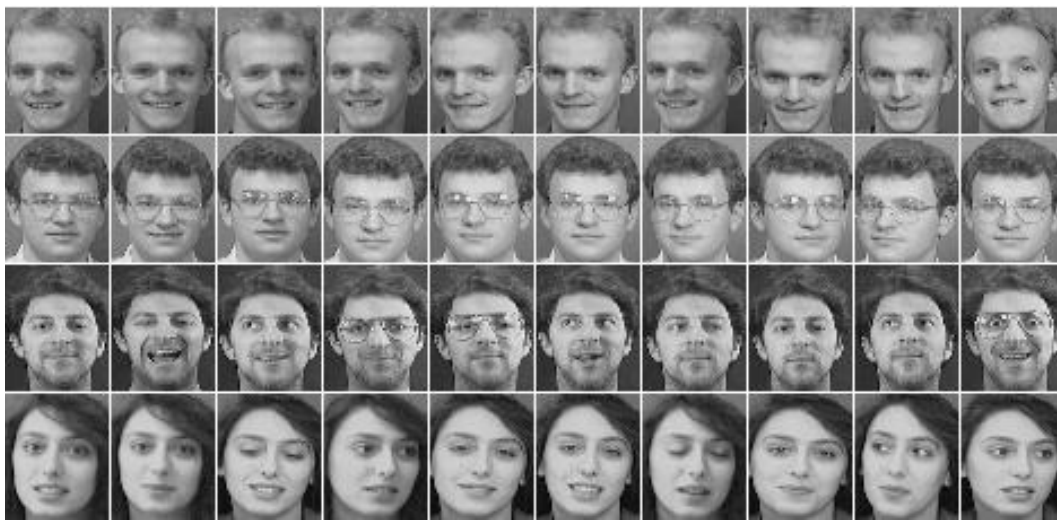


Figura 5. Imagens de amostra do banco ORL.



Figura 6. Imagens de amostra do banco FEI.

- **The IMM Face Database** (Nordstrom, et al., 2004), é um banco de imagens faciais estáticas composto de 240 imagens de 40 indivíduos (43 homens e 7 mulheres). No banco existem imagens com variação de expressões faciais e rotações de perfil. A seguir se detalham as características das imagens de cada pessoa.

- 1 imagem de face frontal com expressão neutra.
- 1 imagem de face frontal com expressão alegre.
- 2 imagens de  $\pm 30^\circ$  de rotação no perfil com respeito à frontal.
- 1 imagem de face frontal com expressão neutra e iluminação não uniforme.
- 1 imagem com variações em expressão e rotação de perfil.

A Figura 7, apresenta as imagens do individuo número 02 com suas 6 imagens faciais.



Figura 7. Imagens de amostra do banco IMM.

- **The VIDTIMIT Audio-Video Dataset** (Sanderson, 2008), é um banco de vídeos gerado pela gravação de cenas com 43 pessoas em 3 sessões cada, apresentando movimentos de cabeça da esquerda para direita e de acima para baixo, ambos com retorno a posição frontal. Os vídeos têm uma taxa de amostragem de 28 quadros/segundo e estão armazenados em sequências de imagens com resolução de 512x384 pixels. Algumas imagens exemplo correspondentes a primeira sessão são apresentadas na Figura 8.



Figura 8. Imagens de amostra do banco VidTIMIT.

- **ChokePOINT Dataset** (Wong, et al., 2011), é uma coleção de vídeos auspiciada pelo NICTA<sup>5</sup>. O banco tem gravações realizadas em 10 sessões distintas com 6 câmeras distribuídas em 2 portões (3 câmeras por portão) e localizadas na parte superior dos portões.

O conjunto de vídeos consiste de 25 participantes (19 homens e 6 mulheres) no portão 1 e 29 participantes (23 homens e 6 mulheres) no portal 2, sendo que para 8 das 10 sessões, os participantes aparecem no quadro de vídeo um a cada vez, ver Figura 9 (a), e nas 2 sessões restantes, um cenário de multidão foi simulado, ver Figura 9 (b).

Os vídeos têm uma taxa de amostragem de 30 quadros/segundo, com uma resolução de 800x600 pixels e estão armazenados em diferentes sequências de imagens por câmera e sessão.

<sup>5</sup> Australia's Information Communications Technology (ICT) Research Centre of Excellence.

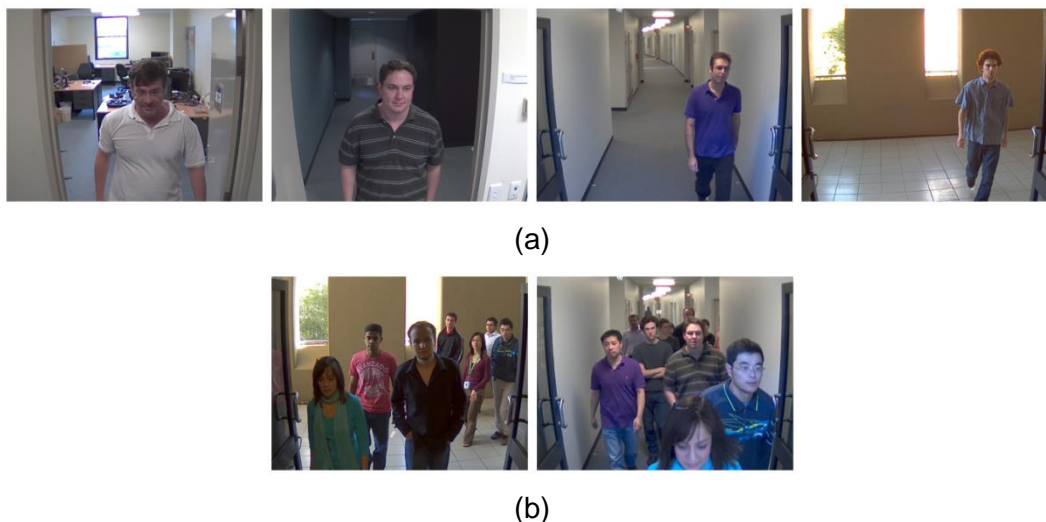


Figura 9. Imagens de amostra do banco ChokePOINT.

- **The Honda/UCSD Video Database** (Lee, et al., 2003) e (Lee, et al., 2005), é um banco de vídeos que contém gravações de 35 pessoas. No banco existem, em alguns casos, até 3 sequências de vídeo por pessoa com uma resolução de 640x480 pixels a uma taxa de amostragem de 15 quadros/segundo com duração de entre 15 a 20 segundos, algumas imagens são apresentadas na Figura 10.



Figura 10. Imagens de amostra do banco HONDA.

- **MOBIO Database** (McCool, et al., 2012), consiste na gravação dum conjunto de 192 vídeos e áudios de cada uma das 152 pessoas (100 homens e 52 mulheres) que compõem o banco; o acesso é público, desde que se tenha a permissão explícita do *Idiap Dataset Distribution Portal*.

As gravações foram adquiridas em 6 localidades distintas de 5 países diferentes com um telefone celular Nokia N93i em posições



variadas, o que realça as variações de iluminação, posição, entre outras, presentes no ambiente, ver Figura 11.



Figura 11. Imagens de mostra do banco MOBIO.

#### 4.2. Detector de Face Frontal

O detector de face frontal é um classificador binário encarregado de etiquetar imagens faciais como “frontal” ou “não frontal” com respeito à orientação da cabeça relativamente à câmera. Admite como variáveis de entrada os atributos extraídos a partir dos ângulos internos do triângulo formado pelos centros dos olhos e a boca.

O detector foi criado com a finalidade de melhorar o desempenho de um protótipo de Vídeo Monitoramento desenvolvido anteriormente no LVC e está fundamentado nos trabalhos realizados por (Abdel-Mottaleb, et al., 2007) e (Pinho, 2012), que avaliam as imagens faciais com respeito às variações em iluminação, nitidez, *orientação da face*<sup>6</sup> (ver Figura 16), entre outras, antes de serem submetidas ao reconhecimento em sim.

Segundo (Pinho, 2012), uma imagem facial com vista frontal tem uma maior probabilidade de ser identificada, seja qual for o algoritmo de reconhecimento empregado. Assim, para a criação do detector se usaram os bancos de imagens FEI, IMM e as sequências de imagens de uma das sessões do VidTIMIT. Além disso, se adotou o algoritmo LBP (*Local Binary Pattern*) como técnica de reconhecimento de referência.

<sup>6</sup> Arfagem, rolagem e guinada, são as variações em rotação de um objeto respeito dos eixos vertical, normal e horizontal do plano da imagem respetivamente.

A criação do detector envolveu três etapas, conforme ilustrado na Figura 12:

- a detecção,
- extração de atributos e
- determinação do estado frontal.

