

LBP Estatístico Aplicado ao Reconhecimento de Expressões Faciais

Vagner do Amaral¹, Gilson A. Giraldi², Carlos E. Thomaz¹

¹ Depto. de Engenharia Elétrica - Centro Universitário da FEI
São Bernardo do Campo - São Paulo - Brasil

²Laboratório Nacional de Computação Científica (LNCC)
Petrópolis - Rio de Janeiro - Brasil

{vamaral, cet}@fei.edu.br, gilson@lncc.br

Abstract. Facial expressions play an important role on interpersonal relations by providing relevant information about the emotions and intentions of individuals during their communication process. In this paper, we propose and implement a new approach to improve the extraction and classification processes of Local Binary Patterns (LBP) on frontal face images. Based on the definition of groups of interest, composed of images with neutral expression or smiling, a spatial mapping of the statistically significant regions has been translated into prior information to identify the most relevant facial features. The experimental results, using two public face databases, have shown that our statistical prior-driven approach not only improves the classification accuracy of LBP, overcoming state-of-the-art corresponding methods, but also provides a methodology for computing the spatial weighting map for each application.

Resumo. Expressões faciais desempenham um papel importante nas relações interpessoais fornecendo informações relevantes sobre as emoções e intenções de um indivíduo. Este artigo propõe e implementa uma nova abordagem para aprimorar o processo de extração e classificação do LBP em análise de imagens frontais de faces. Com base na definição de grupos, compostos por faces com expressão neutra ou sorrindo, realizou-se um mapeamento das regiões estatisticamente significantes e utilizou-se esta informação para identificar as características faciais mais relevantes neste contexto. Os resultados obtidos demonstram que a abordagem proposta melhora a taxa de classificação do LBP e fornece uma metodologia automática para computar a ponderação espacial específica para cada aplicação de interesse.

1. Introdução

Expressões faciais desempenham um papel importante nas relações interpessoais fornecendo informações relevantes sobre as emoções e intenções de um indivíduo durante o processo de comunicação. O reconhecimento automático de tais variações provê inúmeras possibilidades para aplicações de interação humano-computador e tem atraído a atenção da comunidade científica nas últimas décadas [Fasel and Luettin 1999, Pantic and Rothkrantz 2000, Pantic and Rothkrantz 2003].

No final dos anos 90 e início da década passada, trabalhos sobre este tema investigaram diversas técnicas para extração de características e classificação de padrões em

expressões faciais, tais como Redes Neurais [Tian 2004, Zhang et al. 1998], Máquinas de Vetor de Suporte [Bartlett et al. 2003, Hadid et al. 2004], Análise de Discriminantes Lineares [Lyons et al. 1999] e Redes Bayesianas [Cohen et al. 2003]. Nos últimos anos, o método LBP, do inglês, Local Binary Patterns, originalmente proposto por [Ahonen et al. 2004], tem sido empregado com êxito na extração de características para reconhecimento de expressões, devido a sua capacidade para representação de micro padrões, presentes em imagens de face, e baixa complexidade computacional. Essas vantagens têm permitido desenvolver estudos sobre reconhecimento de padrões fisionômicos que utilizam regiões específicas da face e em imagens com baixa resolução, muitas vezes oriundas de ambientes reais [Shan et al. 2005, Fu and Wei 2008, Shan et al. 2009, Liu et al. 2009, Shoyaib et al. 2010, Burcin and Vasif 2011, Shan 2012]. Contudo, boa parte desses estudos tem se dedicado à extração de características e redução de dimensionalidade dos descritores de textura, atrelados à aplicação de classificadores robustos [Shan et al. 2005, Zilu and Xieyan 2008, Shan et al. 2009, Shoyaib et al. 2010, Nanni et al. 2011, Moore and Bowden 2011, Shan 2012], ignorando a possibilidade de utilização de informação a priori, oriunda das amostras, para ponderação espacial das regiões faciais mais relevantes ao problema.

