

Extração e Comparação de Características Locais e Globais para o Reconhecimento Automático de Imagens de Faces

Vagner do Amaral e Carlos Eduardo Thomaz
Centro Universitário da FEI
Av. Humberto de Alencar Castelo Branco, 3972 - Assunção
São Bernardo do Campo, São Paulo, Brasil
vamaral@fei.edu.br e cet@fei.edu.br

Resumo

Recent studies on face recognition have highlighted that local and global features are both useful depending on the application context. This paper proposes and implements a comparison between the Local Binary Pattern (LBP) and Principal Components Analysis (PCA) techniques in order to understand the behavior of each feature extraction method when using pre-processed and previously spatially normalized face images. Our results show a good performance of LBP only in the experimental situations where the training set is composed of individuals in frontal pose with necessarily the neutral facial expression available. However, using a training set with distinct facial expressions randomly chosen the PCA has been more efficient in classifying the top ranked individuals.

1. Introdução

Há muitos anos os desafios presentes na área de reconhecimento de faces fomentam o surgimento de conferências, publicações e protocolos específicos, destinados a padronização de métricas para avaliação de diversas técnicas de classificação de padrões.

Segundo Zhao et al. [19], existem pelo menos duas razões para esta tendência. A primeira é a disponibilidade de tecnologias viáveis após 40 anos de pesquisa, e a segunda é a vasta gama de aplicações. Com o surgimento do protocolo FERET [13] os estudos sobre reconhecimento facial puderam empregar um modelo de avaliação padronizado, com o propósito de facilitar a comparação entre abordagens distintas [4, 18, 6, 12, 1, 2, 15, 7]. Entretanto, aplicações reais de reconhecimento automático de faces apresentam características particulares, ausentes no modelo FERET, que serão abordados no escopo deste trabalho.

Tem-se por objetivo implementar a comparação entre métodos de extração e comparação de características faciais

locais e globais, predominantes na literatura afim. Mais especificamente, pretende-se avaliar as técnicas *Local Binary Pattern* (LBP) e *Principal Components Analysis* (PCA) visando analisar as características faciais de maior relevância no processo de reconhecimento e indicar a melhor solução para o problema de restrição amostral.

A estrutura deste artigo é composta por 5 seções. Na seção seguinte, descrevem-se as técnicas de extração de características empregadas neste estudo, sob o contexto do problema de reconhecimento facial. Na seção 3 são apresentadas as métricas utilizadas para classificação das amostras. Os experimentos e resultados obtidos são discutidos na seção 4. Por fim, na última seção, conclui-se este trabalho.

2. Extração de características

Os métodos de extração empregados em reconhecimento facial costumam atender situações específicas e podem ser classificados em três categorias: locais, globais ou híbridos [9, 10, 15], conforme segue:

- Métodos locais: Estes métodos possuem duas abordagens principais. A primeira utiliza características geométricas, como distância e forma, para calcular medidas relativas entre componentes faciais particulares. No entanto, os processos automáticos para demarcação das características são complexos e imprecisos, tornando-os dependentes de procedimentos manuais. A segunda abordagem baseia-se na aparência de elementos independentes, ou seja, utiliza informações globais de regiões específicas da face, como, por exemplo, os olhos, nariz e boca. Permite-se, deste modo, explorar os fatores de relevância cognitiva de cada elemento facial [5, 15].
- Métodos globais: Métodos globais, também conhecidos como holísticos, dominam este campo de estudo desde 1990 [15] e representam cada amostra como uma matriz bidimensional [5]. Esta estrutura permite

o desenvolvimento de sistemas autônomos utilizando técnicas de aprendizagem. No entanto, estes métodos são altamente dependentes da quantidade e relevância das amostras de treinamento e exigem uma capacidade computacional maior para serem implementados, devido ao padrão de armazenamento dos dados.

