

3 Fundamentos Teóricos

Neste capítulo são apresentados os principais algoritmos de reconhecimento facial disponíveis na literatura, alguns dos quais, como por exemplo, *Probabilistic Linear Discriminant Analysis*, *Inter Session Variability Modeling*, *Gasussian Mixture Models* e *Gabor Graphs*, conformam o estado da arte.

Do ponto de vista do aprendizado de máquina, os algoritmos de reconhecimento precisam ser treinados para criar modelos a partir dos quais seja possível gerar uma medida de similaridade, neste contexto é preciso cumprir quatro requerimentos: (1) ter um banco de dados disponível para treinar os algoritmos; (2) normalizar as imagens faciais do banco de dados; (3) extrair atributos característicos das imagens faciais normalizadas; (4) treinar os algoritmos de reconhecimento em base aos atributos extraídos.

Nas seções seguintes, são apresentados os algoritmos, assim como, a abordagem empregada por cada um deles para gerar modelos das faces que permitem realizar o reconhecimento.

3.1. Autofaces

As *autofaces* são os vetores que caracterizam as variações contidas numa coleção de imagens faciais, os quais permitem codificar e comparar as imagens de uma forma individual e holística. Matematicamente, as *autofaces* representam as componentes principais da distribuição de um conjunto de imagens faciais, ou equivalentemente, os autovetores da matriz de covariância destas imagens. Desta forma, qualquer imagem facial pode ser representada pela combinação linear das *autofaces* associadas aos autovalores com maior valor.

Inicialmente, Kirby e Sirovich (Kirby, et al., 1990) usaram as componentes principais para representar as imagens faciais num espaço N-dimensional. Assim, uma imagem normalizada da face $I_N(x, y)$ de tamanho $h \times w$ pixels pode ser representada por um vetor em um espaço hw -dimensional simplesmente escrevendo o valor de cada um de seus pixels num determinado ordem.

Turk e Pentland (Turk, et al., 1991), aproveitaram esta aproximação para detectar e reconhecer as faces contidas numa imagem qualquer. Assim, dada uma matriz \mathbf{F} de $N \times M$ dimensões, cujas colunas $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_M$ representam às imagens faciais vetorizadas, M é o número de faces (classes, e neste contexto as identidades) e $N = hw$ o número total de pixels de cada imagem, as *autofaces* podem ser calculadas da seguinte forma:

1. Calcula-se a média $\boldsymbol{\mu}$ das imagens faciais:

$$\boldsymbol{\mu} = \frac{1}{M} \sum_{i=1}^M \mathbf{f}_i \quad (3-1)$$

2. Calcula-se a diferença $\boldsymbol{\phi}_i$ entre cada imagem e $\boldsymbol{\mu}$:

$$\boldsymbol{\phi}_i = \mathbf{f}_i - \boldsymbol{\mu} \quad (3-2)$$

3. Calcula-se a matriz de covariância $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \frac{1}{M} \sum_{i=1}^M \boldsymbol{\phi}_i \boldsymbol{\phi}_i^T = \mathbf{A} \mathbf{A}^T, \quad (3-3)$$

onde, $\mathbf{A} = [\boldsymbol{\phi}_1 \ \boldsymbol{\phi}_2 \ \dots \ \boldsymbol{\phi}_M] \in \mathbb{R}^{N \times M}$.

4. Computam-se a matriz de autovetores $\mathbf{V}_{pca} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_e]$ e o vetor de autovalores $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_e]$ correspondentes à matriz de covariância $\boldsymbol{\Sigma}$, onde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_e \geq 0$, e $e \leq M$.

Aqui, \mathbf{V}_{pca} contém os autovetores \mathbf{v}_e (*autofaces*) correspondentes aos maiores λ_e (autovalores) que melhor representam a variabilidade em \mathbf{F} , e que formam o subespaço e -dimensional de nome *espaço de faces* com origem em $\boldsymbol{\mu}$.