Em 2007, Zhang et al [Zhang et al. 2007] propuseram um método de reconhecimento facial baseado no aprendizado estatístico de descritores locais, gerados pela técnica Local Gabor Binary Patterns (LGBP), visando tratar especificamente problemas de oclusão parcial em imagens frontais. Recentemente, Seo e Park [Seo and Park 2011, Seo and Park 2013] empregaram a mesma abordagem também à variações de expressão, utilizando a técnica Scale-Invariant Feature Transform (SIFT) para extração de característica e um classificador de vizinho mais próximo. De forma análoga, este artigo propõe e implementa uma nova técnica para ponderar a relevância dos descritores de textura gerados pelo LBP. A abordagem apresentada permite evidenciar características mais discriminantes entre dois grupos distintos de amostras, utilizando informação a priori obtida pela significância estatística das diferenças existentes explicitamente nos dados. Essas informações correspondem a atributos visuais de alto nível, como, por exemplo, elementos da face descritos de forma supervisionada por categorização específica das amostras em grupos de imagens frontais de face com expressão neutra ou sorrindo. Os resultados experimentais, obtidos com a técnica desenvolvida neste trabalho, mostram uma melhora considerável na taxa de reconhecimento em comparação com os métodos correspondentes já publicados na literatura científica afim.

A estrutura deste artigo é composta por 5 seções. Na seção seguinte, apresenta-se o método LBP original e a abordagem estatística proposta. Os experimentos são descritos na seção 3 e os resultados são apresentados e discutidos na seção 4. Por fim, na última seção, conclui-se o trabalho.

2. Local Binary Pattern (LBP)

Inicialmente desenvolvido para análise de texturas o LBP tem sido empregado, com sucesso, para extração de características no processo de reconhecimento e classificação de imagens de face. Isso ocorre pelo fato de que as faces podem ser vistas como uma composição de micro-padrões que são bem descritos por esta técnica [Ojala et al. 1996, Ahonen et al. 2004, Ahonen et al. 2006].

Neste método a intensidade v_c de cada pixel de uma imagem é substituída por um vetor binário b , determinado pela comparação entre a intensidade do pixel e as intensidades dos pixels vizinhos:

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

onde v_{ij} é valor de um pixel na posição (i, j) e $1 \leq i, j \leq N$, sendo N o tamanho da vizinhança (máscara). 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 ilustra este processo, para uma máscara de 3×3 pixels, e apresenta o resultado gerado por esta transformação.

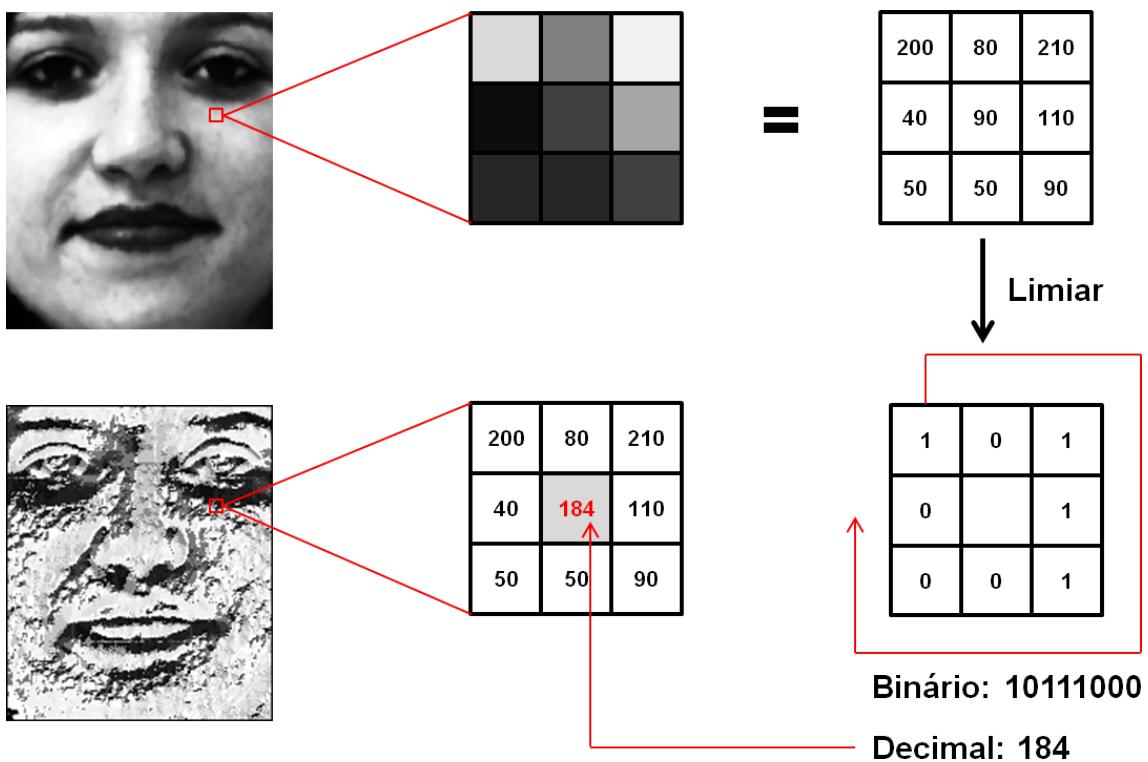


Figura 1. Operação básica do LBP, com valor decimal resultante 184.