- Métodos híbridos: Tais métodos empregam informações locais e globais. Nesta abordagem, visa-se preservar as vantagens e reduzir as desvantagens de ambos. Entretanto, este processo de otimização requer uma análise precisa sobre quais características deverão ser combinadas [15].

2.1. Análise de Componentes Principais (PCA)

O excesso de informação redundante permite que cada imagem de face seja globalmente representada pela sua variância em relação a média [14, 16, 17]. Dada a matriz de imagens de face X de dimensão $N \times n$, onde N é o total de imagens e n a quantidade total de pixels de cada imagem, obtém-se o vetor média através da seguinte equação:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i. \quad (1)$$

O desvio global de cada face é $z_i = x_i - \bar{x}$, onde $i = 1, 2, \dots, N$. Se então aplicarmos esse cálculo a todas as faces da matriz X , teremos uma matriz Z que contém apenas as variações de cada face em torno da média \bar{x} , conforme

$$Z = [z_1, z_2, \dots, z_N]^T. \quad (2)$$

As componentes principais podem ser descritas pelos autovetores e autovalores da matriz de covariância de Z . No entanto, é necessário um esforço computacional enorme para calcular a matriz de covariância $S = Z^T Z$, pois em geral $n \gg N$. Contudo, é possível calcular as componentes principais com esforço muito menor, a partir de uma matriz $N \times N$, por meio da equação:

$$S^* = \left(\frac{Z \cdot Z^T}{N - 1} \right). \quad (3)$$

Entretanto, os autovetores Φ^* da matriz S^* não são ortonormais. Pode-se recalculá-los, conforme a equação abaixo [8]:

$$\Phi = \frac{Z \cdot \Phi^* \cdot \Lambda^{*-1/2}}{\sqrt{N - 1}}, \quad (4)$$

onde Λ^* são os autovalores da matriz S^* e Φ é a matriz de transformação, que descreve as faces do conjunto de treinamento nesse novo subespaço de dimensão menor, cujos autovetores são ortogonais e ortonormais.

Tendo em vista a comparação e classificação das imagens em um processo de reconhecimento de padrões, as amostras de teste também devem ser projetadas neste mesmo subespaço, conforme a seguinte equação:

$$x_{pca} = (x_{teste} - \bar{x})^T \cdot \Phi, \quad (5)$$

onde x_{teste} é uma amostra de teste e x_{pca} é a representação dessa amostra no subespaço das componentes principais.

2.2. Padrão Binário Local (LBP)

Imagens de faces podem ser vistas como uma composição de micro-padrões que são bem descritos pelo LBP [2]. Neste método cada pixel de uma imagem é substituído por um valor binário. Este valor é determinado pela comparação de uma matriz quadrada contendo os pixels vizinhos, onde cada vizinho é comparado com o valor central, conforme a seguinte condição:

$$b_{ij} = \begin{cases} 0, & v_{ij} < v_c \\ 1, & v_{ij} \geq v_c \end{cases}, \quad (6)$$

onde v_{ij} é o valor de um pixel na posição (i, j) e v_c é o valor central.

Os valores obtidos para cada vizinho são concatenados e o número binário gerado é convertido na base decimal para substituir o valor central v_c . A Figura 1 exemplifica este processo, para uma matriz 3×3 de pixels vizinhos. Contudo, o tamanho e o formato da vizinhança podem variar.

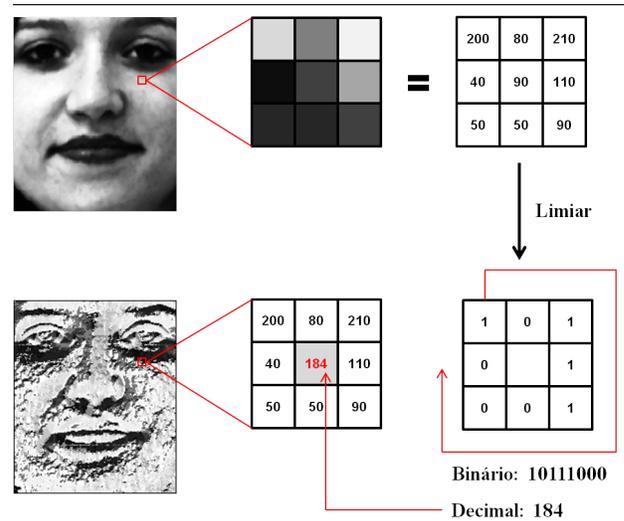


Figura 1. Operação básica do LBP: valor binário resultante 10111000 e decimal 184.