O reconhecimento se realiza pela comparação entre as projeções \mathbf{x} e \mathbf{x}_i numa imagem de teste \mathbf{f} e cada uma das imagens faciais em \mathbf{F} no espaço de faces. As projeções \mathbf{x} e \mathbf{x}_i vêm dadas pelas Equações (3-4) e (3-5) respectivamente.

$$\mathbf{x} = \mathbf{V}_{pca}^T (\mathbf{f} - \boldsymbol{\mu}) \quad (3-4)$$

$$\mathbf{x}_i = \mathbf{V}_{pca}^T (\mathbf{f}_i - \boldsymbol{\mu}) \quad (3-5)$$

Finalmente, uma identidade de \mathbf{F} é atribuída a \mathbf{f} se a distância euclidiana entre as projeções \mathbf{x} e \mathbf{x}_i , denotada pela Equação (3-6), é menor que determinado limiar S_0 .

$$d^2(\mathbf{x}, \mathbf{x}_i) = \|(\mathbf{x} - \mathbf{x}_i)\|^2 \quad (3-6)$$

3.2. Fisherfaces

O algoritmo dos *fisherfaces* procura pelo subespaço \mathbf{W} que melhor caracterize as variações intra-classe e inter-classe de um conjunto de imagens faciais a fim de obter uma melhor capacidade discriminativa. Os vetores de base deste subespaço são conhecidos como *fisherfaces* e permitem comparar as imagens faciais de uma forma global.

Ao contrário das *autofaces*, onde o foco é procurar pelo subespaço que melhor represente à variabilidade global contida num conjunto de imagens faciais, o principal interesse dos *fisherfaces* é determinar o subespaço onde as amostras das imagens faciais duma mesma classe (identidade) fiquem mais próximas entre si, enquanto que aquelas de diferentes classes estejam as mais afastadas possíveis. A técnica usada para conseguir este objetivo é conhecida como *análise de discriminante*.

Belhumeur e Hespanha (Belhumeur, et al., 1997), além de Zhao coautores (Zhao, et al., 1998), usaram o Análises Linear de Discriminante⁴ para encontrar o subespaço \mathbf{W} que maximize a razão entre as matrizes de covariância inter-classe Σ_B e intra-classe Σ_W dadas pelas Equações (3-7) e (3-8) respectivamente, esta razão vem dada pela Equação (3-9).

$$\Sigma_W = \sum_{j=1}^M \sum_{i=1}^{M_j} (\mathbf{f}_{ij} - \boldsymbol{\mu}_j)(\mathbf{f}_{ij} - \boldsymbol{\mu}_j)^T \quad (3-7)$$

$$\Sigma_B = \sum_{j=1}^M M_j (\boldsymbol{\mu}_j - \boldsymbol{\mu})(\boldsymbol{\mu}_j - \boldsymbol{\mu})^T \quad (3-8)$$

onde:

- M , é o número de classes.
- M_j , é o número de amostras da classe j .
- \mathbf{f}_{ij} , é a amostra i da classe j (vector face de tamanho $h \times w$).
- $\boldsymbol{\mu}_j$, é o vetor de média para as amostras pertencentes à classe j .
- $\boldsymbol{\mu}$, é o vetor de média para todas as amostras.

$$\hat{\mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmax}} \left(\frac{|\mathbf{W}^T \Sigma_B \mathbf{W}|}{|\mathbf{W}^T \Sigma_W \mathbf{W}|} \right) \quad (3-9)$$

Estimar $\hat{\mathbf{W}}$ a partir da equação (3-9) conduz à solução de um problema de autovetores generalizado, ou seja:

⁴ Discriminante Linear de Fisher é uma técnica clássica em reconhecimento de padrões, desenvolvida inicialmente por Robert Fisher para classificação taxonômica (Fisher, 1936).

$$\Sigma_W^{-1} \Sigma_B \widehat{\mathbf{W}} = \boldsymbol{\gamma} \widehat{\mathbf{W}}, \quad (3-10)$$

onde, $\widehat{\mathbf{W}}$ é a matriz de autovetores de $\Sigma_W^{-1} \Sigma_B$, e cujas colunas \mathbf{w}_i são os vetores base que definem o subespaço \mathbf{W} e correspondem aos autovalores γ maiores do que zero.

Quando o número de dimensões $h \times w$ é maior do que o número de amostras disponíveis Σ_W resulta numa matriz singular, caso em que o computo de $\widehat{\mathbf{W}}$ a partir da Equação (3-10) não é possível. Para solucionar este problema, pode se realizar uma redução de dimensionalidade baseada na PCA.

