A Statistical Quadtree Decomposition to Improve Face Analysis

Vagner Amaral[1], Gilson A. Giraldi[2] and Carlos E. Thomaz[1]

[1]*Department of Electrical Engineering, Centro Universitario da FEI, Av. Humberto de Alencar Castelo Branco 3972, Sao Bernardo do Campo, Sao Paulo, Brazil*

[2]*Department of Computer Science, LNCC, Av. Getulio Vargas 333, Petropolis, Rio de Janeiro, Brazil*

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

Keywords: Spatial Mapping, Task Driven, Face Recognition, Local Binary Pattern, Quadtree.

Abstract: The feature extraction is one of the most important steps in face analysis applications and this subject always received attention in the computer vision and pattern recognition areas due to its applicability and wide scope. However, to define the correct spatial relevance of physiognomical features remains a great challenge. It has been proposed recently, with promising results, a statistical spatial mapping technique that highlights the most discriminating facial features using some task driven information from data mining. Such priori information has been employed as a spatial weighted map on Local Binary Pattern (LBP), that uses Chi-Square distance as a nearest neighbour based classifier. Intending to reduce the dimensionality of LBP descriptors and improve the classification rates we propose and implement in this paper two quad-tree image decomposition algorithms to task related spatial map segmentation. The first relies only on split step (top-down) of distinct regions and the second performs the split step followed by a merge step (bottom-up) to combine similar adjacent regions. We carried out the experiments with two distinct face databases and our preliminary results show that the top-down approach achieved similar classification results to standard segmentation using though less regions.

1 INTRODUCTION

The feature extraction is one of the most important steps in face analysis applications and this subject always received attention in the computer vision and pattern recognition areas due to its applicability and wide scope. However, defining the correct spatial relevance of physiognomical features remains a great challenge (Blais et al., 2012). In the last few years a method called Local Binary Pattern (LBP) has been successfully used in face analysis research (Shan et al., 2009; Pietikäinen et al., 2011; Shan, 2012; Torrisi et al., 2015; Santarcangelo et al., 2015). Nevertheless, many studies on this method have ignored the contribution provided by the contextual information (Shan et al., 2005; Shan et al., 2009; Shan, 2012). Thus, a recently proposed approach (Amaral et al., 2013) has used a statistical technique, employing some task driven information from data mining, to highlight the most discriminant facial features and provide a spatial weighted map to LBP. This approach has enabled subsequent studies to explore the relevance of physiognomical features according to the task under investigation (Amaral et al., 2014; Amaral et al., 2015). At first, they employed the uniform grid, that consists of a square non-overlapped segmentation of the face images to extract the features descriptors. But then, they also investigated a non uniform segmentation and concluded that it would be an interesting way to describe facial features with few regions.

In this context, the aim of this paper is to improve the segmentation of the task-driven statistical spatial maps, intending to reduce the dimensionality of LBP spatial feature descriptors and improve the accuracy of classification in face analysis applications. We expect to show as a result that the adaptive decomposition emphasises the facial features by their contextual relevance, providing more consistent spatial segments, that is, segments that contain more pixels with similar values. More specifically, in this study we investigate the spatial segmentation of facial features in gender and facial expression classifications.

This paper is organised as follows. Next, in section 2, we review the LBP. In section 3 we show the statistical spatial mapping. Then, section 4 describes the quadtree image segmentation techniques and their application in this study. Experiments and results have been explained in sections 5 and 6, respectively. Finally, in section 7, we conclude the paper, discussing its main contribution and future works.

