

Proposta de uma Metodologia para a Construção de um Banco de Imagens Faciais Normalizadas

Flávio das Chagas Prodossimo
Universidade Tecnológica Federal do Paraná
Av. Sete de Setembro, 3165
frovas@gmail.com

Chidambaram Chidambaram
Universidade do Estado de Santa Catarina
Rua Luiz Fernando Hastreiter, 180
chidambr@gmail.com

Heitor Silvério Lopes
Universidade Tecnológica Federal do Paraná
heitorslopes@gmail.com

Abstract

The area of computer vision is one of the most studied and is the basis for studies of face recognition. This area requires facial images under different conditions. Thus, the objective of this paper is to show the methodology applied to capture facial images which will be used in the experiments of various applications, for example, standardization, detection and face recognition. In this work, equipment configuration, the image measures and different poses for the capture of face images are presented. Several face images were captured and at then they were normalized according to international standards. From this work, it can be noted that predefined procedure and standards are important to capture images so that face recognition and other related tasks can be effectively performed.

1. Introdução

Com surgimento de novas tecnologias e o aumento da preocupação nas questões de segurança, a identificação e o reconhecimento de pessoas usando imagens passou a ser uma das tarefas mais comuns e, ao mesmo tempo, se tornou essencial tanto na área de segurança privada quanto na pública. Para construção de sistemas de reconhecimento são necessárias várias imagens em diversas condições. Portanto, a construção de base de imagens se torna necessária a fim de realizar os experimentos utilizando diferentes métodos da área de reconhecimento facial.

No presente trabalho, o objetivo é reunir as informações necessárias para captura de imagens faciais com base em documentos nacionais e internacionais e, a partir disto, propor uma metodologia para capturar e normalizar estas imagens.

O estabelecimento de padrões tem como objetivo facilitar a análise detalhada e a verificação de imagens. Isso

pode auxiliar tanto os examinadores da segurança pública quanto a identificação através de métodos computacionais. Esses padrões objetivam definir um conjunto de orientações e recomendações que farão com que as imagens capturadas tenham uma qualidade mínima e suficiente para a verificação e a identificação de imagens faciais.

O trabalho também apresenta a metodologia que foi utilizada na construção de uma base de imagens faciais. No decorrer do documento é possível ver a forma como o ambiente deve ser organizado com os equipamentos para que as imagens sejam capturadas. Está descrito, também, como as imagens são capturadas, as rotações e translações da face, variações de iluminação, e a sequência em que isto é feito. Essas imagens estão sendo utilizadas em estudos de processamento facial, tais como normalização, detecção e reconhecimento.

Outra tarefa importante tratada neste trabalho é a normalização das imagens. A normalização das imagens inclui ajuste da rotação, escala, posicionamento da face, iluminação e cor do fundo da imagem [9]. Esse processo consta da padronização das imagens seguindo as padrões de normalização.

A principal motivação para realizar este estudo é a de ter encontrado somente um estudo semelhante em âmbito nacional [3]. Outra motivação que levou a este estudo é a falta de bases de imagens faciais que possuam imagens coloridas e com resolução grande, acima das resoluções das bases existentes e acima do indicado em [7] que é de 768x1024 pixels. Segundo o documento da NPPIA [7], para que o reconhecimento automático seja implementado amplamente, o mais provável é que os padrões atuais devem ser atualizados, iniciando pelo aumento de resolução, captura de múltiplas imagens em vários ângulos e até as imagens 3D. Foram pesquisadas diversas bases na internet e no estudo de Gross [6], mas não foi encontrada nenhuma base pronta

com imagens nas condições esperadas. Portanto, decidiu-se a construção de uma base com condições específicas alinhadas às necessidades do projeto de reconhecimento facial que está em desenvolvimento.

2. Fundamentação Teórica

Para estudo de técnicas de visão computacional é necessário capturar imagens em diversas condições distintas. Para projetos de reconhecimento facial as bases de imagens devem ser compostas de várias posições das faces de pessoas, com isso é possível testar algoritmos de reconhecimento. A partir do momento em que as imagens estão disponíveis, existem etapas importantes para o reconhecimento facial. Uma delas é a normalização de imagens, que pode ser entendida como um pré-processamento das faces para utilizar os métodos de reconhecimento. Tanto para a captura quanto para a normalização de imagens faciais existem normas a ser seguidas [7] e [9].