Dada uma matriz \mathbf{F} de $N \times M_T$ dimensões, cujas colunas $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_{M_T}$ são as imagens faciais vetorizadas, M_T é o número de amostras e $N = hw$ o número de pixels total de cada imagem, os *fisherfaces* se calculam da seguinte forma:

1. Projetar \mathbf{F} no espaço e -dimensional, onde $e = (M_T - M)$ é o número de autovetores \mathbf{v}_e obtidos pelo PCA e M o número de classes (identidades).
Daqui para diante a projeção das amostras (faces) \mathbf{f}_i são representadas por \mathbf{y}_i .
2. Computar o vetor de média $\boldsymbol{\mu}$ para todas as amostras.

$$\boldsymbol{\mu} = \frac{1}{M_T} \sum_{i=1}^{M_T} \mathbf{y}_i \quad (3-11)$$

3. Calcular os vetores de média $\boldsymbol{\mu}_j$ para as amostras pertencentes a cada classe j .

$$\boldsymbol{\mu}_j = \frac{1}{M_j} \sum_{i=1}^{M_j} \mathbf{y}_{ij} \quad (3-12)$$

4. Calcular a matriz de covariância intra-classe Σ_W .

$$\Sigma_W = \sum_{j=1}^M \sum_{i=1}^{M_j} (\mathbf{y}_{ij} - \boldsymbol{\mu}_j)(\mathbf{y}_{ij} - \boldsymbol{\mu}_j)^T \quad (3-13)$$

5. Computar a matriz de covariância inter-classe Σ_B .

$$\Sigma_B = \sum_{j=1}^M M_j (\boldsymbol{\mu}_j - \boldsymbol{\mu})(\boldsymbol{\mu}_j - \boldsymbol{\mu})^T \quad (3-14)$$

6. Calcular os autovetores de $\Sigma_W^{-1} \Sigma_B$

$$\Sigma_W^{-1} \Sigma_B \widehat{\mathbf{W}} = \boldsymbol{\gamma} \widehat{\mathbf{W}} \quad (3-15)$$

e organizar a matriz de autovetores \mathbf{W} de forma a obter $\mathbf{W}_{lda} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_f]$, com $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_e \geq 0$, além $f = e - 1$.

7. Projetar cada \mathbf{f}_i de \mathbf{F} no novo espaço f -dimensional.

$$\mathbf{x}_i = \mathbf{V}_{pca}^T \mathbf{W}_{lda}^T (\mathbf{f}_i - \boldsymbol{\mu}) \quad (3-16)$$

Finalmente, uma identidade de \mathbf{F} é atribuída à imagem de teste \mathbf{f} se a distância euclidiana entre as projeções \mathbf{x} e \mathbf{x}_i , denotada pela Equação (3-6), é menor do que determinado limiar S_0 .

$$d^2(\mathbf{x}, \mathbf{x}_i) = \|\mathbf{x} - \mathbf{x}_i\|^2 < S_0 \quad (3-17)$$

3.3. Bayesian Intrapersonal/Extrapersonal Classifier

O algoritmo *Bayesian Intrapersonal/Extrapersonal Classifier* (bic) é um classificador binário que, ao contrario, das *autofaces* e *fisherfaces* usa uma medida de similaridade probabilística para discriminar entre classes. Tal medida está baseada numa estimativa paramétrica da densidade de probabilidade que caracteriza as variações típicas em aspecto das imagens faciais de um mesmo indivíduo.

Segundo Moghaddam e Pentland (Moghaddam, et al., 1998), a diferença das intensidades entre duas imagens faciais \mathbf{f}_1 e \mathbf{f}_2 permite capturar as variações particulares dos indivíduos, isto é, o vetor de atributos $\mathbf{x} = \mathbf{f}_1 - \mathbf{f}_2$. Estas variações são divididas em duas classes: “variações intrapessoais” (Ω_I) entre imagens dum mesmo indivíduo, originadas, por exemplo, por mudanças na iluminação do entorno de captura ou diferentes expressões faciais; e as “variações extrapessoais” (Ω_E) que resultam da comparação entre imagens faciais de diferentes indivíduos. A medida de similaridade $S(\mathbf{x})$ é definida pela probabilidade a posteriori de que \mathbf{x} pertença à classe Ω_I , conforme a regra de Bayes definida na Equação (3-18).

$$S(\mathbf{x}) = P(\Omega_I | \mathbf{x}) = \frac{P(\mathbf{x} | \Omega_I)P(\Omega_I)}{P(\mathbf{x} | \Omega_I)P(\Omega_I) + P(\mathbf{x} | \Omega_E)P(\Omega_E)}, \quad (3-18)$$

onde $P(\mathbf{x} | \Omega)$ são as funções de densidade de probabilidade, cujos parâmetros são estimados a partir dos dados disponíveis no banco de dados, e $P(\Omega)$ são as probabilidades a priori, cujos valores são estabelecidos conforme as características operacionais do banco de dados.