Amaral, V., Giraldi, G. and Thomaz, C.
A Statistical Quadtree Decomposition to Improve Face Analysis.
In *Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2016)*, pages 375-380
ISBN: 978-989-758-173-1

2 LOCAL BINARY PATTERN

Initially implemented as a texture operator (Ojala et al., 1996), LBP has been widely employed in face image processing due to its low computational complexity and invariance to monotonic gray level changes (Ahonen et al., 2006). In short, the original approach labels the pixels of an image to encode the local structure around each pixel by thresholding the neighbourhood by its center pixel value:

$$b_{ij} = \begin{cases} 0, v_{ij} < v_c \\ 1, v_{ij} \geq v_c, \end{cases} \tag{1}$$

where v_c is center pixel value, v_{ij} is the pixel value in (i, j) position, wherein $1 \leq i, j \leq N$ and N is the neighbourhood size. The b_{ij} values are concatenated as a binary number and converted to decimal basis to label the central pixel. Figure 1 shows the images before and after this process. The output image is divided into R_j regions, $j = 1, 2, ..., N$, usually arranged in a regular grid. The textures descriptors are taken out from each region R_j by their histograms of LBP labels that are grouped in a single feature vector. In the classification process the distinct relevance of physiognomical features are often emphasized (Zhao et al., 2003). Therefore specific w_j weights are defined for each R_j region. In this work, for example, we used a Chi-Square distance (Ahonen et al., 2006):

(a) (b)

Figure 1: LBP initial step: a) Original image; b) LBP pixel labeled image.

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

where x and y are feature vectors to be compared, $x_{i,j}$ is the i histogram bin corresponding to j-th region and w_j its pre-defined weight.

2.1 Uniform Patterns

In this work, we have implemented an useful extension of the original LBP operator called uniform pattern (Ojala et al., 2002; Ahonen et al., 2006), which reduces the length of the feature vector and provides a simple rotation invariant descriptor. This approach is motivated by the fact that some binary patterns occur more frequently in texture images than others. A pixel neighbourhood is called uniform if the pattern contains at most two binary transitions. Using this extension, the length of the feature vector histograms for a $3x3$ kernel reduces from 256 values to 59 values, being 58 bins for uniform patterns and 1 bin to represent all non uniform patterns.

3 STATISTICAL SPATIAL MAP

The possibility to emphasise some physiognomical features among others, provided by LBP, allows us to improve the classification step in face image analysis. Thus, a recent work proposed a method that highlights more relevant facial regions in according to the task, employing the statistical significance extracted from pixel intensity of samples (Amaral et al., 2013). This approach consists of calculating the t-Student test from two distinct face image sample groups to generate a statistical spatial map, as follows:

$$T = \frac{\overline{X_1} - \overline{X_2}}{S_{X_1X_2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \tag{3}$$

where X_1 and X_2 are face image groups, n_1 is the total number of samples from group X_1 and n_2 is the total number of samples from group X_2. $S_{X_1X_2}$ is given by:

$$S_{X_1X_2} = \sqrt{\frac{(n_1 - 1)S^2_{X_1} + (n_2 - 1)S^2_{X_2}}{n_1 + n_2 - 2}}, \tag{4}$$

where $S^2_{X_1}$ and $S^2_{X_2}$ are the variances of the X_1 and X_2 groups, respectively.

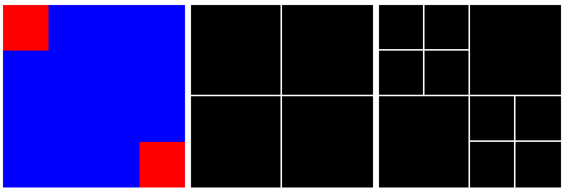
In the uniform segmentation procedure the map is divided in a regular grid, composed of rectangular regions. Then, for each j region, we calculate the absolute mean value for T and apply this information as w_j weight in Chi-Square distance (Equation 2) to compare two feature vectors x and y. However, another approaches that generate non-regular regions have been analyzed intending to optimize the use of statistical spatial map (Amaral et al., 2014) as well.