Várias bases de imagens faciais foram analisadas para se obter informações sobre as imagens disponíveis para realização de reconhecimento facial, as bases mais relevantes podem ser vistas na Tabela 1. Levando em consideração estes dados e, através do estudo de Gross [6], no decorrer do artigo, procura-se expor como uma base de imagens foi construída.

Base de Dados	Pessoas	Imagens	Resolução	Poses
FERET [8]	1199	30000	256x384	20
CVL Face [10]	114	798	640x480	
PIE [5]	68	41368		13
BioID [1]	23	1521	384x286	
M2VTS [2]	295	2460	720x576	
FEI [3]	200	2800	640x480	

Tabela 1. Comparativo entre bases de dados de imagens faciais.

Nesta tabela pode-se observar a variedade de poses e o grande número de pessoas que participaram da base de imagens FERET [8], base mais conceituada e utilizada em estudos da atualidade para pesquisas de reconhecimento facial. Isso pode ser visto nos artigos científicos relacionados ao reconhecimento facial, como [11], [12] e [14]. Outras bases bastante conhecidas são mostradas na Tabela 1, com menor número de pessoas mas com um grande número total de imagens, como é o exemplo da PIE [5]. Outra base mostrada na tabela é a base brasileira da FEI [3], que conta com 14 imagens por pessoa totalizando 2800 imagens. Porém todas as bases estudadas possuem pequena resolução.

Para a organização do ambiente de captura de imagens foi utilizado como base o experimento de Rubinfeld [13], que mostra os principais detalhes sobre a captura de imagens utilizada por ele. Este estudo ainda mostra alguns experimentos com câmeras diferentes, formas diferentes de iluminação, mas sempre tentando manter um padrão entre as imagens capturadas.

A normalização de imagens é tratada, principalmente, no artigo da NIST [9], onde também são definidos os padrões que foram adotados para a padronização de imagens faciais. Os autores de [3] também tratam da captura e alinhamento das imagens. Neste trabalho, o alinhamento das imagens foi feito manualmente.

3. Normas para Captura e Normalização de Imagens

Com o objetivo de criar um banco de imagens para projetos de pesquisa na área de reconhecimento facial, as imagens foram capturadas de acordo com algumas normas, estabelecidas pelos documentos [7] e [9] e pelo relatório técnico M1/03-0114 [4].

Durante a captura da imagem facial, a pessoa deve olhar diretamente para câmera, sem que tenha qualquer tipo de inclinação do ombro ou corpo. Se existir qualquer inclinação, deve ser inferior a +/- 5° em relação à posição frontal, ou seja, tanto no lado esquerdo quanto no lado direito. A imagem deve mostrar claramente os olhos, principalmente a íris e a pupila. A presença de cabelos longos pode prejudicar as imagens obstruindo a parte da face. Isso deve ser evitado prendendo o cabelo ou colocando atrás das orelhas. Para mostrar claramente os olhos também é aconselhável não utilizar óculos [7].

Além da imagem frontal, outras imagens devem ser capturadas segundo Islam [7]: imagens das laterais direita e esquerda da face com rotação de 45° e 90°, e imagens frontais da face com inclinação entre 20° e 45° para baixo e para cima.

Outras normas importantes para o processo de captura de imagens são em relação ao equipamento a ser utilizado, à iluminação e ao tipo de fundo da imagem. Uma condição de iluminação adequada deve ser empregada para proporcionar uma iluminação uniforme à imagem que está sendo capturada e ao fundo. A iluminação deve ser distribuída igualmente em todos os lados da imagem sem gerar sombra e o local de captura deve estar totalmente isolado de iluminação natural (iluminação externa) para que não comprometa o experimento. O fundo da imagem deve ser limpo e sem objetos adicionais, para que, depois de capturada, a imagem mostre claramente as bordas da face em relação ao fundo. O fundo deve ter 18% de cinza e deve ser posicionado a aproximadamente 70 centímetros da pessoa.

O passo mais importante para a normalização de imagens é o posicionamento dos olhos. Através de dois pontos (pontos centrais das duas pupilas) é possível realizar grande parte da normalização. Segundo as normas, os olhos devem estar alinhados na mesma altura, para isso é utilizada a rotação da imagem. Em seguida, segundo o documento da NIST [9], a distância entre os olhos deve ser de 60 pixels, ou seja, é necessário realizar o redimensionamento da imagem. A imagem normalizada deve ter a resolução na proporção de 1:1.25 (ou 4:5) [7], com isso é necessário recortar a imagem para ficar com a resolução recomendada. A face deve ser posicionada na imagem de acordo com as medidas mostradas na figura (Figura 1) abaixo.