Como se vê, esta formulação Bayesiana converte o problema de reconhecimento facial num problema de classificação de padrões binário. Existe, contudo, um grande problema. “Geralmente não se dispõe de uma quantidade de amostras suficiente para estimar adequadamente os parâmetros das densidades $P(\mathbf{x} | \Omega)$ num espaço de dimensionalidade muito elevada”. Então, para resolver este inconveniente Moghaddam e Pentland (Moghaddam, et al.,

1997), propuseram usar o PCA a fim de dividir o espaço N -dimensional de \mathbf{x} ($\mathbf{x} \in \mathbb{R}^N$), em dois subespaços ortogonais, como é apresentado na Figura 4. O subespaço principal \mathbf{B} que contém os k primeiros componentes principais que servem para estimar os parâmetros de $P(\mathbf{x}|\Omega)$, e o subespaço $\bar{\mathbf{B}}$ que contém os $(N - k)$ componentes restantes. A componente de \mathbf{x} que está no espaço \mathbf{B} é denominado de “distância no espaço de atributos” (DIFS, *distance in feature space*) e a componente que fica no subespaço $\bar{\mathbf{B}}$ é denominado de “distância do espaço de atributos” (DFFS, *distance from feature space*).

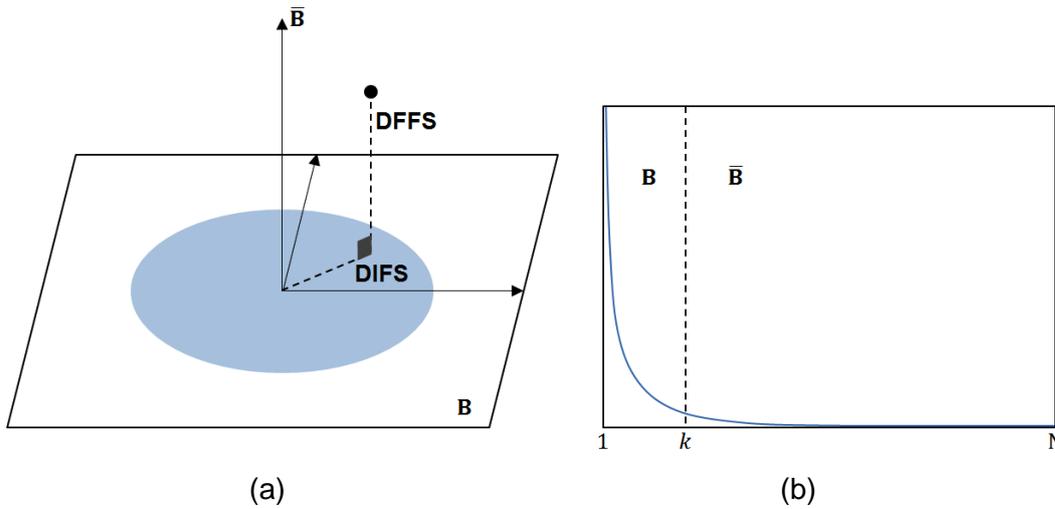


Figura 4. (a) Decomposição de \mathbb{R}^N no subespaço principal \mathbf{B} e seu complemento ortogonal $\bar{\mathbf{B}}$, (b) espectro de autovalores típico e a sua divisão nos dois subespaços ortogonais.

Baseados na decomposição do vetor de atributos \mathbf{x} nos subespaços \mathbf{B} e $\bar{\mathbf{B}}$, Moghaddam e Pentland (Moghaddam, et al., 1997) expressam as densidades de probabilidade $P(\mathbf{x}|\Omega)$ como um produto de duas densidades gaussianas independentes:

$$\hat{P}(\mathbf{x}|\Omega) = \left[\frac{\exp\left(-\frac{1}{2}\sum_{i=1}^k \frac{y_i^2}{\lambda_i}\right)}{(2\pi)^{k/2} \prod_{i=1}^k \lambda_i^{1/2}} \right] \left[\frac{\exp\left(-\frac{\epsilon^2(\mathbf{x})}{2\rho}\right)}{(2\pi\rho)^{(N-k)/2}} \right] = P_{\mathbf{B}}(\mathbf{x}|\Omega) \hat{P}_{\bar{\mathbf{B}}}(\mathbf{x}|\Omega) \quad (3-19)$$

onde, $P_{\mathbf{B}}(\mathbf{x}|\Omega)$ é a densidade marginal em \mathbf{B} , $\hat{P}_{\bar{\mathbf{B}}}(\mathbf{x}|\Omega)$ é a densidade marginal estimada em $\bar{\mathbf{B}}$, y_i são as componentes principais, $\epsilon^2(\mathbf{x})$ é o resíduo de reconstruir \mathbf{x} com os k primeiros autovetores, e o valor de ρ é simplesmente a média dos autovalores no espaço $\bar{\mathbf{B}}$, como é definido pela Equação (3-20):

$$\rho = \frac{1}{N-k} \sum_{i=k+1}^N \lambda_i \quad (3-20)$$

Neste algoritmo, o reconhecimento da identidade numa imagem facial se consegue pela seleção da medida de similaridade, entre \mathbf{f} e \mathbf{f}_i , com maior valor entre todas as comparações.