A imagem gerada é dividida em regiões, habitualmente organizadas em grades. Entretanto, é possível empregar outras formas com, ou sem, sobreposição. O descritores de textura são extraídos de cada região isoladamente, calculando-se o histograma de intensidade dos pixels. Os vetores resultantes são concatenados em uma única matriz para classificação das amostras.

3. Classificação das Amostras

Em amostras processadas com o PCA habitualmente emprega-se um classificador de vizinho mais próximo. No presente trabalho, adotaram-se as distâncias Euclidiana, Mahalanobis convencional, Mahalanobis ponderada e Erro Médio Quadrático modificado, descritas a seguir, considerando que $m \leq n$, x é a imagem de teste e y uma imagem da base de treinamento, ambas projetadas no subespaço gerado pelo PCA:

- Euclidiana

$$d(x, y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2}; \quad (7)$$

- Mahalanobis

$$d(x, y) = (x - y)S^{-1}(x - y)^T; \quad (8)$$

- Mahalanobis ponderada

$$d(x, y) = - \sum_{i=1}^m x_i y_i \sqrt{1/\lambda_i}; \quad (9)$$

- Erro médio quadrático modificado

$$d(x, y) = \frac{\sum_{i=1}^m \sqrt{1/\lambda_i} (x_i - y_i)^2}{\sqrt{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2}}. \quad (10)$$

Para classificar as amostras geradas pelo LBP utilizam-se, basicamente, métricas que permitam explorar a relevância de algumas características faciais sobre outras, no processo de reconhecimento [19]. Deste modo, é possível definir pesos específicos para cada região da grade de histogramas. Portanto, adotou-se neste trabalho a distância Qui-Quadrado, também empregada por Ahonen et. al. [1, 2], definida abaixo:

$$\chi_w^2(x, y) = \sum_{i,j} w_j \frac{(x_{i,j} - y_{i,j})^2}{x_{i,j} + y_{i,j}}, \quad (11)$$

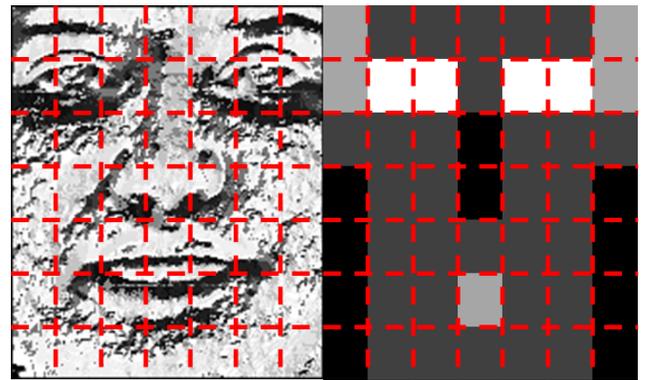
onde x e y são as matrizes de histogramas a serem comparadas, i é um elemento do histograma correspondente a j -ésima região e w_j é o peso pré-definido para a região j .