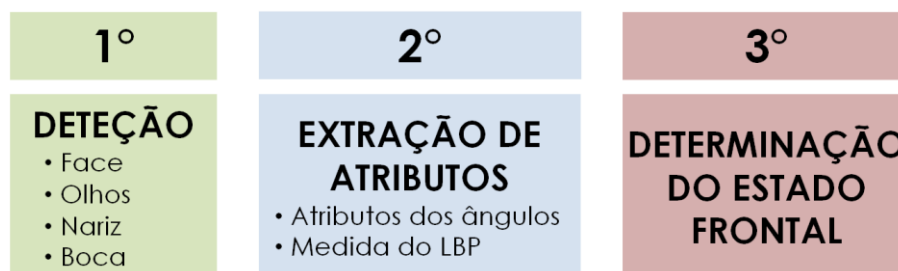


Figura 12. Etapas para a criação do detector de face frontal.

Na detecção, se localizam os pontos que representam aos olhos e a boca. A extração de atributos consiste em medir os ângulos internos do triângulo formado pelos centros dos olhos e da boca e ainda a medida de similaridade da imagem com uma imagem frontal da mesma pessoa, conforme medida pelo algoritmo de reconhecimento LBP. Finalmente, a determinação do estado frontal, cuja análise foi realizada com base nas curvas ROC (*Receiver Operating Characteristic*), como se descreve mais adiante.

#### 4.2.1. Detecção

Nesta primeira etapa, se localizaram os pontos que representam aos olhos e a boca em cada uma das imagens que compõem os três bancos; para isto, se usaram os detectores de objetos do modulo de classificação em cascata do OpenCV<sup>7</sup> (Haar Feature-based Cascade Classifier for Object Detection<sup>9</sup>). Estes detectores foram treinados com o algoritmo de Viola & Jones (Viola, et al., 2001) com o fim de encontrar as regiões na face que contém os olhos, nariz e boca.

Na Figura 13(a), se observa a face localizada para uma imagem exemplo. Em 11(b), os quadros em vermelho definem as regiões de busca para a

<sup>7</sup> Open Source Computer Vision, disponível em: <http://opencv.org>, último acesso em 19 de Fevereiro 2014.

<sup>9</sup> Disponível em: [http://docs.opencv.org/modules/objdetect/doc/cascade\\_classification.html](http://docs.opencv.org/modules/objdetect/doc/cascade_classification.html), último acesso em 19 de Fevereiro 2014.

localização das componentes faciais na imagem facial extraída. Na figura 11(c), as componentes faciais localizadas são exibidas em cor amarela e seus pontos centrais em vermelho.

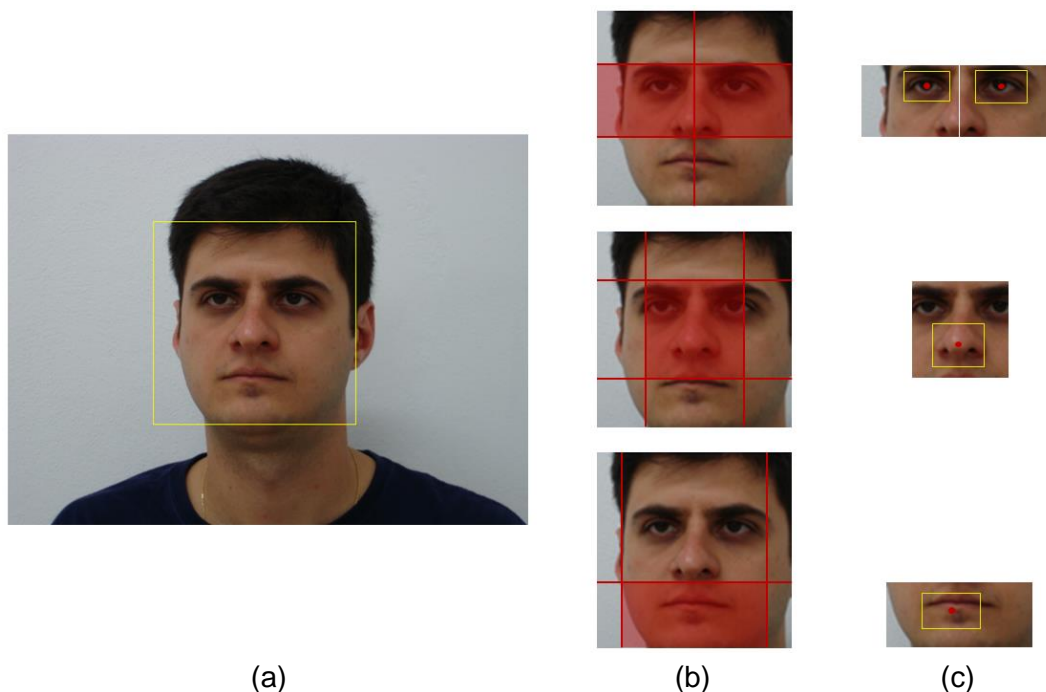


Figura 13. Localização dos pontos centrais dos olhos, nariz e boca.

Verificou-se experimentalmente que, em imagens de faces em posição frontal, os pontos fornecidos pelos detectores são localizados com boa exatidão. Assim, os pontos que correspondem a nariz e a boca estão verticalmente alinhados, e a linha que os conecta é ortogonal à linha que une aos olhos (ver Figura 14(a)). No entanto, isto nem sempre acontece para as imagens que apresentam pequenas variações em arfagem, onde em geral, o ponto fornecido pelo detector da boca se encontra deslocado relativamente ao ponto do nariz. A solução adotada para este problema consistiu em projetar o ponto da boca detectado sobre a linha reta que passa pelo nariz e é ortogonal à linha que une os olhos, como é apresentado na Figura 14(b).

Os pontos em verde da Figura 14 representam aos pontos bem localizados pelos detectores; o ponto preto em (b) corresponde ao ponto da boca com localização errada, e o ponto azul é o corrigido.



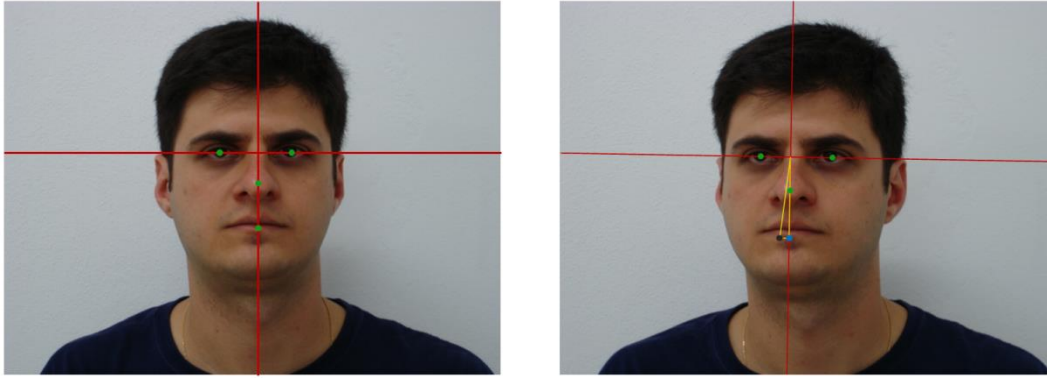


Figura 14. (a) Imagem com face frontal, (b) Imagem com variação na arfagem.

#### 4.2.2. Extração de Atributos

A união dos pontos referentes aos olhos e à boca, localizados na etapa anterior, definem um triângulo isósceles quando se trata de imagens frontais devido à simetria facial. Mas, quando a face apresenta variações em arfagem ou guinada o triângulo nem sempre é isósceles (ver Figura 15). Portanto, quanto mais o triângulo formado se afasta da forma isósceles, mas afastada a face estará da posição frontal.

Seguindo esta linha de raciocínio, deverá ser possível determinar faces frontais a partir dos ângulos internos do triângulo (Equações (4-1) – (4-4)), e o escore de reconhecimento, a saber:

$$A_{01} = \max(\alpha, \beta) \quad (4-1)$$

$$A_{02} = \min(\alpha, \beta) \quad (4-2)$$

$$A_{03} = |\alpha - \beta| \quad (4-3)$$

$$A_{04} = \gamma \quad (4-4)$$

onde,

- $\alpha, \beta$  e  $\gamma$ , são os ângulos correspondentes aos olhos e à boca.
- $A_{01}$  é o máximo entre os ângulos dos olhos.
- $A_{02}$  é o mínimo entre os ângulos dos olhos.
- $A_{03}$  é a diferença entre os ângulos dos olhos.
- $A_{04}$  é o ângulo da boca.