A imagem gerada é dividida em regiões, habitualmente organizadas em uma malha regular. Entretanto, é possível empregar outras formas com, ou sem, sobreposição, conforme ilustrado na Figura 2. Os descritores de textura são extraídos de cada região isoladamente, calculando-se o histograma de intensidade dos pixels e os valores resultantes são concatenados em um único vetor para comparação das amostras.

No processo de classificação utilizam-se, basicamente, métricas que permitem explorar a relevância de algumas características faciais sobre outras [Zhao et al. 2003]. Deste modo, é possível definir pesos específicos para cada região j da grade de histogramas, utilizados, por exemplo, na distância Qui-Quadrado [Ahonen et al. 2006]:

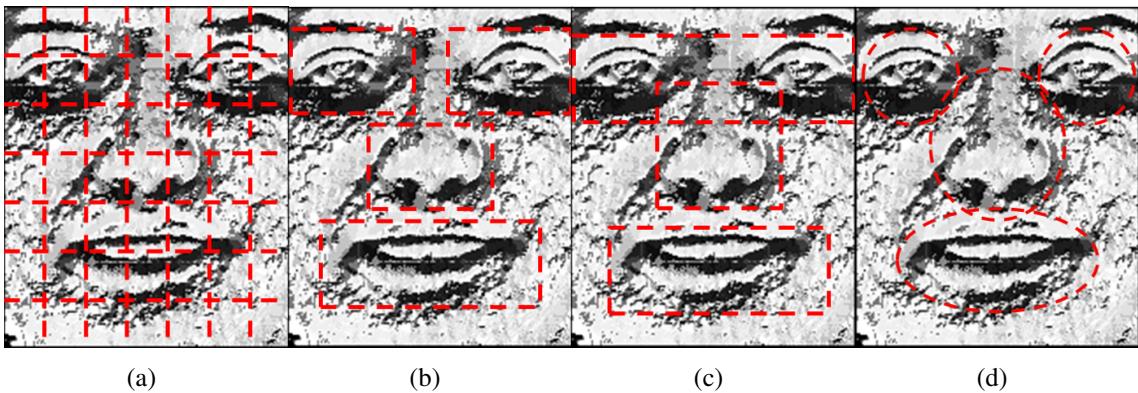


Figura 2. Regiões para extração dos histogramas: a) Malha regular 7×7 ; b / c / d) Formas alternativas com, ou sem, sobreposição.

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

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 .

2.1. Ponderação Estatística de Pesos Espaciais

A possibilidade de enfatizar características faciais específicas durante o processo de classificação dos descritores de textura gerados pelo LBP viabiliza a adaptação deste método de acordo com a sua aplicação. Trabalhos anteriores utilizaram heurísticas para a definição de pesos espaciais, destinados às tarefas de reconhecimento de indivíduos [Ahonen et al. 2004, Ahonen et al. 2006] e classificação de expressões [Shan et al. 2005]. Contudo, tais pesos foram determinados a partir do conhecimento de especialistas, sem o auxílio de nenhum procedimento automático para mineração de dados. Trabalhos mais recentes [Zhang et al. 2007, Seo and Park 2013] propuseram técnicas de aprendizado estatístico para ponderação de descritores locais pré-definidos, visando minimizar o efeito negativo das variações parciais em problemas de reconhecimento facial.

Neste trabalho, diferentemente dos demais, propõe-se acentuar as características mais relevantes ao processo de classificação de expressões, utilizando informação a priori obtida através de mineração estatística da intensidade dos pixels presentes nas amostras de interesse. Deste modo a ponderação espacial independe do descritor local empregado. Acredita-se que esta abordagem seja mais semelhante ao processo cognitivo humano de percepção facial, pois se baseia em valores presentes nas imagens originais que correspondem à atributos visuais de alto nível [Bruce and Young 2012].

Com o propósito de realizar um mapeamento espacial das regiões mais discriminantes em reconhecimento de expressões faciais, categorizadas aqui como amostras de imagens de face com expressão neutra (X_1) e sorrindo (X_2), calculou-se a imagem média de cada grupo de interesse para as duas bases utilizadas neste artigo [Phillips et al. 2000, Oliveira Junior and Thomaz 2006, Thomaz and Giraldi 2010], conforme ilustrado na Figura 3.