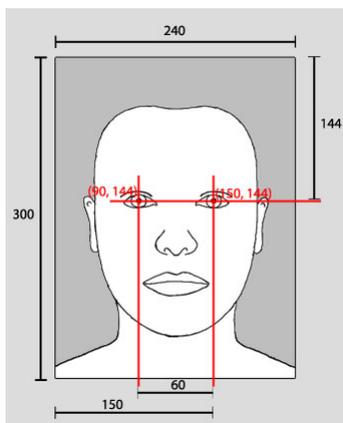


Figura 1. Dimensões de uma imagem normalizada canônica (em pixels).

Na Figura 1 pode-se notar que a distância entre os olhos é de 60 pixels e o tamanho total da imagem obedece a proporção 1:1.25, 240x300. Além disso o posicionamento da face é centralizada na imagem com os olhos localizados nos pontos (91, 144) e (150, 144). A Figura 1 representa uma imagem normalizada canônica, se for preciso uma resolução maior a imagem normalizada deve seguir as proporções indicadas das medidas desta imagem.

4. Metodologia

Nas subseções seguintes são mostrados os procedimentos utilizados para a captura e normalização de imagens neste trabalho.

4.1. Captura de Imagens

4.1.1. Equipamentos Para a captura de imagens foi utilizada uma câmera digital posicionada sobre um tripé para

que a distância entre a câmera e as pessoas não fosse alterada. A resolução de cada foto é de 2592x1944 pixels. Esta resolução foi escolhida por ser maior do que as existentes na literatura. E também permite ajustes posteriores quanto ao tamanho e à resolução, ou seja, será possível processar a imagem capturada para obter um tamanho menor, sem perder informações.

A câmera utilizada foi uma Sony Cyber-Shot (DSC-S3000) de 10.1 Mega Pixels. Porém, as imagens foram capturadas com 5 Mega Pixels. Para todas as posições de face a câmera estava com um valor de ampliação de três vezes (*zoom* de 3x). A ampliação é necessária para obter somente a face da pessoa, mas, ao mesmo tempo, a câmera deve ser posicionada numa distância da face para obter uma melhor nitidez e para evitar o aparecimento de sombras.

Um plano de fundo de cor fixa foi utilizado para a captura. A cor do fundo é cinza, sendo que a porcentagem aconselhada para captura de imagens faciais é de 18%, segundo Rubinfeld e Wilson [13].

A iluminação do ambiente de captura das imagens foi feita artificialmente, algumas imagens com fonte de iluminação própria para captura de imagens e outras com a iluminação da sala, caracterizando dois ambientes de captura de imagens: ambiente controlado e ambiente não controlado. Além dos ambientes controlado e não controlado, foram capturadas imagens com iluminação lateral para gerar sombra somente em um lado da face. O local de captura foi isolado de iluminação externa para não comprometer o experimento.

4.1.2. Posicionamento do Equipamento Com base nos estudos de Gross [6] e Rubinfeld [13] os equipamentos devem ser posicionados como mostrado na Figura 2.

Primeiramente, a cadeira foi posicionada um pouco a frente do painel de fundo para evitar sombras. Os pés da cadeira estão a 1,2 metros do painel, fazendo com que a pessoa fique a aproximadamente 70 centímetros do fundo.

A câmera foi posicionada a 0,85 metros do centro da cadeira, de modo que somente a face e uma parte do ombro da pessoa apareçam na imagem adquirida. Posteriormente a câmera foi colocada a 1,75 metros do fundo da cadeira, para algumas fotos em escala. A variação das distâncias foi escolhida para ter uma maior variabilidade de imagens que podem ser utilizadas nos estudos de reconhecimento facial.

Com o objetivo de capturar as imagens com variações de iluminação, utilizou-se duas fontes de iluminação próprias, além da própria iluminação do ambiente da sala. Uma das fontes de iluminação foi posicionada a 2,5 metros do fundo da cadeira, de forma centralizada com a face. Esta posição evitou que aparecessem sombras no painel. A outra fonte foi localizada na lateral da cadeira, a 1 metro de distância e 1,7 metros de altura. Quando a fonte está ligada causa som-

bra em um lado da face. Neste estudo, isto serve para causar uma variação significativa de iluminação na imagem.