Por outro lado, o algoritmo de reconhecimento facial baseado no LBP, gera valores de similaridade  $S$  próximos a zero, quando compara duas imagens

frontais de um mesmo indivíduo. A similaridade tende a crescer quando pelo menos uma das imagens se afasta da posição frontal de referência (ver Figura 16). Então, do ponto de vista do classificador, é possível estabelecer relações entre os atributos  $A_{01}$ ,  $A_{02}$ ,  $A_{03}$ ,  $A_{04}$  e o escore  $S$  para determinar quando uma imagem facial se encontra em posição frontal, sempre que os pontos dos olhos e boca sejam conhecidos.

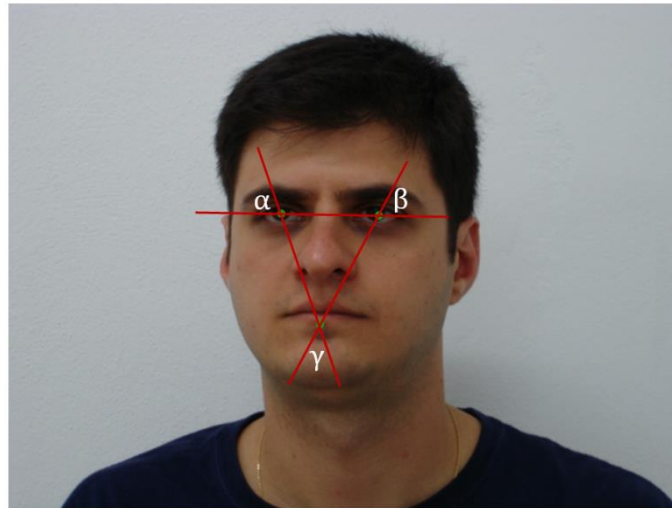


Figura 15. Triângulo e os ângulos gerados pela união dos olhos e boca

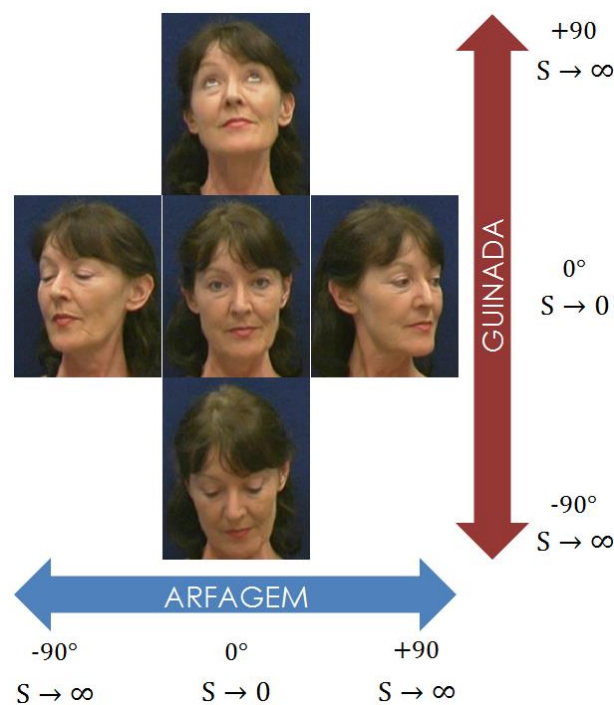


Figura 16. Variação da medida de similaridade  $S$  com respeito das variações em arfagem e guinada.

### 4.2.3. Determinação do Estado Frontal

Tal como mencionado na seção anterior, o algoritmo de reconhecimento facial LBP retorna valores de similaridade baixos, quando as imagens de referência e prova são frontais e pertencem à mesma pessoa. Estes valores aumentam a medida que as imagens de prova apresentam variações em arfagem e guinada; ou simplesmente quando a imagem da prova pertence a outra pessoa.

O método aqui proposto baseia-se nesta ideia e consiste em determinar limites para os valores dos atributos dentro dos quais uma face pode ser considerada frontal. A medida de “frontalidade” de uma face de prova é dada pela similaridade  $S$  desta imagem com uma imagem frontal do mesmo indivíduo escolhida a priori.

Assim, para selecionar a faixa de valores dos atributos mais relevantes que caracterizam uma face como frontal, foi concebido um classificador que recebe como entrada aos atributos e fornece uma saída binária (“frontal” ou “não frontal”). O ajuste final destes limites baseia-se na análise das curvas ROC (*Receiver Operating Characteristic*), as que foram construídas em base às taxas dos verdadeiros positivos e falsos positivos avaliados pra cada atributo.

A Figura 17 apresenta a abordagem proposta para determinar quando uma imagem de prova é considerada como verdadeira positiva ou falsa positiva. Nesta figura, os pontos em azul em (a) e aqueles em vermelho em (b) representam às imagens de prova cujos valores de similaridade  $S$  estão associados aos valores de um atributo em particular. Assim, os verdadeiros positivos ocorrem quando os valores de  $S$ , para um mesmo indivíduo, estão abaixo de um determinado limiar na medida de similaridade e à esquerda de um valor fixo no atributo (Figura 17(a)). Os falsos positivos, em contraste aos verdadeiros positivos, são avaliados para diferentes indivíduos, no entanto, os limiares na medida de similaridade e no atributo devem ser mantidos.

Nas Figura 18 à Figura 20, se apresentam os gráficos de similaridade relativamente a cada atributo para os bancos de imagens IMM, FEI e VidTIMIT. Na parte superior das figuras, são representados os valores de similaridade versus o valor máximo e mínimo dos ângulos dos olhos respectivamente (atributos  $A_{01}$  e  $A_{02}$ ). Similarmente, na parte inferior das figuras são representados os valores de similaridade versus a diferença dos ângulos dos olhos e o ângulo da boca respectivamente (atributos  $A_{03}$  e  $A_{04}$ ).

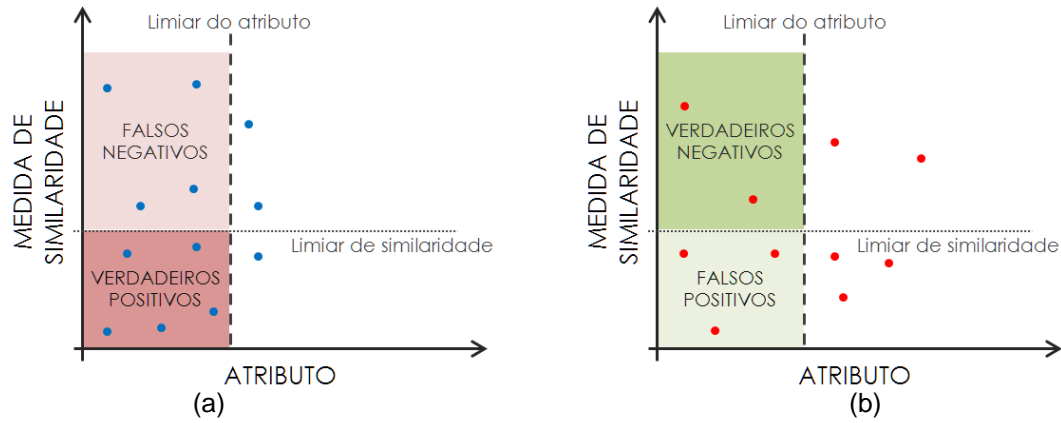


Figura 17. (a) Medidas de similaridade para o mesmo indivíduo com variações de arfagem e guinada, (b) Medidas de similaridade para diferentes indivíduos com posição frontal.

No caso dos bancos de imagens IMM e FEI (Figura 18 e Figura 19) se avaliaram imagens com variações em arfagem e no VidTIMIT (Figura 20) se avaliaram arfagem e guinada. A partir dos valores apresentados nestes gráficos é possível observar que, para que uma imagem facial possa ser considerada frontal, o valor do ângulo máximo dos olhos (atributo  $A_{01}$ ) está entre  $60^\circ$  e  $80^\circ$ , enquanto que o ângulo mínimo dos ângulos dos olhos (atributo  $A_{03}$ ) está compreendido entre  $60^\circ$  e  $70^\circ$ .