4. Experimentos e Resultados

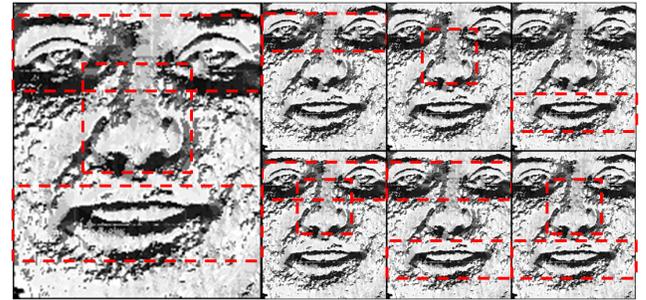
Para este estudo foram utilizadas 400 imagens de face da base FERET e 900 imagens de face da base FEI [11], todas frontais com pequenos desvios de perfil e diferentes expressões faciais, normalizadas de acordo com a metodologia proposta por Amaral et al.[3]. Para ambas as bases, adotaram-se duas configurações distintas, denominadas homogênea e heterogênea, conforme segue:

- Homogênea: emprega o protocolo FERET. O conjunto de treinamento é gerado apenas com imagens frontais, em expressão neutra, e os testes são realizados com as demais amostras presentes em cada base;
- Heterogênea: utiliza validação cruzada *leave-one-out*. O conjunto de treinamento é composto por todas amostras de cada base, e suas respectivas variações, com exceção da amostra de teste determinada em cada rodada.

Nos experimentos com LBP, foram empregados, para a extração dos histogramas, os padrões 7×7 , 7×7 ponderado e mais sete composições de regiões faciais específicas, conforme ilustrado nas Figuras 2(a) e 2(b).



(a) Grades 7×7 e 7×7 ponderada. As regiões preta, cinza escuro, cinza claro e branco, com pesos 0, 1, 2 e 4, respectivamente.



(b) Composições de características faciais específicas.

Figura 2. Regiões empregadas no LBP.

Os resultados descritos na Figura 3 ilustram a taxa de classificação, entre os 10 indivíduos mais próximos, para as grades e regiões faciais específicas empregando-se os olhos, o nariz e a boca, independentemente ou em composição. Verificou-se um bom desempenho em relação à grade ponderada, principalmente em ambientes heterogêneos, devido à valorização de regiões mais discriminantes e a suavização de desvios amostrais. Contudo, os resultados obtidos no ambiente heterogêneo são muito baixos e ficam aquém das taxas geradas no ambiente homogêneo e apresentadas na literatura relacionada. Os ensaios acerca das regiões faciais específicas são favoráveis ao uso exclusivo dos olhos e do nariz como regiões mais discriminantes.

Para comparar os métodos LBP e PCA, gerou-se uma taxa de classificação cumulativa, considerando o reconhecimento de uma amostra de teste ao longo das cinquenta amostras de treinamento mais semelhantes [13, 1, 2]. Foram avaliadas as grades 7×7 e 7×7 ponderada para a técnica LBP e as métricas Mahalanobis e Mahalanobis ponderada, com todas componentes principais para o PCA, pois estas foram as configurações que apresentaram os melhores resultados correspondentemente. Os resultados dessa comparação são descritos nas Figuras 4(a), 4(b), 4(c) e 4(c). Em geral, o desempenho do PCA foi melhor que o LBP. Pode-se observar que em ambientes homogêneos os resultados de ambas técnicas foram semelhantes. No entanto, para ambientes heterogêneos o PCA apresentou taxas de reconhecimento superiores com as duas métricas propostas para esta análise.