3.4. Local Gabor Binary Pattern Histogram Sequence

O *Local Gabor Binary Pattern Histogram Sequence* (lgbphs) é uma técnica de reconhecimento facial que combina os filtros de Gabor e a codificação binária do operador LBP (*Local Binary Pattern*) para caracterizar a textura da face e criar modelos robustos e invariantes às mudanças de iluminação e expressões faciais (Zhang, et al., 2005). A principal vantagem do lgbphs, é que permite representar uma imagem facial, como uma sequência de histogramas sem a necessidade de treinamento prévio, o que evita os problemas de generalização do qual sofrem as técnicas baseadas na análise em espaços discriminantes.

Para criar um modelo da face, isto é, o vetor de atributos \mathbf{x} , primeiro deve se realizar a convolução entre a imagem normalizada da face e um conjunto de filtros de Gabor. Os resultados de cada uma destas operações são denominados Imagens de Magnitude de Gabor (GMPs, *Gabor Magnitude Pictures*). Em seguida, se transforma cada imagem GMP em um mapa de intensidades codificado por médio do operador LBP, descrito sucintamente mais adiante. Tal mapa recebe o nome de mapa LGBP (*Local Gabor Binary Pattern*). No próximo passo, se dividem os mapas LGBP em pequenas regiões retangulares de tamanho fixo, as quais podem ou não estar sobrepostas. Finalmente, o modelo da face é formado pela concatenação dos histogramas de todas as regiões dos mapas LGBP.

As Imagens de Magnitude de Gabor (GMPs) são o resultado da convolução entre uma imagem facial normalizada $I_N(x,y)$ com cada um dos filtros de Gabor $\psi_{\mathbf{k}}(\mathbf{z})$, denotados pela Equação (3-21).

$$G_j(\mathbf{z}) = (I_N * \psi_{\mathbf{k}_j})(\mathbf{z}) \quad (3-21)$$

onde, $G_j(\mathbf{z})$ denota uma imagem GMP, $\mathbf{z} = (x,y)$. Os filtros de Gabor são definidos como:

$$\psi_{\mathbf{k}}(\mathbf{z}) = \frac{\mathbf{k}^2}{\sigma^2} \exp\left(-\frac{\mathbf{k}^2 \mathbf{z}^2}{2\sigma^2}\right) \left[\exp(i\mathbf{k}^T \mathbf{z}) - \exp\left(-\frac{\sigma^2}{2}\right) \right] \quad (3-22)$$

Comumente, uma família de filtros de Gabor é gerada pela definição de 40 diferentes vectores de parâmetros \mathbf{k}_j , com cinco escalas $\mu = 0, \dots, 4$ e oito orientações $\nu = 0, \dots, 7$, assim:

$$\mathbf{k}_j = \mathbf{k}_{\mu,\nu} = \begin{pmatrix} k_\mu \cos \varphi_\nu \\ k_\mu \sin \varphi_\nu \end{pmatrix}; k_\mu = 2^{-\frac{2+\mu}{2}} \pi; \varphi_\nu = \frac{\nu\pi}{8}; j = 8\mu + \nu \quad (3-23)$$

Devido a que os filtros de Gabor contém componentes complexos, neste algoritmo, as imagens GMP são criadas considerando só a magnitude de $G_j(\mathbf{z})$.