Pontos fixos foram definidos no chão para que o tripé/câmera, cadeira e iluminação sejam posicionados sempre no mesmo lugar. Isto é feito para que a distância entre a câmera e a face das pessoas seja sempre a mesma e a iluminação não sofra alterações. Além destes pontos, mais dois foram definidos para as fotos laterais (90° e 105-115°). De fato, estes pontos servem para auxiliar na captura da lateral da face. Para as fotos com a face voltada para cima foi definido um ponto junto à luz artificial para que as pessoas olhassem sempre para o mesmo ponto. Para as fotos com a face voltada para baixo outro ponto foi definido em um dos pés do tripé.

As posições do tripé assim como o posicionamento dos outros equipamentos podem ser melhor visualizados na Figura 2.

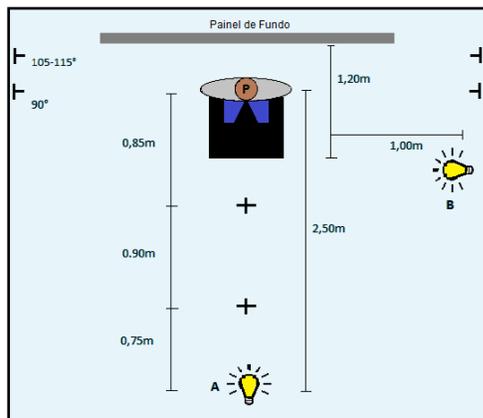


Figura 2. Posicionamento do equipamento.

Na Figura 2 os símbolos marcados com 'A' e 'B' são as iluminações próprias para a captura de imagens. A letra 'P' representa o local onde fica a pessoa. As marcações com o sinal de '+' são as duas posições do tripé e as marcas laterais (90° e 105-115°) são marcações para o voluntário olhar durante a captura das imagens laterais.

4.1.3. Poses das Faces As imagens foram capturadas da seguinte forma: imagens com iluminação controlada com a fonte de iluminação 'A' acesa e a iluminação ambiente desligada; imagens com iluminação não controlada com as fontes de iluminação 'A' e 'B' desligadas e a iluminação do ambiente ligada.

Durante a captura, é importante verificar que não existe cabelo cobrindo partes importantes como sobrancelhas e olhos e, principalmente, nas fotos laterais, a orelha deve aparecer descoberta.

O número total de imagens capturadas por pessoa neste estudo foi de 27, contando com todas as variações de iluminação e de pose. De modo geral, as imagens capturadas podem ser classificadas em duas categorias: 14 imagens com iluminação controlada e 13 imagens com iluminação não controlada. Foram capturadas imagens frontais com inclinações, imagens laterais e outras imagens com a câmera em posições diferentes. As poses das imagens capturadas são mostradas na Tabela 2.

4.1.4. Procedimento para Captura das Imagens Antes de iniciar a captura das imagens, num primeiro momento, os participantes devem assinar um termo de consentimento, dizendo que concordam e autorizam o uso de suas imagens para fins científicos.

Na Tabela 2 pode-se observar a codificação utilizada para identificar a pose de cada foto. Estes códigos são precedidos por outros números que representam o tipo de iluminação e a pessoa, de modo que a identidade possa ser preservada. Os códigos estão na sequência em que as imagens foram capturadas.

Código	Posição
01	90° Esquerda
02	105° Esquerda
03	Frontal
04	Frontal com Expressão
05	Inclinação Direita
06	Inclinação Esquerda
07	Face voltada para Cima
08	Face voltada para Baixo
09	90° Direita
10	105° Direita
11	2,5m Frontal
12	2,5m Inclinação Direita
13	2,5m Inclinação Esquerda
14	Frontal com luz lateral

Tabela 2. Sequência de captura das Imagens.

A captura é realizada em duas etapas, uma com iluminação artificial ou controlada e outra com iluminação do ambiente ou não controlada. Apenas a imagem 14 (Tabela 2) é capturada somente com iluminação controlada utilizando, também, a fonte lateral de iluminação. Nas imagens 01 e 02, captura-se a face lateral direita do voluntário em 90 graus e com uma leve inclinação para direita em 105 a 115 graus. O mesmo é feito para as imagens 09 e 10, desta vez com a face lateral esquerda. As imagens 03 até 06 são frontais, com variações de inclinação e expressão facial. A imagem 07 é capturada com a pessoa olhando para cima e a 08 olhando para baixo. As ima-

gens 11, 12 e 13 são semelhantes às imagens 03, 05 e 06, respectivamente, porém com a câmera com uma distância de 1,75 metros. A última, número 14, é capturada com a pessoa olhando para a câmera e com uma luz lateral gerando sombra em um dos lados da face.