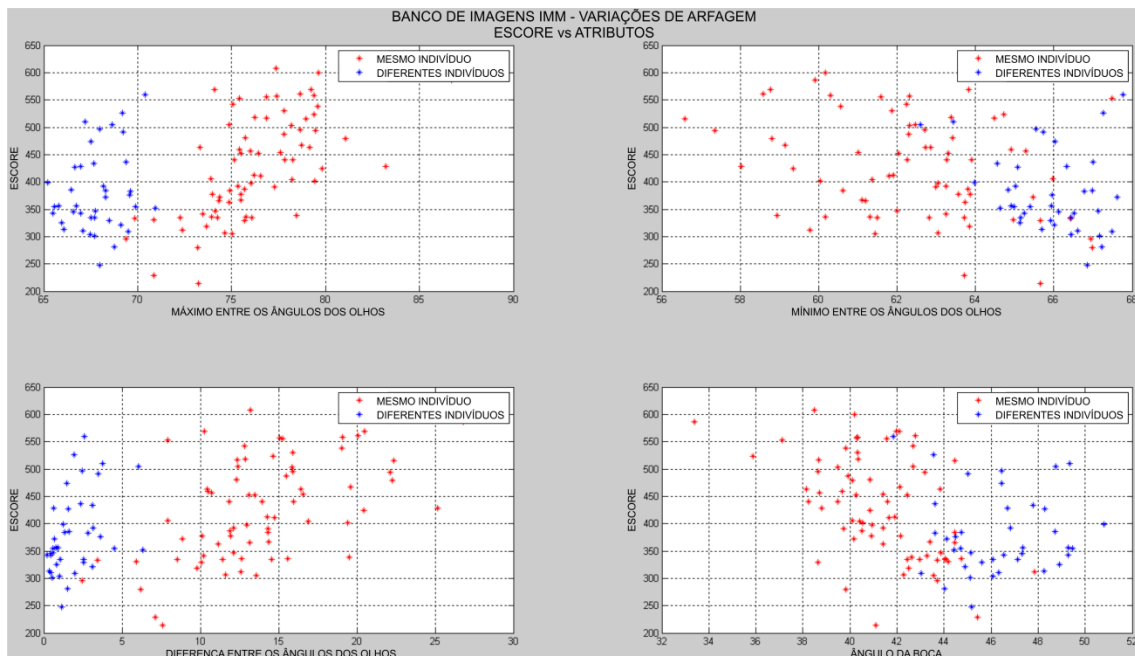


Figura 18. Medidas de similaridade versus atributos para o banco de imagens IMM. Em vermelho, mesmos indivíduos; em azul, indivíduos diferentes.

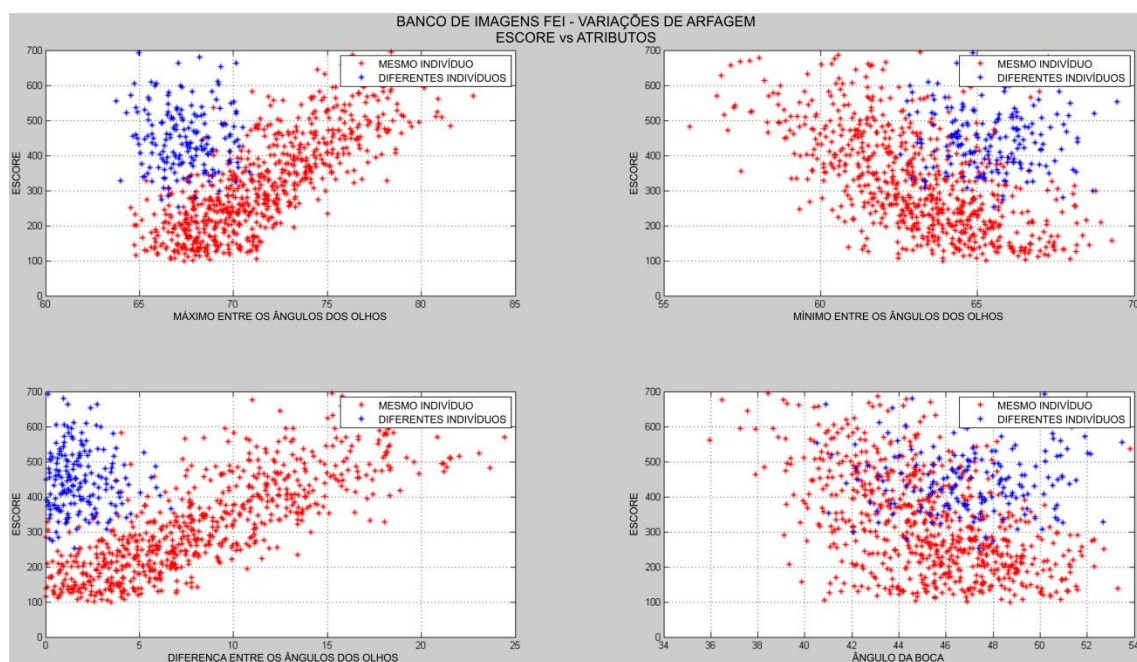


Figura 19. Medidas de similaridade versus os atributos para o banco de imagens FEI. Em vermelho, mesmos indivíduos; em azul, indivíduos diferentes.

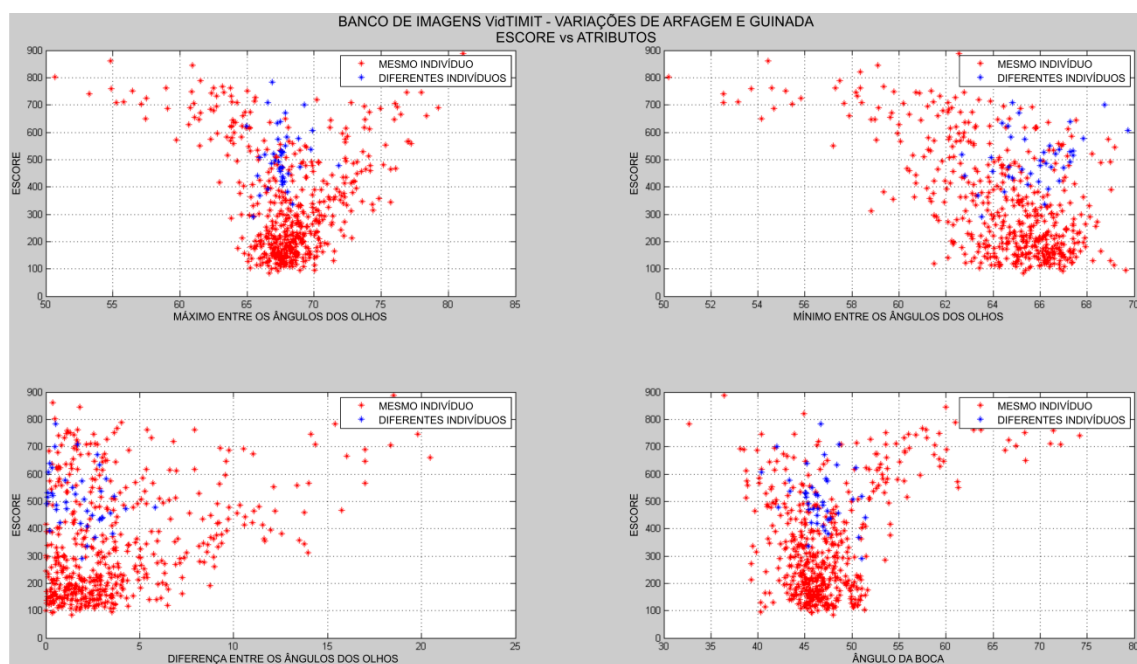


Figura 20. Medidas de similaridade versus os atributos para o banco de imagens VidTIMIT. Em vermelho, mesmos indivíduos; em azul, indivíduos diferentes.



A diferença entre os ângulos dos olhos (atributo  $A_{03}$ ) deve estar na faixa entre  $0^\circ$  a  $15^\circ$ , e o ângulo da boca (atributo  $A_{04}$ ) entre  $40^\circ$  e  $50^\circ$ .

Por outro lado, os atributos ângulo máximo dos olhos (atributo  $A_{01}$ ) e à diferença entre os ângulos dos olhos (atributo  $A_{03}$ ) se perfilam como os atributos que melhor separam as imagens faciais frontais.

As curvas ROC para os gráficos anteriores foram construídas levando em consideração a razão dos verdadeiros positivos frente à razão dos falsos negativos, quando se varia o limiar de similaridade para valores nos atributos fixados a priori. A Tabela 1, apresenta estas configurações nos parâmetros dos atributos que permitiram criar as curvas ROC. Por exemplo, o atributo  $A_{01}$ , referente ao ângulo máximo dos olhos, adota valores desde  $60^\circ$  até  $80^\circ$  sexagesimais com intervalos (passos) de  $1^\circ$  em  $1^\circ$ .

ATRIBUTOS	VALORES MÍNIMOS	VALORES MÁXIMOS	PASSO
A01 – Ângulo máximo dos olhos	$60^\circ$	$80^\circ$	$1^\circ$
A02 – Ângulo mínimo dos olhos	$60^\circ$	$70^\circ$	$1^\circ$
A03 – Diferença entre os ângulos dos olhos	$1^\circ$	$20^\circ$	$2^\circ$
A04 – Ângulo da boca	$40^\circ$	$52^\circ$	$1^\circ$

Tabela 1. Parâmetros dos atributos usados para a construção das curvas ROC.