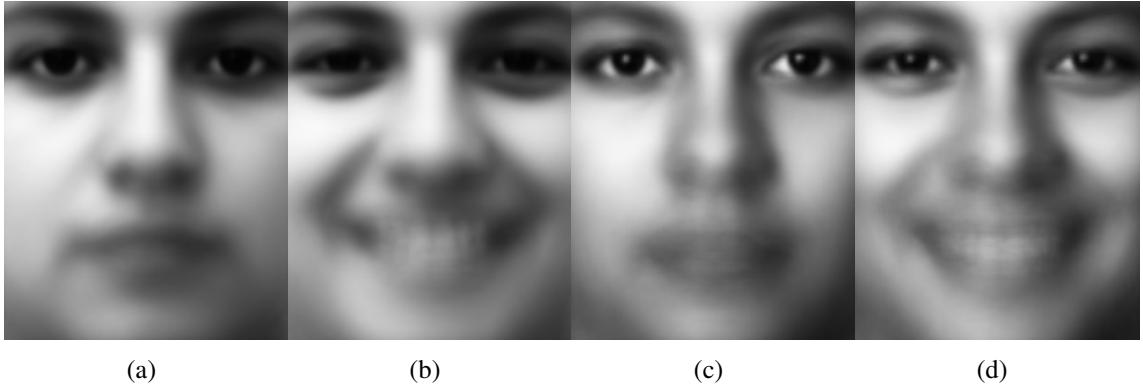


Figura 3. Imagens médias dos grupos de interesse: a) Expressão neutra, FEI Face Database; b) Expressão sorrindo, FEI Face Database; c) Expressão neutra, FERET Database; c) Expressão sorrindo, FERET Database.

Em seguida, aplicou-se o Teste t de Student [Spiegel and Stephens 2007] para cada pixel das médias geradas, conforme segue:

$$T = \frac{\overline{X}_1 - \overline{X}_2}{S_{X_1 X_2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad (3)$$

onde X_1 e X_2 são os grupos de imagens de face, n_1 é o número total de amostras do grupo X_1 e n_2 é o número total de amostras do grupo X_2 . $S_{X_1 X_2}$ pode ser definido como:

$$S_{X_1 X_2} = \sqrt{\frac{(n_1 - 1)S_{X_1}^2 + (n_2 - 1)S_{X_2}^2}{n_1 + n_2 - 2}}, \quad (4)$$

sendo $S_{X_1}^2$ e $S_{X_2}^2$ a variância dos grupos X_1 e X_2 , respectivamente. As figuras 4(a) e 4(b) ilustram este processo e exibem os valores de T obtidos para as bases analisadas.

Por fim, calcula-se a média dos valores absolutos de T em cada região j , da malha pré-definida para extração dos histogramas, e utiliza-se este valor como fator de ponderação w_j para o cálculo da distância Qui-Quadrado (ver Equação 2).

3. Experimentos

Nos experimentos foram utilizadas duas bases públicas: FEI Face Database [Oliveira Junior and Thomaz 2006, Thomaz and Giraldo 2010] e FERET Database [Phillips et al. 2000]. Em ambas as bases, selecionaram-se 400 imagens de face de 200 indivíduos em posição frontal com expressão neutra e sorrindo. Todas imagens foram previamente normalizadas pelo procedimento descrito em [Amaral et al. 2009].

Para definição das regiões de extração dos histogramas, empregou-se uma grade 7×7 devido ao bom desempenho da mesma em aplicações de reconhecimento facial automático [Ahonen et al. 2004, Ahonen et al. 2006, Amaral and Thomaz 2012]. As imagens foram divididas em dois grupos, de acordo com as expressões faciais presentes em cada base, e classificaram-se todas as amostras utilizando como referência o vizinho mais próximo, determinado pela distância Qui-Quadrado. Visando comparar o