Contando com a utilização de iluminação artificial o flash da câmera deve estar desativado e o tempo do obturador da máquina, tempo de exposição, *zoom*, entre outras características da câmera devem ser fixos. Com isso, será possível ter um controle maior sobre as imagens capturadas. Este controle é necessário para as devidas comparações com outras bases de dados ou entre algoritmos e para validação da base. Outra forma de validação é a utilização de rotações e iluminação fixas. Porém são capturadas imagens com rotações diferentes e variações de iluminação para se aproximar da realidade, onde as imagens são capturadas sem controle.

4.2. Normalização das Imagens

Nas subseções seguintes são apresentados os procedimentos utilizados para a normalização das imagens capturadas. O algoritmo de normalização foi implementado em C/C++ utilizando a biblioteca OpenCV¹. Neste trabalho não foram implementadas a normalização de iluminação e fundo. E o centro dos olhos são obtidos manualmente.

Como especificado em [7] e [9] as imagens podem ter o tamanho de 768 X 1024 pixels mas, como já foi dito, as imagens neste trabalho foram capturadas com uma resolução maior. Porém, as novas medidas devem obedecer à proporção com as medidas das normas. As imagens devem ser normalizadas no tamanho da imagem canônica (Figura 1) ou com as medidas proporcionais, se desejável uma resolução maior. Neste estudo as imagens foram normalizadas com as medidas cinco vezes maior que a imagem canônica, logo a distancia entre os olhos é de 300 pixels, sendo que os olhos devem estar nas coordenadas (450, 720) e (750, 720) e o tamanho da imagem é de 1200x1500 pixels.

4.2.1. Rotação Esta etapa consta em alinhar os dois olhos da face. Para isso calculou-se a diferença de pixels entre os dois olhos. Primeiramente a diferença entre os valores x dos pontos e em seguida os valores y, representando os catetos de um triângulo. Com os catetos é possível calcular a hipotenusa e com esses valores é possível calcular o ângulo entre os olhos através da equação trigonométrica do arco cosseno.

4.2.2. Redimensionamento Para que existam 300 pixels entre os olhos é necessário redimensionar a imagem. Neste

passo é feita a divisão da distância entre os olhos por 300. Ou seja, como tem-se o valor da hipotenusa basta dividi-lo por 300 obtendo, assim, o valor de escala necessário para causar a ampliação ou redução da imagem.

4.2.3. Centralização e Recorte da Face Através do valor da escala da etapa anterior é possível calcular onde os pontos centrais estão localizados na nova imagem. Com isso é possível calcular o ponto central dos olhos para centralizar a face na imagem. Com isso basta cortar a imagem nos redores da face, mantendo ela centralizada e obedecendo as medidas citadas anteriormente. A imagem final, normalizada, deve ter o tamanho de 1200x1500 pixels.

5. Resultados

Na Figura 3 é mostrado o resultado da captura das 27 imagens de um voluntário. As imagens são mostradas na sequência determinada pela Tabela 2, sendo as 14 primeiras com iluminação controlada e as outras 13 com iluminação não controlada.

Encontrou-se dificuldade neste trabalho em relação ao posicionamento dos equipamentos, principalmente, para eliminar qualquer sombra na imagem. Este problema envolvia distância entre a câmera e a pessoa, entre a pessoa e o painel de fundo e altura da a fonte de iluminação.

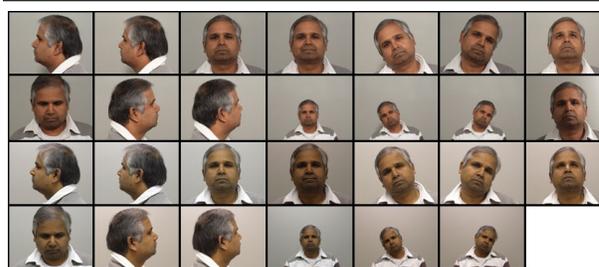


Figura 3. Resultado da captura de imagens.

A Tabela 3 mostra um resumo das características da base criada. Dentre o total de imagens que foram capturadas atualmente 65% são de pessoas do sexo masculino e 35% do sexo feminino. Quase 54% da base de imagens é formada por pessoas de faixa etária entre 21 e 30 anos. As imagens da base contém três variações de iluminação, levando em consideração a iluminação lateral, e um total de 14 poses diferentes para cada voluntário.