Nas Figura 21 à Figura 23 se apresentam os gráficos das curvas ROC com respeito a cada atributo para os bancos de imagens IMM, FEI e VidTIMIT. Na superior das figuras encontram-se as curvas ROC para os ângulos máximo e mínimo dos olhos. Da mesma maneira, na parte inferior das figuras, se apresentam as curvas ROC para a diferença dos ângulos dos olhos e o ângulo da boca.

É importante ressaltar que, conforme as curvas ROC vão se aproximando ao canto superior esquerdo dos gráficos, melhor é a capacidade discriminativa dos classificadores.

Assim, os resultados apresentados nestes gráficos indicam que os atributos ângulo máximo dos olhos (atributo  $A_{01}$ ) e a diferença entre os ângulos dos olhos (atributo  $A_{03}$ ) têm o melhor poder de discriminar as imagens frontais. No caso do banco de imagens IMM, estes atributos apresentam uma discriminação fraca. Isto é devido à pouca quantidade de imagens do banco, e à presença de imagens com variações consideráveis em arfagem.

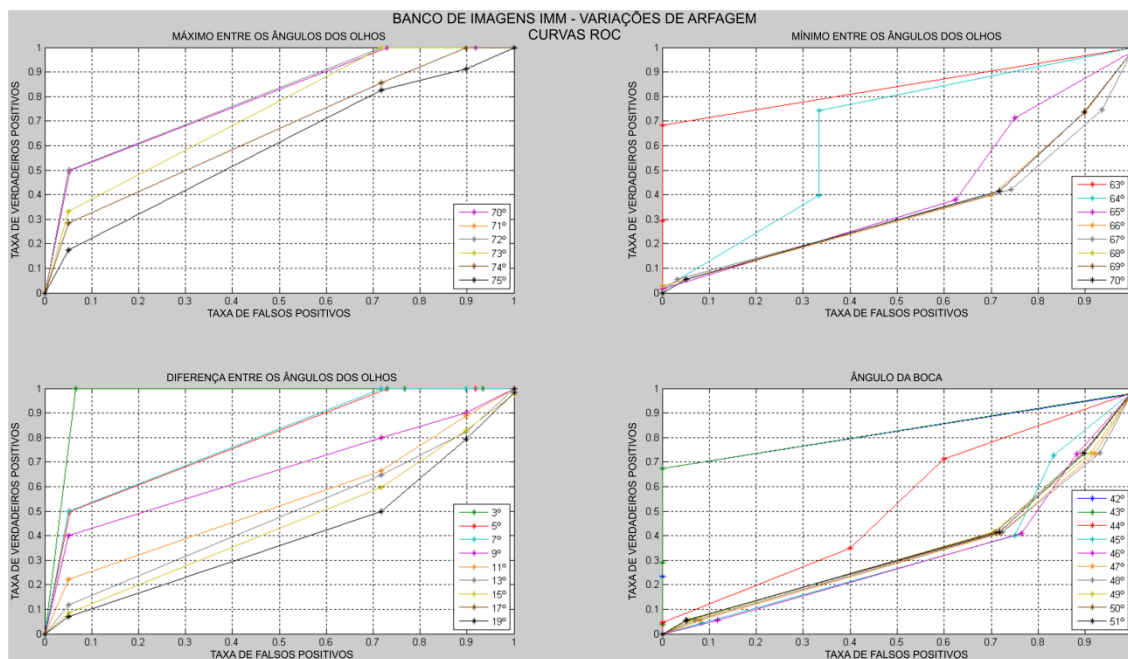


Figura 21. Curvas ROC para os atributos no banco de imagens IMM.

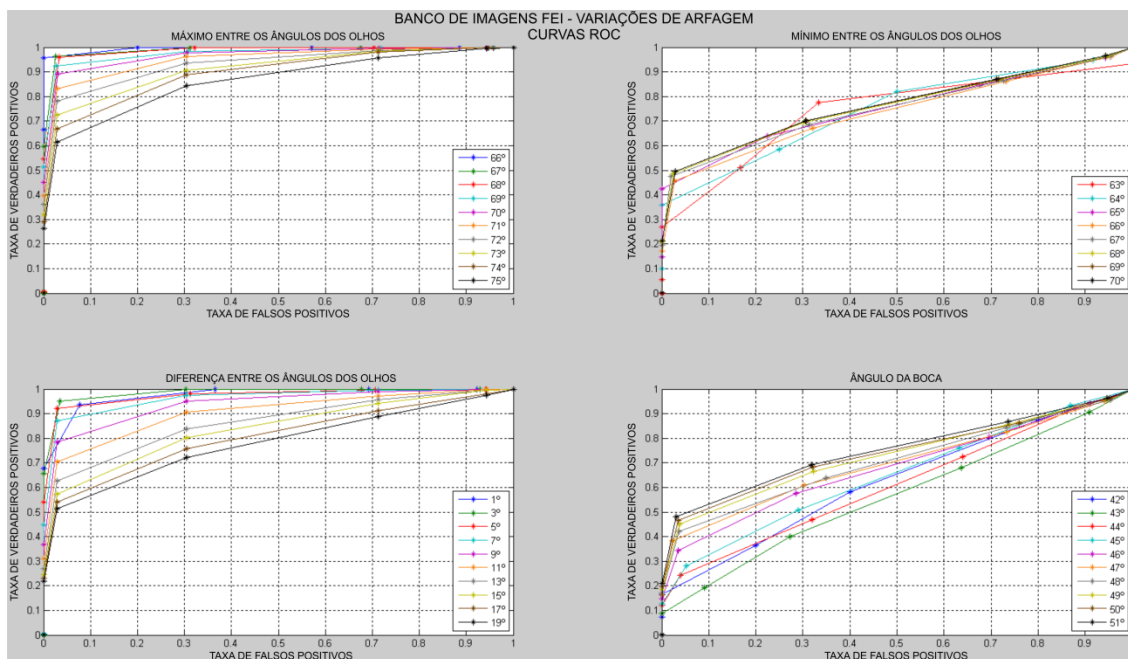


Figura 22. Curvas ROC para os atributos no banco de imagens FEI.

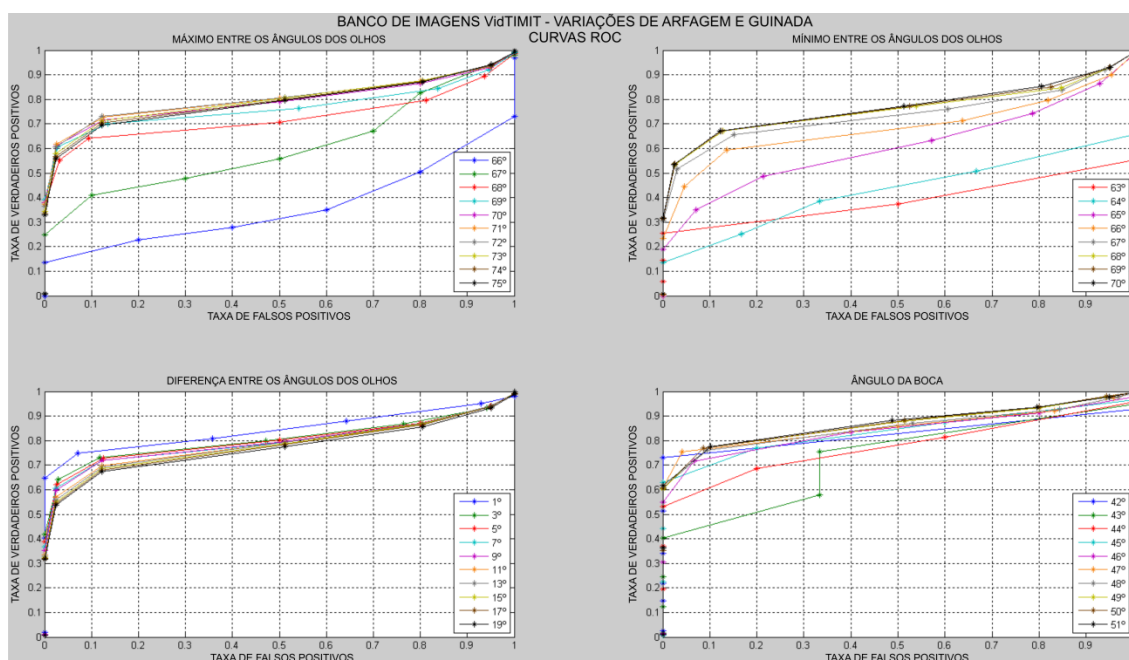


Figura 23. Curvas ROC para os atributos no banco de imagens VidTIMIT

Por outro lado, uma análise das curvas de ROC permitiu estabelecer os limites dos valores destes atributos dentro dos quais uma face pode ser considerada frontal. Assim, o ângulo máximo dos olhos deve estar na faixa compreendida entre  $63^\circ$  e  $71^\circ$  e a diferença entre os ângulos dos olhos não deve ser maior do que  $5^\circ$ .