5. Conclusão

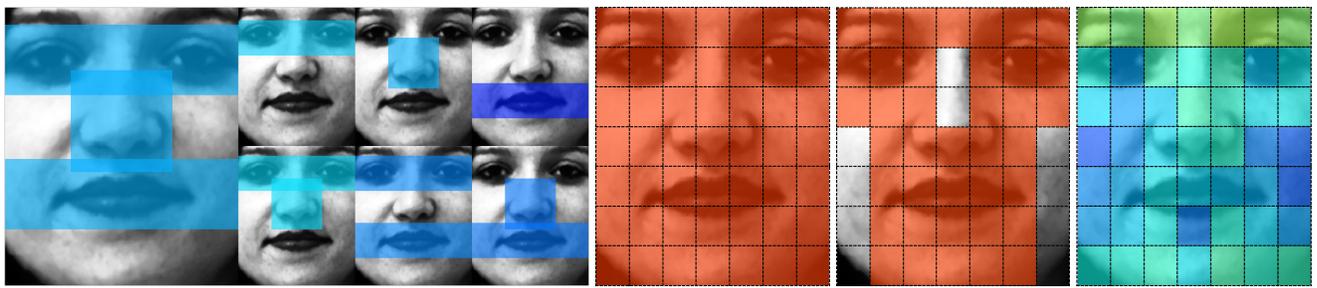
Ambientes homogêneos, compostos por imagens frontais neutras, são ideais para a implementação de sistemas de reconhecimento automático de faces e esta abordagem é apropriada para situações onde há restrição amostral. No entanto, aplicações reais frequentemente proporcionam a criação de um ambiente heterogêneo, pois não há critério para aquisição das imagens de treinamento. Enfim, a utilização de múltiplas imagens por pessoa otimiza a classificação caso todos indivíduos possuam as mesmas variações.

No presente trabalho, verificou-se, que para imagens pré-processadas, utilizando todas as componentes principais disponíveis, o PCA foi melhor que o LBP na maioria dos experimentos realizados. Contudo, o LBP é mais adequado para tratar variações de captura e problemas de oclusão parcial da face, pois permite a valorização de elementos mais discriminantes e a suavização de áreas ruidosas ao processo, melhorando as taxas de reconhecimento. Este método não requer treinamento e também é invariante ao problema de limitação amostral. No entanto, observou-se uma fragilidade desta técnica a variações na base de treinamento. Portanto, recomenda-se o uso de bases homogêneas ou pe-

sos que neutralizem essas variações para aplicação do LBP em reconhecimento de faces.

Referências

- [1] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. *Computer Vision ECCV 2004*, 3021:469–481, 2004.
- [2] T. AHONEN, A. HADID, and M. PIETIKAINEN. Face description with local binary patterns: Application to face recognition. *IEEE Transaction Pattern Analysis and Machine Intelligence*, 28:2037–2041, 2006.
- [3] V. AMARAL, C. FIGARO-GARCIA, G. J. F. GATTÁS, and C. E. THOMAZ. Normalização espacial de imagens frontais de face em ambientes controlados e não-controlados. *FaSci-Tech*, 1(1), 10 2009.
- [4] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski. Face recognition by independent component analysis. *IEEE Transactions on Neural Networks*, 13(6):1450–1464, 2002.
- [5] R. BRUNELLI and T. POGGIO. Face recognition: Features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):805–821, 10 1993.
- [6] K. Delac, M. Grgic, and S. Grgic. Independent comparative study of pca, ica, and lda on the feret data set. *International Journal of Imaging Systems and Technology*, 15(5):252–260, 2005.
- [7] W. DENG, J. HU, J. GUO, W. CAI, and D. FENG. Robust, accurate and efficient face recognition from a single training image: A uniform pursuit approach. *Pattern Recognition*, 43(5):1748–1762, 2010.
- [8] K. FUKUNAGA. *Introduction to statistical pattern recognition (2nd ed.)*. Academic Press Professional, Inc., San Diego, CA, USA, 1990.
- [9] S. GUTTA and H. WECHSLER. Face recognition using hybrid classifiers. *Pattern Recognition*, 30(4):539–553, 7 1997.
- [10] H. F. LIAU, K. P. SENG, L. ANG, and S. W. CHIN. New parallel models for face recognition. In K. DELAC, M. GRGIC, and M. S. BARTLETT, editors, *Recent Advances in Face Recognition*, volume 1, pages 15–28. In-teh, 2008.
- [11] L. L. OLIVEIRA JUNIOR and C. E. THOMAZ. Captura e alinhamento de imagens: Um banco de faces brasileiro. Technical report, Centro Universitário da FEI, 6 2006.
- [12] V. PERLIBAKAS. Distance measures for pca-based face recognition. *Pattern Recognition*, 25(6):711–724, 2004.
- [13] P. J. Phillips, H. Moon, S. A. Rizvi, and P. Rauss. The feret evaluation methodology for face-recognition algorithms. In *IEEE Transaction on Pattern Analysis and Machine Intelligence*, volume 22, pages 1090–1104, Washington, DC, USA, 10 2000. IEEE Computer Society.
- [14] L. SIROVICH and M. KIRBY. Low-dimensional procedure for the characterization of human faces. *Journal of Optical Society of America A*, 4(3):519–524, 1987.
- [15] X. TAN, S. CHEN, Z. ZHOU, and F. ZHANG. Face recognition from a single image per person: A survey. *Pattern Recognition*, 39(9):1725–1745, 2006.
- [16] M. TURK and A. PENTLAND. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