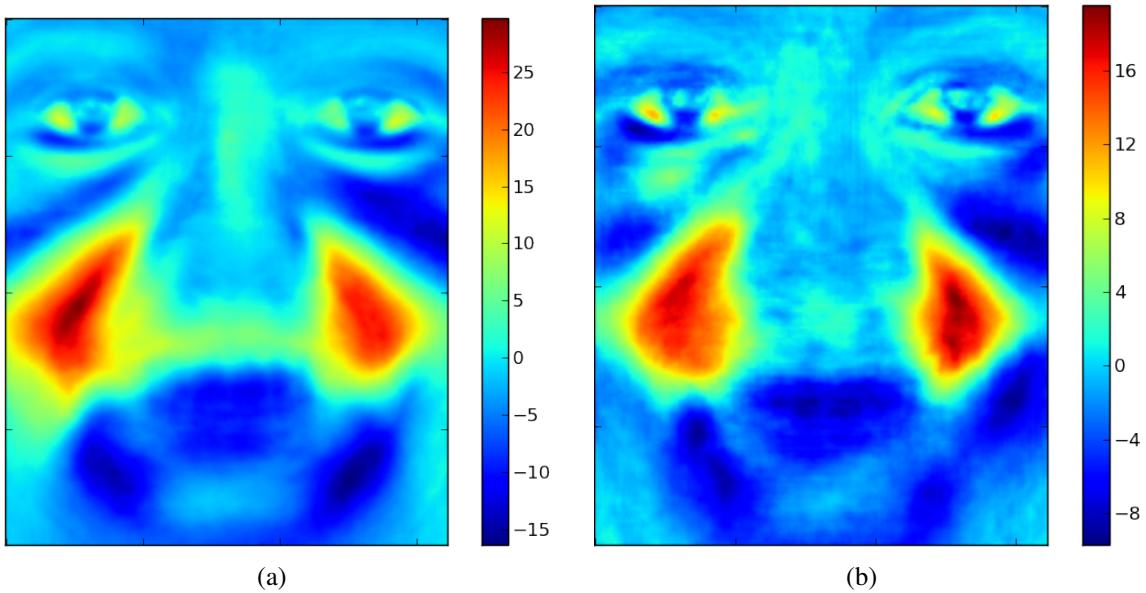


Figura 4. Mapeamento espacial das regiões faciais estatisticamente discriminantes entre imagens neutras e sorrindo: a) FEI Face Database [Oliveira Junior and Thomaz 2006, Thomaz and Giraldo 2010]; b) FERET Database [Phillips et al. 2000].

método proposto, para o cálculo dos pesos w_j , com as heurísticas vigentes na literatura [Ahonen et al. 2004, Ahonen et al. 2006, Shan et al. 2005], avaliaram-se quatro mapas de ponderação espacial, conforme ilustrado na Figura 5.

4. Resultados

Nos experimentos realizados verificou-se a taxa de classificação para cada um dos quatro modelos de grade pré-definidos: Padrão, sem ponderação espacial; Ahonen, heurística para reconhecimento facial [Ahonen et al. 2004, Ahonen et al. 2006]; Shan, heurística para classificação de expressões [Shan et al. 2005]; Estatístico, ponderado com base nos grupos de interesse.

Os resultados gerados são descritos na Tabela 1, onde pode-se verificar, em destaque, que a abordagem estatística proposta obteve os melhores resultados, com um aumento significativo no percentual de acertos para as duas bases analisadas. A heurística de Shan [Shan et al. 2005] também apresentou resultados superiores ao LBP padrão e a heurística de Ahonen [Ahonen et al. 2004, Ahonen et al. 2006], enfatizando os benefícios da utilização de uma ponderação espacial específica para cada aplicação.

Base	Grade / Mapa de Ponderação Espacial			
	Padrão	Ahonen	Shan	Estatístico
FEI	70%	58%	74%	88%
FERET	50%	40%	53%	71%

Tabela 1. Taxas de reconhecimento obtidas nos experimentos.

As tabelas 2 e 3 apresentam visualmente exemplos de vizinhos mais próximos para os quatro modelos de grade avaliados. Em alguns casos, o LBP padrão, que não