A Figura 24 mostra um exemplo de aplicação destes limites para imagens pertencentes a um banco conhecido como *BioID Face Database*<sup>10</sup>, que foram etiquetados como frontais (FRONTAL) e não frontais (NON FRONTAL) pelo classificador.

É importante destacar que, o detector é flexível. Os valores limites indicados aqui podem ser relaxados dependendo da robustez do algoritmo de reconhecimento utilizado e das exigências da aplicação.

<sup>10</sup> Disponível em: <http://www.bioid.com/downloads/software/bioid-face-database.html>, último acesso em 3 de Março 2014.



Figura 24. Imagens classificadas de frontais e não frontais.

### 4.3. Protótipo de Vídeo Monitoramento – LVC

O Protótipo de Vídeo Monitoramento, denominado deste ponto em diante pelo termo “protótipo”, é um sistema de reconhecimento facial automático em vídeo baseado na arquitetura paralela proposta em (Bonfá, 2013). Este protótipo foi desenvolvido utilizando-se a linguagem em C++, e o ambiente de interface gráfica do Visual Studio 2010.

A estrutura simplificada do protótipo é apresentada na Figura 25. Envolve uma etapa de detecção de face e um rastreador, seguido de um detector de face frontal para avaliar a postura da imagem facial. Somente quando se detecta uma imagem facial frontal, executa-se o identificador, que envolve a normalização da face e extração dos atributos particulares do algoritmo de reconhecimento, como apresentado na seção 2.1.

Neste protótipo, o detector de faces usa um detector de objetos treinado com o algoritmo de Viola & Jones para localizar rostos (Viola, et al., 2001). O rastreador, baseado no algoritmo TLD<sup>11</sup> persegue as faces localizadas pelo detector de faces nos quadros de vídeo subsequentes. O detector de faces frontais avalia as faces fornecidas pelo rastreador para determinar se estas se encontram em posição frontal. Finalmente o identificador, realiza o reconhecimento usando o algoritmo baseado no LBP.

<sup>11</sup> *Tracking-Learning-Detection*, algoritmo de modelagem de objeto em tempo de execução.

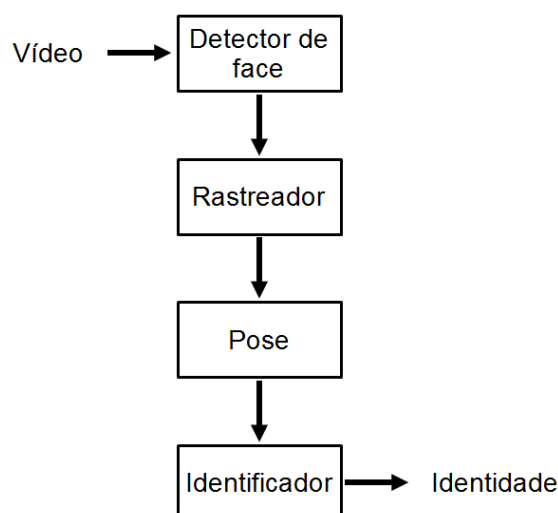


Figura 25. Modelo de reconhecimento facial do protótipo de Vídeo Monitoramento, adaptado de (Bonfá, 2013).

É importante ressaltar que a Figura 25 é uma representação bastante simplificada da arquitetura do protótipo, já que, todos os processos que o compõem operam de forma paralela.

Por outro lado, a interface gráfica do protótipo tem cinco espaços bem definidos, como é ilustrado na Figura 26: um painel de aquisição de dados, outro de registro de usuários, configuração de parâmetros de reconhecimento, aquisição de faces frontais e outro onde mostra métricas de desempenho.

No painel de aquisição de dados, o usuário pode escolher entre imagens provenientes da câmera, ou vídeos para realizar o reconhecimento, além de inicializar ou parar o sistema.

O painel de registro permite o cadastramento de novos usuários, o qual pode ser efetuado a partir de uma foto ou de uma câmera ligada ao computador; a eliminação de usuários do cadastro também é possível.

O painel de configuração de parâmetros para reconhecimento é composto pelos módulos de detecção facial e de detecção frontal. Em ambos dos casos o usuário pode modificar os valores para limitar a janela de procura rostos do detector de faces ou limitar a faixa de aceitação das faces em posição frontal.

O painel de aquisição de faces frontais permite estabelecer o diretório no qual as imagens faciais serão salvas.

Finalmente, o painel de métricas de desempenho apresenta ao usuário o desempenho computacional do protótipo em termos de quadros por segundo relativamente ao processamento da detecção e rastreamento de faces.



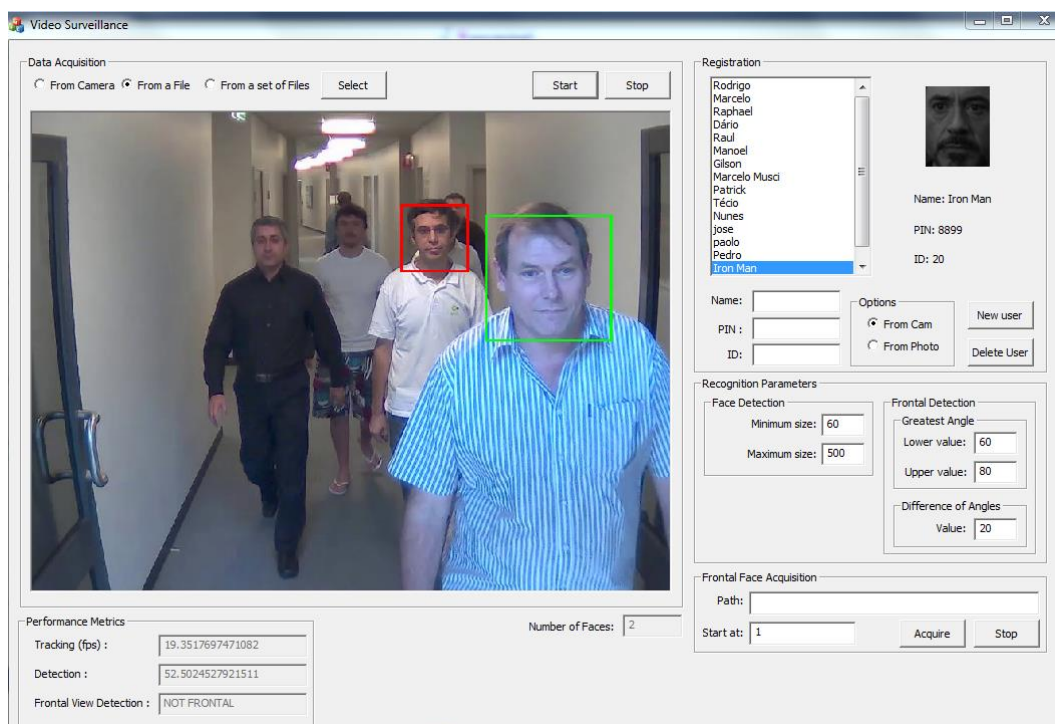


Figura 26. Protótipo de Vídeo Monitoramento, interface gráfica.

Nesta dissertação, o protótipo de Vídeo Monitoramento do LVC foi usado para coletar as imagens faciais normalizadas a partir de várias sequências de vídeo que superaram o teste de frontalidade conforme aos valores estabelecidos no limiar dos atributos para o detector de face frontal (ver seção 4.5.1).

#### 4.4. Biblioteca FACERECLIB

Facereclib<sup>13</sup> é uma ferramenta de código aberto construída sobre a biblioteca de processamento de sinais e aprendizado de máquina BOB<sup>14</sup>; que foi desenvolvido pelo Grupo de Biometria no Instituto de Pesquisa IDIAP (*Idiap Research Institute*) (Günther, et al., 2012).

A ferramenta foi desenvolvida em python e está projetada para realizar comparações entre uma variedade de algoritmos reconhecimento facial. A sua vantagem está na implementação de alguns dos algoritmos que conformam o estado da arte em reconhecimento facial, como:

- Autofaces,

<sup>13</sup> Disponível em: <https://github.com/idiap/facereclib>, último acesso em 26 de Março 2014.

<sup>14</sup> Disponível em: <https://github.com/idiap/bob>, último acesso em 26 de Março 2014.

- Fisherfaces,
- Bayesian Intrapersonal/Extrapersonal Classifier,
- Local Gabor Binary Pattern Histogram Sequences,
- Graph Matching,
- Gaussian Mixture Modeling,
- Inter – Session Variability Modeling,
- Probabilistic Linear Discriminant Analysis.