4 QUADTREE DECOMPOSITION

Quadtree decomposition has been widely used in digital image analysis to define regions of interest for further processing as image segmentation, feature selection, object detection, sample annotation, among others (Conde-Marquez et al., 2011). This technique consists of a hierarchical data structure whose nodes are recursively subdivided in four parts until a predefined split condition is not satisfied. This process performs a top-down approach beginning with a single node, that represents the entire image, and give a structure with variable block sizes at the end, where smaller blocks describe fine details more accurately and bigger blocks incorporates similar regions to represent them with fewer information as possible (Samet, 1984; Muhsin et al., 2014). In this work we have used a simple pixel value homogeneity criteria to perform the blocks partition, defined as follow:

$$h = \begin{cases} True, max(r) - min(r) \geq t \\ False, max(r) - min(r) < t, \end{cases} \tag{5}$$

where r is the analysed region and t is the threshold value used as split criteria to set block homogeneity. Figure 2 illustrate this process.



(a) (b) (c)

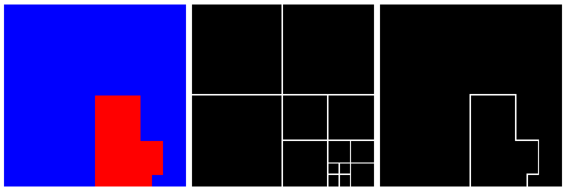
Figure 2: Quadtree segmentation: a) Image sample; b) First slice throughout entire image; c) Second slice only in specific blocks that satisfy the segmentation criteria.

To establish a parsimonious feature vector representation, we have defined the minimum block size in $8x8$ pixels, i.e. 64 values, due to the length of local texture descriptor histogram in uniform LBP method that contains 59 bins (Ojala et al., 2002).

4.1 Bottom-up Strategy

In order to overcome the limitations of quadtree segmentation and to reduce dimensionality of data structure, we perform a second processing step based on a bottom-up strategy (Usó, 2003; Fu et al., 2013; Scholefield and Dragotti, 2014), that consists of merging similar adjacent regions from quadtree that satisfies a joint condition, providing non square regions as shown in Figure 3.

Our implemented bottom-up approach receives a pre-segmented quadtree as input, then it performs a search throughout the tree to find the best combination between two adjacent blocks that satisfy the join



(a) (b) (c)

Figure 3: Segmentation differences: a) Image sample; b) Default quadtree segmentation with symmetric slices, composed by 13 square blocks; c) Bottom-up strategy with merge step, composed of only 2 non square blocks.

condition and repeat this procedure iteratively until the quadtree hasn't a pair of blocks that comply with the merge criteria, defined as bellow:

$$h = \begin{cases} True, max(r_1, r_2) - min(r_1, r_2) \leq t \\ False, max(r_1, r_2) - min(r_1, r_2) > t, \end{cases} \tag{6}$$

where r_1 and r_2 are the adjacent regions and t is the threshold used as homogeneity condition to join.

5 EXPERIMENTS

In this section we describe automatic gender and facial expression classification experiments carried out to compare the top-down and bottom-up quadtree approaches in different face analysis tasks.

5.1 Databases and Setup

To perform the proposed experiments we used two public available sample sources that meet the necessary requirements for the experiments. The first is the FEI Face Database (Thomaz and Giraldi, 2010), employed to training and testing the proposed approaches. And the second is the Grayscale FERET (Phillips et al., 2000), used to validate the best results obtained with FEI Face Database. In this study we employ only frontal face images, of both genders, and two samples for each subject, one with neutral facial expression and the other with smiling facial expression, providing a total of 400 images from each database. All the images were normalized to reduce sample variability by the following steps: rotation, until to align the both pupils with the horizontal axis; resize, to adjust the interpupillary distance; cutting to specified measures; conversion to grey scale, between 0 and 255; and finally histograms equalization of pixels. Figure 4 shows the adjusting dimensions.