As imagens frontais passaram por um processo de validação. Este processo consta em verificar se elas estão dentro dos limites estabelecidos pelas normas citadas anteriormente. Esta validação utiliza o algoritmo de normalização de faces, sendo assim, as imagens foram validadas e a normalização foi testada. A tolerância

¹ *Open Source Computer Vision*. Disponível em: <http://opencv.willowgarage.com/wiki/>.

Resolução	2592x1944
Número de Pessoas	150
Número de Poses	14
Formas de Iluminação	3
Imagens por Pessoa	27
Total de Imagens	4050

Tabela 3. Características da Base.

de inclinação para os lados é de 5° e, em aproximadamente 9% das imagens esse valor foi maior. Dentre as imagens dentro do permitido a maior parte estava com inclinação abaixo de 3°. A maior inclinação das imagens frontais capturadas foi de 8,5° e a menor inclinação foi de 0,6°. Com isso é possível notar que a captura de imagens foi eficiente para grande parte das faces.

Seguindo o padrão das normas e através das equações citadas anteriormente as imagens da base de faces foram normalizadas. Vários testes foram feitos com as imagens rotacionadas para melhor visualização do processo de normalização. Na Figura 4 são mostrados dois exemplos deste processo. A primeira e a segunda imagem rotacionadas, sendo que a segunda também está em escala menor.



Figura 4. Resultado da normalização de imagens.

Nesta figura é possível ver a normalização passo a passo: na primeira coluna estão as imagens originais; na segunda coluna, as imagens rotacionadas com os olhos na mesma linha; na terceira, o redimensionamento da imagem para a diferença entre os olhos ficar em 300 pixels; e por último a imagem recortada para o tamanho de 1200x1500. Neste exemplo, a rotação para as imagens foi de 22° e 18°, respectivamente, e o zoom necessário para redimensionar as imagens para o tamanho correto foi de 0,82 e 1,57.

6. Conclusões

O objetivo deste artigo foi mostrar uma metodologia de captura de imagens faciais para a construção de uma base de imagens. Este estudo também apresenta a adoção de normas

e orientações e, ainda, reforça a necessidade de diversos cuidados para capturar as imagens. As imagens estão sendo utilizadas em estudos de normalização, detecção e reconhecimento facial. A principal dificuldade neste trabalho foi em relação ao posicionamento dos pontos de iluminação para eliminar a sombra nas imagens. Outro ponto que exigiu um esforço adicional foi em encontrar voluntários para colaborar na captura de imagens. O processo de normalização deste trabalho não tratou da normalização da iluminação nem do fundo de imagem. Isso será tratado em trabalhos futuros. Embora a base tenha alcançado, até o momento, um número razoável de imagens, o processo de captura ainda continua na tentativa de obter uma proporção equilibrada nas imagens em algumas categorias. Com este estudo percebe-se que a criação de um banco de imagens de alta qualidade é um procedimento complexo e criterioso.

A continuação deste trabalho tem como objetivo a detecção automática dos olhos. Com isso todo o processo de normalização passa a ser automático, facilitando o estudo das demais técnicas de reconhecimento facial.

Referências

- [1] BioID. Bioid face database, 2011.
- [2] C. H. Cham. The extended m2vts database, 2006.
- [3] L. L. de Oliveira Junior and C. E. Thomaz. Captura e alinhamento de imagens: Um banco de faces brasileiro. Technical report, Centro Universitário da FEI, 2005.
- [4] P. Griffin. Face recognition format for data interchange. Technical report, Identix Corporate Research Center, 2003.
- [5] R. Gross. Pie database, 2000.
- [6] R. Gross. Face databases. *Handbook of Face Recognition*, 2005.
- [7] R. Islam. *Police Standard for Still Digital Image Capture and Data Interchange of Facial/Mugshot and Scar, Mark and Tattoo Images*. National Policing Improvement Agency, 2.0 edition, Maio 2007.
- [8] NIST. The feret database, 2004.
- [9] NIST. *Face Recognition Format for Data Interchange - Best Practices*, 2007.
- [10] P. Peer. Cvl face database, 1999.
- [11] P. J. Philips and S. A. Rizvi. The feret evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1090 – 1104, 2000.
- [12] S. A. Rizvi, P. J. Philips, and H. Moon. The feret verification testing protocol for face recognition algorithms. Technical report, NIST, 1998.
- [13] M. Rubinfeld and C. Wilson. Gray calibration of digital cameras to meet nist mugshot best practice. 1999.
- [14] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:775 – 779, 1997.