Entre as suas características, pode-se ressaltar a abordagem para dividir as imagens que compõem o banco de dados em grupos de treinamento e desenvolvimento segundo um protocolo dado. As imagens que pertencem ao **treinamento** são usadas para computar os meta-parâmetros correspondentes ao algoritmo de reconhecimento facial em uso, por exemplo, a matriz de projeção no caso das *autofaces* ou *fisherfaces*; ou, o *Universal Background Model* (UBM) no caso das misturas de Gaussianas. As imagens que pertencem ao **desenvolvimento** estão divididas a sua vez em dois subconjuntos: **cadastro**, que são usadas unicamente para criar modelos correspondentes às identidades na galeria; estas imagens podem ser (ou não) distintas das identidades usadas no treinamento; e **teste**, que são utilizadas como imagens de prova.

A Figura 27 exibe de forma aproximada esta abordagem, onde os quadros em vermelho correspondem ao conjunto de treinamento; em verde escuro se representam as imagens usadas para cadastro (galeria) e em verde claro as imagens utilizadas para os testes (provas).

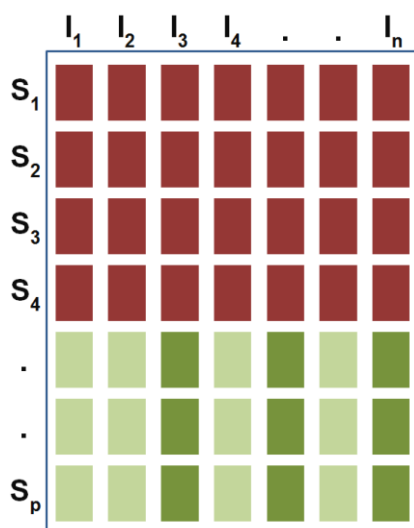


Figura 27. Nas linhas, o número de indivíduos no banco de dados; e nas colunas, o número de imagens por indivíduo no banco de dados.

Recomenda-se formar conjuntos de treinamento, cadastro e teste separados de modo a garantir que os resultados obtidos a partir das experiências com a biblioteca possam ser generalizados. Além disso, a não sobreposição das identidades no treinamento e no cadastro/teste contribui para que os resultados do sistema de reconhecimento não sejam tendenciosos.

#### **4.5.**

### **Procedimento Experimental**

É importante ressaltar que o propósito central desta pesquisa é comparar os algoritmos de reconhecimento facial apresentados no capítulo 3, e a partir desta comparação, determinar o algoritmo que apresente melhor acurácia no reconhecimento facial em vídeo.

Na busca deste objetivo foram consultados diversos bancos de dados para reconhecimento facial a partir de imagens estáticas e sequências de vídeo disponíveis na internet, como detalhado na seção 4.1. No entanto, apenas os bancos de vídeo e o banco de imagens ORL foram utilizados nos experimentos que permitiram avaliar os diversos algoritmos de reconhecimento facial implementados no *facereclib*.

Para avaliar corretamente estes algoritmos, teve-se que:

- Gerar novos bancos de dados com imagens faciais normalizadas, as imagens obtidas foram organizadas por identidade.
- Separar as imagens nos bancos normalizados em conjuntos de treinamento, cadastro (galeria) e teste (prova).
- Configurar os parâmetros corretos para cada algoritmo de reconhecimento.
- Avaliar de forma conjunta, os resultados fornecidos pelos algoritmos.

A seguir se detalha o procedimento experimental adotado na execução dos experimentos, que compreende quatro etapas: geração dos bancos de imagens faciais normalizadas, definição de protocolos, configuração de parâmetros dos algoritmos de reconhecimento e, finalmente, avaliação dos algoritmos de reconhecimento.

#### **4.5.1.**

### **Geração dos Bancos de Imagens Faciais Normalizadas**

Os bancos de imagens faciais normalizados foram criados a partir dos bancos de vídeo MOBIO, HONDA, ChokePOINT e VidTIMIT; além de quatro

fragmentos de filmes com 10 a 15 minutos de duração cada um, isto com o propósito de simular a captura de imagens faciais em ambientes reais não controlados.

Cabem aqui as seguintes considerações para cada banco de vídeo.

- No MOBIO, se selecionaram aleatoriamente quinze sequências de vídeo por participante;
- No HONDA, se consideraram duas sequências de vídeo por pessoa;
- No ChokePOINT, todas as sequências de vídeo foram utilizadas;
- No VidTIMIT, se selecionaram as sequências de vídeo das sessões que não formaram parte da criação do detector de face frontal, conforme detalhado na seção 4.1.

Assim, as imagens faciais normalizadas foram obtidas usando o protótipo de Vídeo Monitoramento do LVC, atribuindo aos parâmetros do detector de face frontal os seguintes valores:

- $A_{01\_min} = 60^\circ$ ;
- $A_{02\_max} = 80^\circ$ ;
- $A_{04} = 20^\circ$ ;

Isto permitiu a captura de doze imagens normalizadas por participante no caso dos bancos de vídeo (vide Figura 28), e sete por participante no caso dos filmes (vide Figura 29).



Figura 28. Da primeira a quarta linha, imagens faciais normalizadas correspondentes aos bancos HONDA, VidTIMIT, ChokePOINT e MOBIO.



Figura 29. Imagens faciais normalizadas correspondentes aos fragmentos de filmes.

No entanto, apenas foi possível extrair esta quantidade de imagens para 68 participantes do banco MOBIO, 39 participantes do VidTIMIT, 18 participantes do ChokePOINT, 19 participantes do HONDA e 67 participantes no caso dos fragmentos de filmes. Finalmente, reuniram-se as imagens faciais correspondentes aos bancos de MOBIO, VidTIMIT, ChokePOINT e HONDA num único banco, denominado de FUSION. De esta forma, além do banco de imagens estáticas ORL, se têm dois bancos de imagens faciais normalizadas: o FUSION que contém 12 imagens por cada um dos 144 participantes, e o FILMS que contém 7 imagens por cada um de seus 77 participantes.

É importante ressaltar que as imagens apresentadas na Figura 28 e Figura 29 foram normalizadas geometricamente (transformação em escala e rotação no plano) no protótipo, estas imagens tem o tamanho final de 80×64 pixels, considerando uma distância final de 33 pixels entre os olhos.

#### 4.5.2. Definição de protocolos

Os bancos de imagens foram organizados de modo a serem compatíveis com o facereclib. Assim, os bancos foram divididos em conjuntos de treinamento e desenvolvimento. No conjunto de **treinamento**, as imagens são usadas para calcular os meta-parâmetros dos algoritmos. No conjunto **desenvolvimento**, as imagens são usadas para compor a galeria e o conjunto de teste.

Duas configurações foram geradas para cada banco de dados:



- **Configuração 1:** nesta configuração, os conjuntos de treinamento e desenvolvimento estão disjuntos; isto significa que as imagens cujas identidades correspondem ao treinamento não formaram parte do desenvolvimento.

A distribuição das imagens para estes conjuntos, a exceção do banco ORL, foi realizada em base à quantidade de indivíduos em cada banco, assim, o 60% dos indivíduos, que correspondem a 322 imagens do banco FUSION e 1032 imagens do banco FILMS, compõem o conjunto de treinamento e o 40% restante pertencem ao conjunto de desenvolvimento.

Os subconjuntos de cadastro e testes, nesta configuração, se compõem do 60% e 40% das imagens, respectivamente, de cada indivíduo no conjunto de desenvolvimento.

No caso do banco ORL, 50% dos indivíduos foram usados para o treinamento e os 50% restantes para o desenvolvimento, neste último conjunto, o 50% das imagens por indivíduo foram considerados no cadastro e os 50% restantes no teste.

Na Tabela 2, se detalha a quantidade de indivíduos que compõem os conjuntos de treinamento e desenvolvimento. Na Tabela 3, é apresentado a quantidade de imagens por indivíduo empregadas para compor o subconjunto de cadastro (galeria) e teste (prova) no conjunto de desenvolvimento.

CONJUNTO	BANCO DE IMAGENS NORMALIZADO		
	ORL	FILMS	FUSION
TREINAMENTO	20	46	86
DESENVOLVIMENTO	20	31	58
TOTAL	40	77	144

Tabela 2. Configuração 1 – Indivíduos por conjunto de Treinamento e Desenvolvimento.

SUBCONJUNTO	BANCO DE IMAGENS NORMALIZADO		
	ORL	FILMS	FUSION
CADASTRO	5	4	7
TESTE	5	3	5
TOTAL	10	7	12

Tabela 3. Configuração 1 – Imagens por indivíduo no Cadastro e Teste.

- **Configuração 2:** nesta configuração, ao contrário da primeira, os indivíduos pertencentes ao conjunto de desenvolvimento também formaram parte do treinamento, mais explicitamente, os 40 indivíduos no banco ORL, 77 indivíduos no banco FILMS e os 144 indivíduos no banco FUSION, pertencem a ambas dos conjuntos.