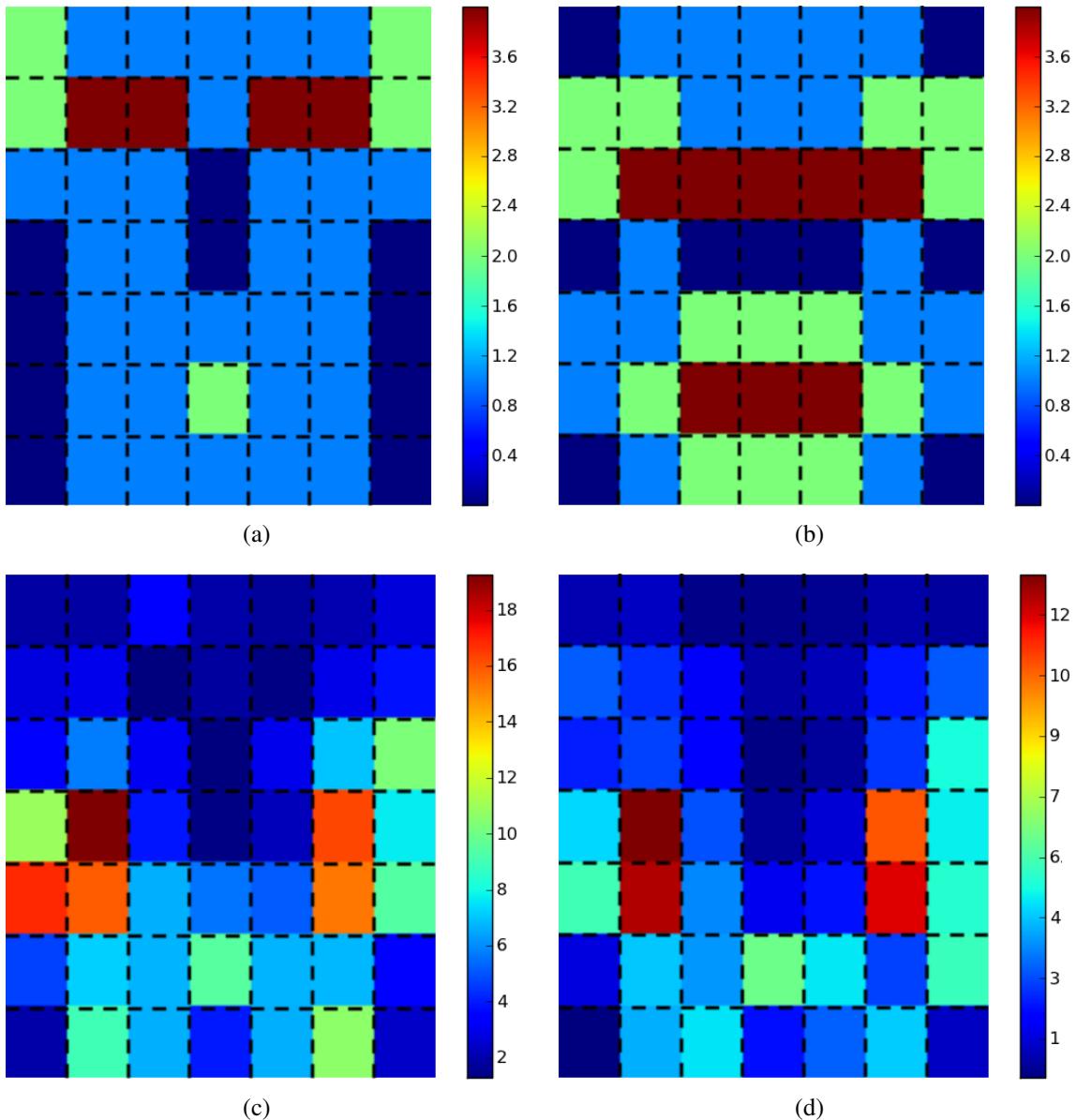


Figura 5. Mapas de ponderação espacial em uma grade 7x7. a) Grade de reconhecimento, proposta por [Ahonen et al. 2004, Ahonen et al. 2006]; b) Grade para classificação de expressão, adaptada de [Shan et al. 2005]; c) Grade estatística proposta, gerada com base nas imagens da base FEI; d) Grade estatística proposta, gerada com base nas imagens da base FERET.

utiliza ponderação espacial, classificou corretamente a expressão facial. No entanto, a heurística de Ahonen [Ahonen et al. 2004, Ahonen et al. 2006], proposta para a tarefa de reconhecimento facial, identificou amostras independentemente da expressão. A ponderação definida por Shan [Shan et al. 2005], destinada ao reconhecimento de expressões faciais, demonstrou ser uma sugestão válida, porém, classificou erroneamente algumas amostras. Pode-se observar que, em geral, o LBP padrão e a heurística de Ahonen [Ahonen et al. 2004, Ahonen et al. 2006] tendem a localizar o indivíduo mais parecido, enquanto a heurística de Shan [Shan et al. 2005] e a ponderação estatística localizam amostras distintas que apresentem a mesma expressão facial.

Amostra de Teste	Vizinho mais próximo			
	Padrão	Ahonen	Shan	Estatístico

Tabela 2. Exemplos de reconhecimento de expressões faciais em amostras da base FEI Face Database.