Os mapas LGBP são gerados a partir da aplicação do operador LBP sobre as imagens GMP. Este operador foi desenvolvido por Ojala e coautores (Ojala, et al., 2002), e codifica uma imagem qualquer a níveis de cinza pela comparação de um pixel central (p_c) com seus oito pixels vizinhos (p_n); esta operação vem definida pela equação (3-24), onde, $LBP(p_c)$ faz referência ao nível de cinza obtido pela aplicação do operador binário $S(p_n - p_c)$ num pixel da imagem GMP.

$$LBP(p_c) = \sum_{n=0}^7 S(p_n - p_c) 2^n \quad (3-24)$$

$$S(p_n - p_c) = \begin{cases} 1, & p_n \geq p_c \\ 0, & p_n < p_c \end{cases} \quad (3-25)$$

O vetor de atributos é formado pela concatenação dos histogramas $\mathbf{H}_{j,r} = [h_0, h_1, \dots, h_{2^n-1}]$ extraídos de cada região ($r: R_0, R_1, \dots, R_{r-1}$) que compõe cada mapa LGBP (j), formalmente:

$$\mathbf{x} = [\mathbf{H}_{0,0}, \mathbf{H}_{0,1}, \dots, \mathbf{H}_{0,r-1}, \mathbf{H}_{1,0}, \mathbf{H}_{1,1}, \dots, \mathbf{H}_{1,r-1}, \dots, \mathbf{H}_{j,0}, \mathbf{H}_{j,1}, \dots, \mathbf{H}_{j,r-1}] \quad (3-26)$$

O reconhecimento é dado pela comparação entre os modelos \mathbf{x}_i e \mathbf{x} correspondentes às imagens faciais normalizadas que compõem o banco de dados e de teste. Esta comparação gera uma medida de similaridade $S(\mathbf{x}, \mathbf{x}_i)$ que pode ser obtida a partir de: (1) o cálculo da área comum a ambos dos modelos, (2) à distância qui-quadrado entre \mathbf{x} e \mathbf{x}_i , (3) soma ponderada dos histogramas gerados para cada região dos mapas LGBP com base nas variações intrapessoais e extrapessoais dos indivíduos (Zhang, et al., 2005). Assim, em qualquer dos três casos, a identidade do modelo cujo escore $S(\mathbf{x}, \mathbf{x}_i)$ tem o maior valor entre todas as comparações é a selecionada.

3.5. Graph Matching

O *Graph Matching* é uma técnica baseada nos *Gabor Graphs* que determina a similaridade entre duas imagens faciais normalizadas (referência e

teste) mediante a comparação de suas respostas aos filtros de Gabor em pontos correspondentes e específicos em ambas as imagens. As posições destes pontos são definidas pela distribuição dos nós de uma malha que é sobreposta a cada uma das imagens, e cuja localização corresponde ao pareamento dos centros dos olhos com dois de seus nós. Um *Gabor Graph* é definido como o conjunto de *Gabor Jets* correspondentes a cada nó da malha; e um *Gabor Jet* é o conjunto de respostas aos filtros de Gabor num nó da malha.

Uma vantagem dos *Gabor Graphs*, e na qual se fundamenta a ideia de usar os filtros de Gabor para reconhecimento facial, é que, dependendo da resolução da malha, pode-se reconstruir uma imagem facial com um grande nível de acurácia, já que a informação da textura está codificada nas componentes de fase dos *Gabor Jets*, mais do que na sua componente magnitude. Desta forma, a identificação de uma imagem facial normalizada de teste é realizada pela comparação de seu modelo *Gabor Graph* com os modelos de cada imagem pertencente à galeria. Esta comparação envolve o cálculo de uma medida de similaridade entre os *Gabor Jet* nos pontos correspondentes ao mesmo nó da malha nas duas imagens.

No entanto, devido às variações existentes nas imagens faciais (rotação da cabeça, expressões faciais) ou erros no processo de normalização das imagens, às vezes os pontos das imagens correspondentes a um mesmo nó se encontram deslocados. Günther e coautores (Günther, et al., 2012) propõem uma solução para este inconveniente por meio da correção da diferença de fases dos *Gabor jets*, realizada diretamente no processo de reconhecimento.

Assim, a medida de similaridade para cada par de *Gabor Jets* $S(J, J')$ vem dada por:

$$S(J, J') = 1 - \frac{1}{N_g} \sum_{j=0}^{N_g-1} \frac{|a_j - a'_j|}{a_j + a'_j} + \frac{1}{N_g} \sum_{j=0}^{N_g-1} \cos(\phi_j - \phi'_j - \mathbf{k}_j^T \mathbf{d}_0) \quad (3-27)$$

onde,

- J e J' são os *Gabor Jet* dum nó nas imagens de referência e teste respectivamente.
- Em coordenadas polares, $J = a_j e^{i\phi_j}$, com fase ϕ_j e amplitudes a_j normalizadas pelo número de filtros de Gabor N_g .
- \mathbf{k}_j é o vetor de parâmetros dos filtros de Gabor, definidos pela Equação (3-23).