Neste caso, as imagens que conformam os conjuntos estão distribuídas nas seguintes proporções: 30% das imagens por indivíduo de cada banco compõem o conjunto de treinamento e os 70% restantes ao conjunto de desenvolvimento, deste ultimo conjunto, o 60% das imagens foram utilizadas para compor a galeria de usuários e 40% pros testes dos algoritmos.

A Tabela 4, apresenta um resumo das imagens por banco distribuídas entre cada conjunto.

CONJUNTOS	BANCO DE IMAGENS NORMALIZADO		
	ORL	FILMS	FUSION
TREINAMENTO	4	3	4
DESENVOLVIMENTO (CADASTRO)	4	3	4
DESENVOLVIMENTO (TESTE)	2	1	4
TOTAL	10	7	12

Tabela 4. Configuração 2 – Imagens por individuo no Treinamento, Cadastro e Teste.

Em ambas das configurações, os indivíduos pertencentes aos bancos de dados FUSION e FILMS foram misturados de forma aleatória a fim de abrangerem uma ampla faixa de variações das imagens faciais, e desta forma simular situações as mais reais possível, considerando que os bancos MOBIO, HONDA, VidTIMIT e ChokePOINT apresentam características distintas de forma independente.

As duas configurações foram criadas para comparar o desempenho dos algoritmos quando se tem conhecimento prévio dos usuários. Isto foi realizado para determinar se os algoritmos fornecem um melhor desempenho em tarefas de verificação, em que um sistema tendencioso poderia apresentar melhor desempenho.

### 4.5.3.

### Configuração dos Parâmetros dos Algoritmos de Reconhecimento

Os desenvolvedores da biblioteca *facereclib* recomendam fazer uso das configurações dos parâmetros que estão incluídas na biblioteca que, segundo os autores, foram obtidas a partir de experiências realizadas em diferentes bancos de dados.

Seguiram-se, nesta dissertação, as recomendações dos autores. Desta forma, a Tabela 5 apresenta o resumo das configurações dos parâmetros para cada algoritmo no *facereclib*.

### 4.5.4.

### Análise dos Resultados e Elaboração das Conclusões

Como foi mencionado na seção 4.4, o *facereclib* fornece escores de reconhecimento por cada par modelo/teste comparado. Desta forma, todas as imagens de teste são comparadas com todos os modelos treinados. Assim, pode-se dividir o conjunto de resultados em dois grupos:

- **Clientes:** onde se tem os escores dos modelos/teste com as mesmas identidades,  $S_{cli}$ .
- **Impostores:** que contém os escores dos modelos/teste cujas identidades são diferentes,  $S_{imp}$ .

Para avaliar os resultados que serão obtidos das experiências se definem três métricas: a taxa de igual erro, curva ROC, e a curva CMC.

Estas métricas estão baseadas no cálculo da taxa de falsa aceitação (FAR, *False Acceptance Rate*) e da taxa de falsa rejeição (FRR, *False Rejection Rate*), definidas nas Equações (4-5) e (4-6) respectivamente, onde  $S_0$  é um limiar variável do escore de reconhecimento.

$$FAR(S_0) = \frac{|\{S_{imp} | S_{imp} \geq S_0\}|}{|\{S_{imp}\}|} \quad (4-5)$$

$$FRR(S_0) = \frac{|\{S_{cli} | S_{cli} \leq S_0\}|}{|\{S_{cli}\}|} \quad (4-6)$$

Num caso ideal, os valores das taxas de FAR e FRR seriam, em princípio, iguais à zero para algum valor de  $S_0$  estabelecido, o que indicaria o desempenho máximo do algoritmo de reconhecimento, já que todas as imagens de teste seriam corretamente identificadas e/ou verificadas. No entanto, na prática isto não acontece, e o que se procura é o valor de  $S_0$  que corresponda a um compromisso entre os valores das taxas de falsa aceitação e falsa rejeição.

ALGORITMO	PARAMETER	ORL	FILMS	FUSION
<b>EIGENFACES</b>	Dimensão do PCA	50	50	50
	Medida de similaridade	Distância Euclidiana		
<b>LDA</b>	Dimensão do PCA	100	100	100
	Dimensão do LDA	50	50	50
	Medida de similaridade	Distância Euclidiana		
<b>BIC</b>	Dimensão do subespaço intrapessoal	30	30	30
	Dimensão do subespaço extrapessoal	30	30	30
	Medida de similaridade	Diferença entre as probabilidades intra e extra pessoais		
<b>LGBPHS</b>	Tamanho dos blocos	10x10	10x10	10x10
	Escalas dos filtros de Gabor	5	5	5
	Orientações dos filtros de Gabor	8	8	8
	Medida de similaridade	Distância qui-quadrada		
<b>GABOR GRAPHS</b>	Escalas dos filtros de Gabor	5	5	5
	Orientações dos filtros de Gabor	2	2	2
	Tamanho da malha	8x11 px.	8x11 px.	8x11 px.
	Medida de similaridade	Média dos Gabor jets		
<b>GMM</b>	Tamanho dos blocos	12	12	12
	Sobreposição dos blocos	11	11	11
	Coefficientes da DCT	45	45	45
	Número de Gaussianas	512	512	512
	Medida de similaridade	Soma ponderada dos classificadores		
<b>ISV</b>	Tamanho dos blocos	12	12	12
	Sobreposição dos blocos	11	11	11
	Coefficientes da DCT	45	45	45
	Número de Gaussianas	512	512	512
<b>PLDA</b>	Dimensão do subespaço de F	16	16	64
	Dimensão do subespaço de G	16	16	64
	Dimensão do subespaço de PCA	150	150	150

Tabela 5. Configuração dos parâmetros para cada algoritmo no facereclib.

A escolha deste valor deve tentar equilibrar os erros de classificação característicos dos sistemas de verificação, já que, o valor da FAR diminui quando o valor da FRR cresce, ou, de forma inversa, FRR diminui quando FAR aumenta.

Um critério largamente utilizado, consiste em determinar do valor de  $S_0$  para o qual as taxas FAR e FRR atendam a Equação (4-7).

$$\hat{S}_0 = \arg \min_{S_0} |FAR(S_0) - FRR(S_0)| \quad (4-7)$$

Neste trabalho se usa a taxa de igual erro (EER, *Equal Error Rate*) como base para comparar o desempenho dos algoritmos de reconhecimento. Neste caso os valores de FAR e FRR são dados pela Equação (4-8). A hipótese subjacente é que, quanto menor o valor da EER melhor será a acurácia do algoritmo de reconhecimento.

$$EER = \frac{FAR(\hat{S}_0) + FRR(\hat{S}_0)}{2} \quad (4-8)$$

As características de operação de qualquer classificador podem ser observadas mediante a comparação da interação existente entre as amostras que são corretamente classificadas e as que são classificadas erroneamente. Esta comparação dá origem à curva ROC (*Receiver Operating Characteristic*), que basicamente relaciona as taxa de correta aceitação (CAR, *Correct Acceptance Rate*), definida pela Equação (4-9), e a taxa FAR para diversos valores de  $S_0$ .

$$CAR(S_0) = 1 - FRR(S_0) \quad (4-9)$$

Assim, nesta dissertação, a curva ROC é adotada como segunda métrica para avaliar o desempenho dos algoritmos de reconhecimento quando operam em tarefas de verificação. Dado que neste modo de operação, os algoritmos só comparam o modelo do cliente diretamente com o modelo do suposto cliente (teste), a análise desta curva é realizada do ponto de vista dos classificadores binários. Assim, um algoritmo de reconhecimento terá bom desempenho sempre que seus valores de FAR estiverem próximos de 0, entanto que seus valores de CAR estiverem próximos de 1.

Finalmente, as curvas CMC (*Cumulative Match Characteristic*) são usadas para avaliar o desempenho dos algoritmos de reconhecimento quando operam em tarefas de identificação; esta curva permite observar o comportamento de cada algoritmo baseado na probabilidade de encontrar a identidade correta entre as primeiras  $r$  identidades com maior coincidência na base de dados.

Nesta curva, os escores de reconhecimento fornecidos para uma imagem de teste são ordenados de forma decrescente, desta forma, se a comparação com maior escore corresponde à identidade correta então se tem um acerto de ordem-1 (rank-1). Se ao contrário, a identidade correta está entre as cinco



primeiras da lista então se tem um acerto de ordem-5 (rank-5). Este processo se repete para todas as imagens de teste e ao final se computa a porcentagem das identificações corretas para cada ordem (rank). O gráfico da curva CMC corresponde às porcentagens acumuladas com respeito de cada ordem (rank).