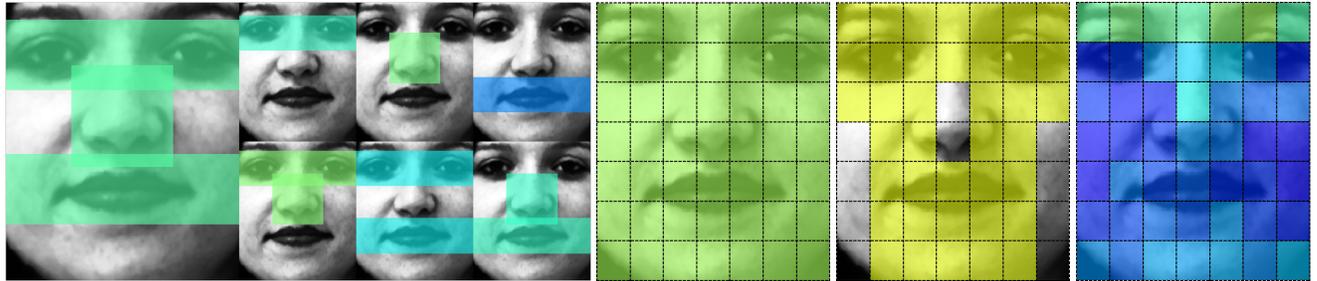


(a) Regiões faciais de interesse.

(b) Grade 7x7.

(c) Grade 7x7 ponderada.

(d) Acertos por região.

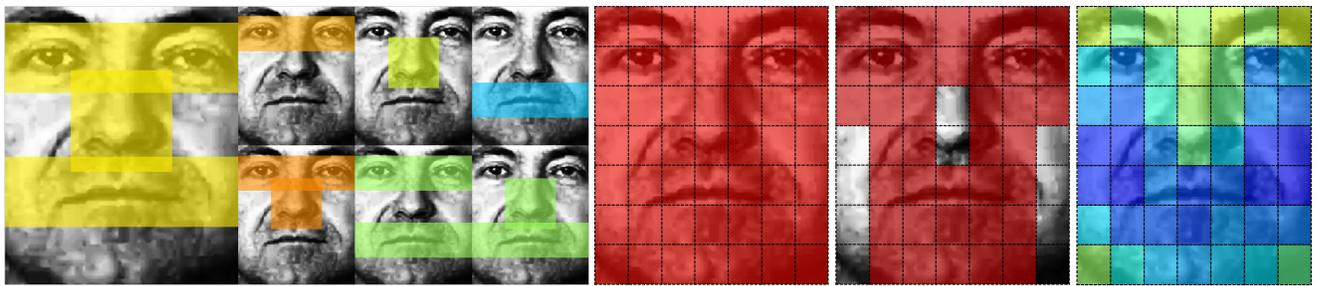


(e) Regiões faciais de interesse.

(f) Grade 7x7.