A correção do deslocamento entre J e J' é realizada pelo “vetor de disparidade” $\mathbf{d}_\mu = [d_x \ d_y]^T = \Gamma_\mu^{-1} \Phi_\mu$, onde Γ_μ e Φ_μ estão definidas pelas Equações (3-27) e (3-28), respectivamente:

$$\Gamma_{\mu;x,y} = \sum_{j=8\mu}^{N_g-1} a_j a'_j k_{j;x} k_{j;y} \quad (3-28)$$

$$\Phi_{\mu;x} = \sum_{j=8\mu}^{N_g-1} a_j a'_j k_{j;x} (\phi_j - \phi'_j - n_j 2\pi) \quad (3-29)$$

Aqui, os valores de n_j para as escalas $\mu = 3, \dots, 0$ nos filtros de Gabor, são definidas como:

$$n_j = \left\lfloor \frac{\phi_j - \phi'_j - \mathbf{k}_j^T \mathbf{d}_{\mu+1}}{2\pi} \right\rfloor : \forall j \in \{8\mu, \dots, 8(\mu + 1) - 1\} \quad (3-30)$$

Finalmente, a similaridade entre os dois *Gabor Graphs* (correspondentes às imagens de referência e teste) é calculada pela similaridade média dos *Gabor Jets* em todas as posições da malha.

3.6. Gaussian Mixture Model

O algoritmo *Gaussian Mixture Model* (GMM) foi inicialmente empregado no reconhecimento de voz (Reynolds, et al., 2000). Devido à sua grande eficiência para esta tarefa, pensou-se em adaptá-lo para verificação de identidade a partir de imagens de faces. McCool e Marcel (McCool, et al., 2009) propuseram combinar vários classificadores baseados em misturas de gaussianas para verificar faces. Este método modela a distribuição dos vetores de atributos que contêm informação espacial e de frequência das imagens faciais por meio de GMMs. A modelagem se faz pela decomposição das imagens num conjunto de blocos e pela coleção de suas respostas à transformada discreta de cosseno (DCT, *Discret Cosine Transform*).

Esta aproximação corresponde aos métodos de reconhecimento baseados em componentes (vide Seção 2.2), que do ponto de vista prático têm a vantagem de serem invariantes a iluminação, rotação de cabeça, expressões faciais e oclusão. Além disso, permite a geração de uma quantidade considerável de vetores de atributos a partir duma única imagem facial, favorecendo assim uma boa estimativa dos parâmetros das Gaussianas.

De uma forma geral, a estimativa dos parâmetros que modela a distribuição dos vetores de atributos de um indivíduo (imagem facial), se obtém através da adaptação dos parâmetros de um GMM base, conhecido como

modelo UBM (*Universal Background Model*). O UBM é essencialmente um modelo geral do universo de imagens faciais (Ω_{mundo}). O modelo de cada cliente ($\Omega_{\text{cliente}}^i$) é gerado por uma adaptação do UBM com base em um conjunto de vetores de atributos extraídos de uma ou mais faces do cliente. Para verificar uma imagem facial, avalia-se a verossimilhança do vetor de atributos \mathbf{x} relativamente a ambos dos modelos (Ω_{mundo} e $\Omega_{\text{cliente}}^i$), as quais são combinadas para produzir um escore $h(\mathbf{x})$ dado pela Equação 3-31. A imagem é verificada se e só se o valor de $h(\mathbf{x})$ é maior do que algum determinado limiar τ .

$$h(\mathbf{x}) = \ln(p(\mathbf{x}|\Omega_{\text{cliente}}^i)) - \ln(p(\mathbf{x}|\Omega_{\text{mundo}})) \quad (3-31)$$

No estudo realizado por McCool e Marcel os vetores de atributos de uma imagem facial normalizada são extraídos das sub-imagens de frequência geradas a partir da coleção dos coeficientes da DCT dos blocos que compõem a imagem facial original, estes vetores de atributos são obtidos da seguinte forma:

1. a imagem facial normalizada é dividida em pequenos blocos de $h_b \times w_b$ pixels, os quais podem estar (ou não) sobrepostos;
2. a DCT é calculada de cada um destes blocos, de cujo resultado se retém os D coeficientes (tipicamente $D = 64$) relativos às mais baixas frequências;
3. cada sub-imagem de frequência é gerada pela coleção dos coeficientes da DCT relativos ao mesma ordem ao longo dos blocos;
4. as filas, ou colunas, ou blocos vetorizados de cada sub-imagem de frequência constituem a descrição de uma imagem facial.