Dentre os resultados incorretos, obtidos com o método proposto, verificou-se que algumas amostras de teste não apresentavam uma expressão facial semelhante ao esperado para cada grupo de interesse, ou seja, face neutra, com características de boca fechada e sem contração das bochechas ou face sorrindo com a boca aberta e bochechas contraídas. Acredita-se que a aplicação de grades mais detalhadas permita a identificação de variações sutis e consequentemente a classificação de micro expressões [Pfister et al. 2011, Li et al. 2013], presentes nesses casos. A Tabela 4 apresenta alguns exemplos identificados nas duas bases.

Por fim, constatou-se também, através das figuras 4(a) e 4(b) que a ponderação gerada pela abordagem proposta neste artigo identificou os maiores valores de significância estatística nas regiões referentes às bochechas e, portanto, evidenciaram-se estas áreas nos mapas 5(c) e 5(d), divergindo das heurísticas ilustradas nas figu-

Amostra de Teste	Vizinho mais próximo			
	Padrão	Ahonen	Shan	Estatístico

Tabela 3. Exemplos de reconhecimento de expressões faciais em amostras da base FERET.

ras 5(a) [Ahonen et al. 2004, Ahonen et al. 2006] e 5(b) [Shan et al. 2005], definidas, respectivamente, para reconhecimento e classificação de expressões.

5. Conclusão

Neste artigo implementou-se uma nova abordagem para ponderação espacial dos descriptores de textura empregados no processo de classificação do LBP visando aprimorar a aplicação desta técnica em reconhecimento de expressões faciais. Com base na definição de grupos de interesse, realizou-se um mapeamento das regiões faciais estatisticamente significantes e utilizou-se esta informação a priori para identificar as características mais relevantes. Os resultados obtidos demonstraram que a abordagem proposta melhora consideravelmente a eficiência do LBP e fornece uma metodologia automática para computar a ponderação espacial específica para cada aplicação.

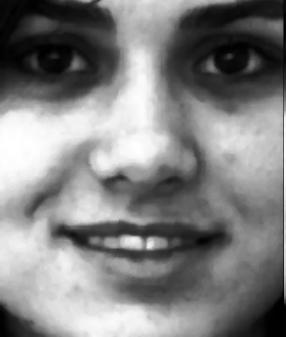
Base FEI		Base FERET	
Amostra	Erro	Amostra	Erro
			
			

Tabela 4. Exemplos de reconhecimentos incorretos realizados pela abordagem proposta.

Em trabalhos futuros, pretende-se explorar outras aplicações para o método proposto, como, por exemplo, classificação de gênero. Acredita-se, também, que a definição de um limiar estatístico, para triagem das regiões mais discriminantes, permita uma redução sensível de dimensionalidade e a utilização de malhas mais adequadas possa beneficiar a ponderação do fator de significância proporcionado pelo mapeamento espacial.

Agradecimentos

O autor principal desse trabalho gostaria de agradecer à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo auxílio fornecido através de bolsa PROSUP.

Referências

- Ahonen, T., Hadid, A., and Pietikainen, M. (2004). Face recognition with local binary patterns. *Computer Vision ECCV 2004*, 3021:469–481.
- Ahonen, T., Hadid, A., and Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Transaction Pattern Analysis and Machine Intelligence*, 28:2037–2041.
- Amaral, V., Figaro-Garcia, C., Gattás, G. J. F., and Thomaz, C. E. (2009). Normalização espacial de imagens frontais de face em ambientes controlados e não-controlados. *FaSCI-Tech*, 1(1).
- Amaral, V. and Thomaz, C. E. (2012). Extração e comparação de características locais e globais para o reconhecimento automático de imagens de faces. In *Proceedings of the VIII Workshop de Visão Computacional*. Universidade Federal de Goias.