(g) Grade 7x7 ponderada.

(h) Acertos por região.

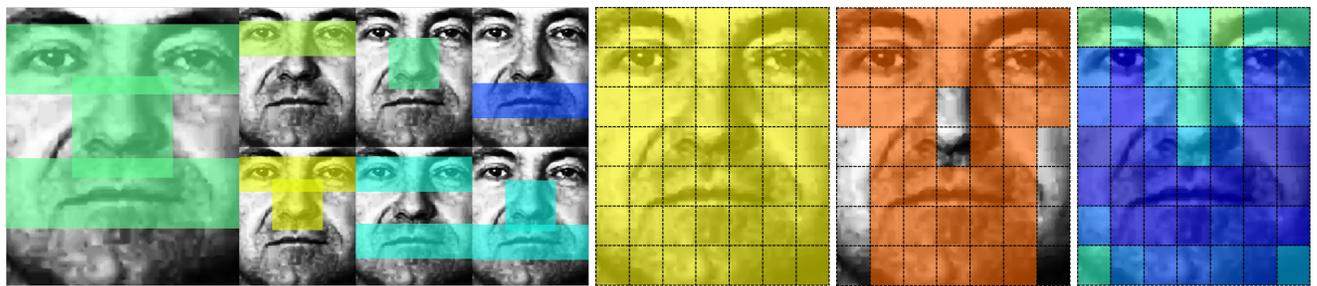


(i) Regiões faciais de interesse.

(j) Grade 7x7.

(k) Grade 7x7 ponderada.

(l) Acertos por região.



(m) Regiões faciais de interesse.

(n) Grade 7x7.

(o) Grade 7x7 ponderada.

(p) Acertos por região.

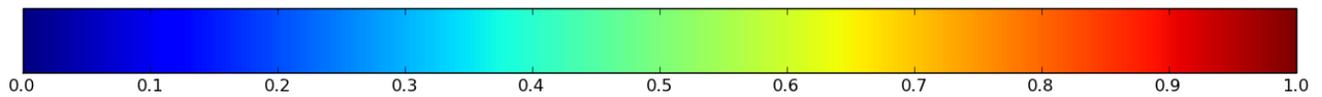
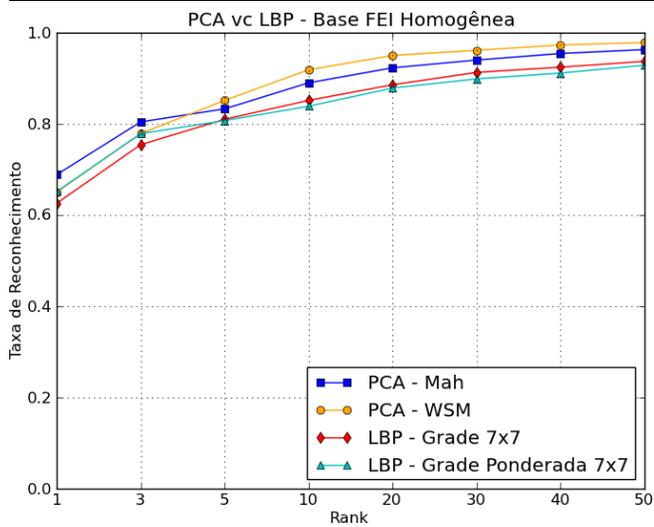
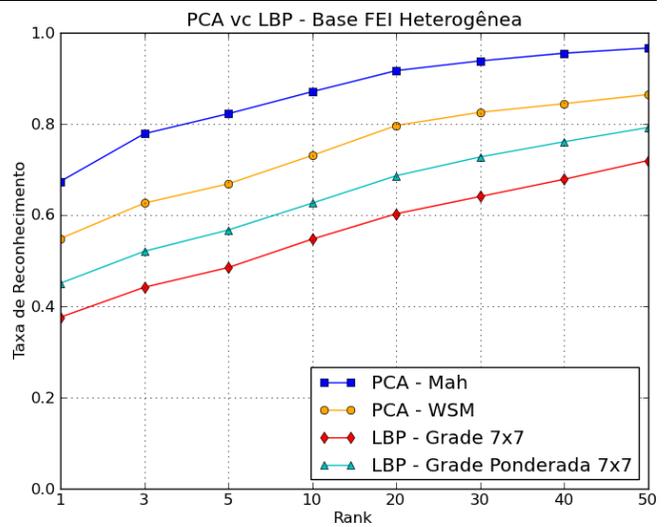