Finalmente, McCool e Marcel usam a abordagem apresentada por (Cardinaux, et al., 2003) para a adaptação do UBM para cada cliente.

3.7. Inter Session-Variability Modeling

O *Inter Session-Variability* (ISV) é uma extensão da técnica GMM Visa melhorar a estimativa dos modelos dos clientes por meio da eliminação das variações existentes nas imagens de um indivíduo em distintas sessões. Basicamente, o ISV trata-se explicitamente as variações entre imagens faciais de um mesmo indivíduo.

Assim, o propósito do ISV é modelar explicitamente as variações que ocorrem devido à iluminação, pose, expressão facial, e excluí-las dos modelos criados durante o processo de cadastramento e teste de uma nova imagem

facial. Em outras palavras, o objetivo do ISV é encontrar o modelo mais fiel dos indivíduos que seja tanto quanto possível independente das condições de aquisição das imagens.

Nesta técnica, a diferença da GMM, os vetores de atributos são extraídos diretamente da DCT como a seguir:

1. a imagem é dividida em pequenos blocos de $h_b \times w_b$ pixels, os quais podem estar (ou não) sobrepostos;
2. a DCT é calculada de cada um destes blocos, de cujo resultado se retêm os D coeficientes relativos as mais baixas frequências;
3. o conjunto de vetores D dimensionais assim produzidos constituem a descrição de uma imagem facial.

Mais informação sobre esta técnica pode ser encontrada no trabalho realizado por Wallace e coautores (Wallace, et al., 2011).

3.8. Probabilistic Linear Discriminant Analysis

Esta técnica foi proposta por Prince e Elder (Prince, et al., 2007), e recebe o nome de Análises Discriminante Linear Probabilístico (PLDA, *Probabilistic Linear Discriminant Analysis*) por ser uma extensão probabilística dos *Fisherfaces*. Nesta técnica uma imagem é descrita como o resultado de um modelo generativo que trata separadamente as variações entre imagens de um mesmo indivíduo e imagens de diferentes indivíduos.

O modelo generativo proposto por Prince e Elder (ver Equação 3-32) tenta explicar as variações da face. Neste modelo o vetor \mathbf{h}_i define a identidade de um indivíduo univocamente. Uma segunda componente, denotada aqui pelo vetor \mathbf{w}_{ij} , descreve todas as possíveis variações que a imagem facial de um mesmo indivíduo pode assumir. Tanto \mathbf{h}_i quanto \mathbf{w}_{ij} são variáveis ocultas (latentes) que, como tais, não podem ser diretamente medidas. A imagem facial, que corresponde à variável observável, é modelada pela expressão:

$$\mathbf{x}_{ij} = \boldsymbol{\mu} + \mathbf{F}\mathbf{h}_i + \mathbf{G}\mathbf{w}_{ij} + \epsilon_{ij} \quad (3-32)$$

onde $\boldsymbol{\mu}$ representa a imagem facial média do universo de indivíduos, \mathbf{F} é uma matriz cujas colunas são os vetores de base do subespaço em que estão modeladas as variações das imagens de diferentes indivíduos, e \mathbf{h}_i representa uma posição neste subespaço. A matriz \mathbf{G} contém vetores de base do subespaço que modela as variações das imagens faciais de um mesmo

indivíduo, \mathbf{w}_{ij} representa um ponto neste subespaço, e ϵ_{ij} explica o erro do modelo (variações inexplicáveis a partir do modelo). Admite-se que ϵ_{ij} tem uma distribuição normal com matriz de covariância Σ diagonal. Os parâmetros $\theta = \{\boldsymbol{\mu}, \mathbf{F}, \mathbf{G}, \Sigma\}$ são estimados a partir de um conjunto de dados de treinamento por meio do algoritmo EM.

Neste modelo, duas imagens faciais que pertencem ao mesmo indivíduo compartilham a mesma variável de identidade \mathbf{h}_i , do contrário, terão diferentes identidades.

Um inconveniente desta abordagem é que não pode ser aplicado em bancos de dados muito grandes, já que envolve a inversão de uma matriz, cujo tamanho cresce de forma quadrática com o número de observações utilizadas no treinamento (Kenny, 2010). Recentemente, El Shafey e coautores. (El Shafey, et al., 2013), propuseram um método mais eficiente computacionalmente na implementação do PLDA. Basicamente, esta melhora consiste numa troca adequada de variável que permite diagonalizar as matrizes, cuja inversa importa calcular. Com isto, a determinação das probabilidades a posteriori passa a ser escalável.