- Bartlett, M. S., Littlewort, G., Fasel, I., and Movellan, J. R. (2003). Real time face detection and facial expression recognition: Development and applications to human computer interaction. In *Conference on Computer Vision and Pattern Recognition Workshop*, volume 5, pages 53–53.
- Bruce, V. and Young, A. (2012). *Face Perception*. Taylor & Francis.
- Burcin, K. and Vasif, N. V. (2011). Down syndrome recognition using local binary patterns and statistical evaluation of the system. *Expert Syst. Appl.*, 38(7):8690–8695.
- Cohen, I., Sebe, N., Chen, L., Garg, A., and Huang, T. S. (2003). Facial expression recognition from video sequences: Temporal and static modelling. In *Computer Vision and Image Understanding*, pages 160–187.
- Fasel, B. and Luettin, J. (1999). Automatic facial expression analysis: A survey. *PATTERN RECOGNITION*, 36(1):259–275.
- Fu, X. and Wei, W. (2008). Centralized binary patterns embedded with image euclidean distance for facial expression recognition. In *Proceedings of the 2008 Fourth International Conference on Natural Computation - Volume 04*, ICNC ’08, pages 115–119.
- Hadid, A., Pietikainen, M., and Ahonen, T. (2004). A discriminative feature space for detecting and recognizing faces. In *Computer Vision and Pattern Recognition, 2004. Proc. of the 2004 IEEE Computer Society*, volume 2, pages II–797–II–804.
- Li, X., Pfister, T., Huang, X., Zhao, G., and Pietikainen, M. (2013). A spontaneous micro-expression database: Inducement, collection and baseline. In *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, pages 1–6.
- Liu, W., juan Li, S., and jiang Wang, Y. (2009). Automatic facial expression recognition based on local binary patterns of local areas. In *Information Engineering, 2009. ICIE ’09. WASE International Conference on*, volume 1, pages 197–200.
- Lyons, M. J., Budynek, J., and Akamatsu, S. (1999). Automatic classification of single facial images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(12):1357–1362.
- Moore, S. and Bowden, R. (2011). Local binary patterns for multi-view facial expression recognition. *Computer Vision and Image Understanding*, 115(4):541 – 558.
- Nanni, L., Brahnam, S., and Lumini, A. (2011). Combining different local binary pattern variants to boost performance. *Expert Systems with Applications*, 38(5):6209 – 6216.
- Ojala, T., Pietikainen, M., and Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59.
- Oliveira Junior, L. L. and Thomaz, C. E. (2006). Captura e alinhamento de imagens: Um banco de faces brasileiro. Technical report, Centro Universitário da FEI.
- Pantic, M. and Rothkrantz, L. J. M. (2000). Automatic analysis of facial expressions: the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445.
- Pantic, M. and Rothkrantz, L. J. M. (2003). Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 91(9):1370–1390.

- Pfister, T., Li, X., Zhao, G., and Pietikainen, M. (2011). Recognising spontaneous facial micro-expressions. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1449–1456.
- Phillips, P. J., Moon, H., Rizvi, S. A., and Rauss, P. (2000). The feret evaluation methodology for face-recognition algorithms. In *IEEE Transaction on Pattern Analysis and Machine Intelligence*, volume 22, pages 1090–1104. IEEE Computer Society.
- Seo, J. and Park, H. (2011). A robust face recognition through statistical learning of local features. In *Proceedings of the 18th international conference on Neural Information Processing*, ICONIP’11, pages 335–341, Berlin, Heidelberg. Springer-Verlag.
- Seo, J. and Park, H. (2013). Robust recognition of face with partial variations using local features and statistical learning. *Neurocomputing*.
- Shan, C. (2012). Learning local binary patterns for gender classification on real-world face images. *Pattern Recogn. Lett.*, 33(4):431–437.
- Shan, C., Gong, S., and McOwan, P. W. (2005). Robust facial expression recognition using local binary patterns. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*. IEEE.
- Shan, C., Gong, S., and McOwan, P. W. (2009). Facial expression recognition based on local binary patterns: A comprehensive study. *Image Vision Comp.*, 27(6):803–816.
- Shoyaib, M., Abdullah-Al-Wadud, M., Youl, J. M., Alam, M. M., and Chae, O. (2010). Facial expression recognition based on a weighted local binary pattern. In *13th International Conference on Computer and Information Technology*, pages 321–324.
- Spiegel, M. and Stephens, L. (2007). *Schaum’s Outline of Statistics*. Schaum’s Outline Series. McGraw-hill.
- Thomaz, C. E. and Giraldi, G. A. (2010). A new ranking method for principal components analysis and its application to face image analysis. *Image Vision Computing*, 28(6):902–913.
- Tian, Y.-L. (2004). Evaluation of face resolution for expression analysis. In *Conference on Computer Vision and Pattern Recognition Workshop*, pages 82–82.
- Zhang, W., Shan, S., Chen, X., and Gao, W. (2007). Local gabor binary patterns based on kullback leibler divergence for partially occluded face recognition. *Signal Processing Letters, IEEE*, 14(11):875–878.
- Zhang, Z., Lyons, M., Schuster, M., and Akamatsu, S. (1998). Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron. In *Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998. Proceedings.*, pages 454–459.
- Zhao, W., Chellappa, R., Phillips, J., and Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458.
- Zilu, Y. and Xieyan, F. (2008). Combining lbp and adaboost for facial expression recognition. In *Signal Processing, 2008. ICSP 2008. 9th International Conference on*, pages 1461–1464.