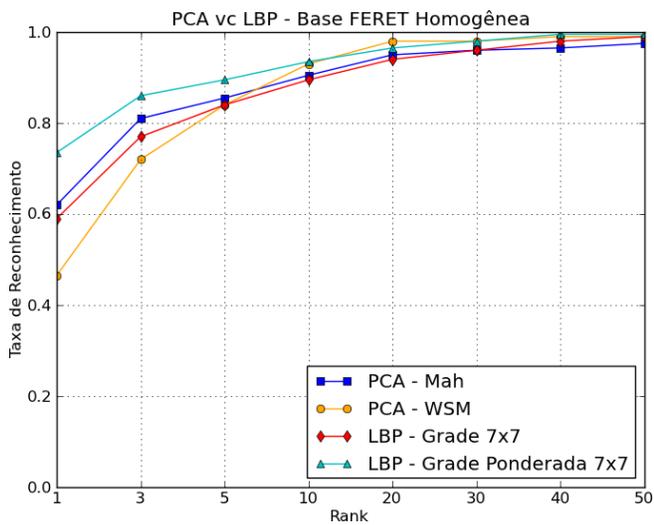
Figura 3. Experimentos realizados, de cima para baixo, bases FEI homogênea, FEI heterogênea, FERET homogênea e FERET heterogênea. Da esquerda para a direita, regiões faciais de interesse, grade 7x7 simples, grade 7x7 ponderada e grade 7x7 independente. A escala de cores indica a taxa de classificação entre as 10 amostras mais semelhantes.



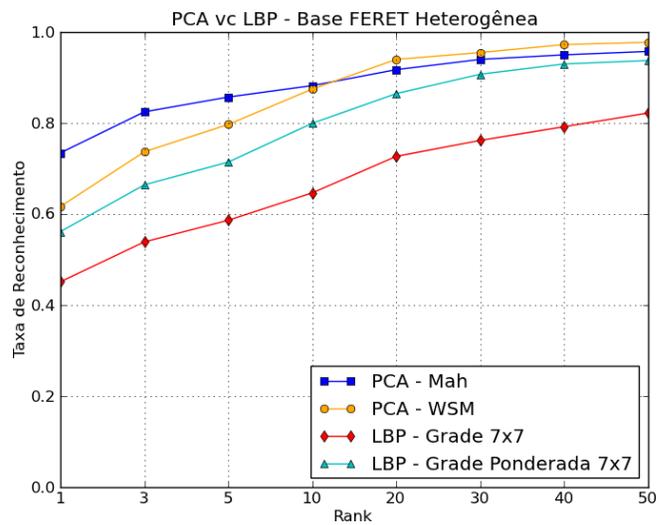
(a) Base FEI homogênea.



(b) Base FEI heterogênea.



(c) Base FERET homogênea.



(d) Base FERET heterogênea.

Figura 4. Comparação entre as técnicas LBP e PCA.

- [17] M. TURK and A. PENTLAND. Face recognition using eigenfaces. *Computer Vision and Pattern Recognition*, pages 586–591, 1991.
- [18] J. Wu and Z. Zhou. Face recognition with one training image per person. *Pattern Recognition Letters*, 23:1711–1719, December 2002.
- [19] W. ZHAO, R. CHELLAPPA, J. PHILLIPS, and A. ROSENFELD. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.