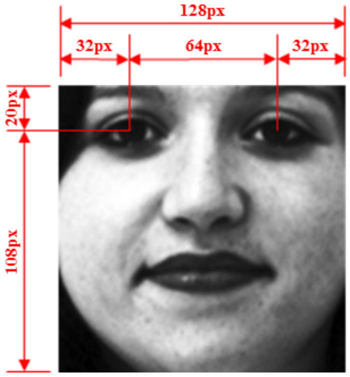
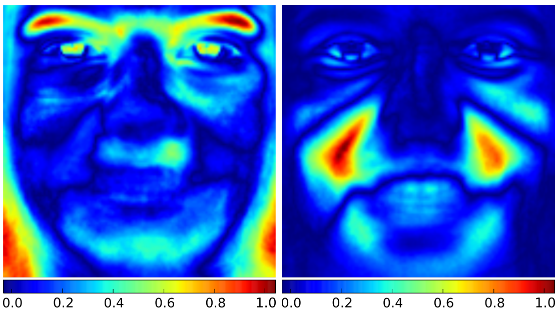


Figure 4: Dimensions used in the normalisation process.

5.2 Procedure

Initially, we generate the statistical spatial maps for gender and facial expression classification, using FEI Face Database images. Figure 5 shown such statistical significance maps.



(a) (b)

Figure 5: Statistical spatial maps: a) Gender (male and female); Facial expression (neutral and smile).

The statistical spatial maps were segmented using 19 distinct threshold values, between 0 and 1, with intervals of 0.05, for each proposed quadtree approach. This process provides 28 maps, containing the facial regions to extract the texture descriptor histograms and calculate spatial weights. Then, we arranged the samples in four classification groups for two classification tasks: gender (male vs. female) and facial expression (neutral vs. smiling). Next, it performed the classification procedure with the specific task driven maps. Thus, each sample was compared to all other samples, except to samples from the same subject, to identify the nearest neighbour by Chi-Square distance (Equation 2) as classification criteria.

To compare with earlier studies (Amaral et al., 2013; Amaral and Thomaz, 2013), we performed the same experiment using 4 uniform grid layouts for each task (2x2, 4x4, 8x8 and 16x16), with and without spatial weights. To validate the results obtained with the FEI Face Database, we have reproduced the experiments with Grayscale FERET employing only the maps that achieved the best classification rates.

6 CLASSIFICATION RESULTS

In this section, we present the classification accuracy for gender and facial expression using a simple quadrtree strategy and a quadtree with subsequent merge step. These approaches were verified with 19 thresholds and the results are shown in figures 6 and 7.

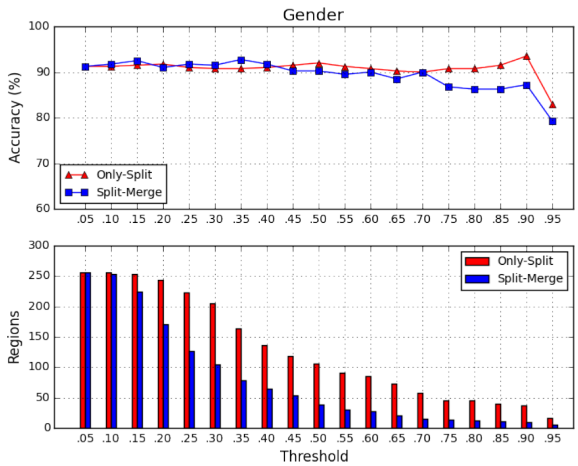


Figure 6: Gender classification rates and their corresponding number of regions for both approaches.

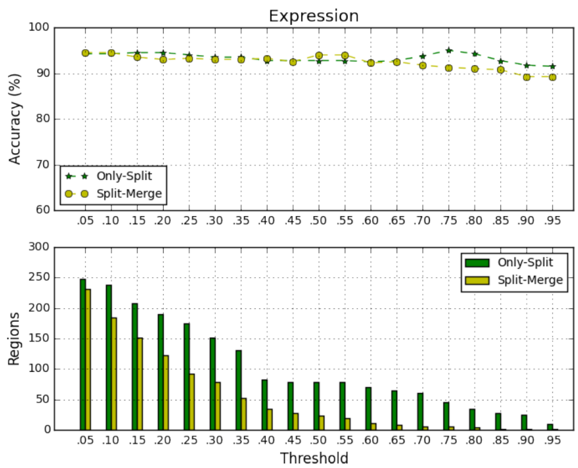
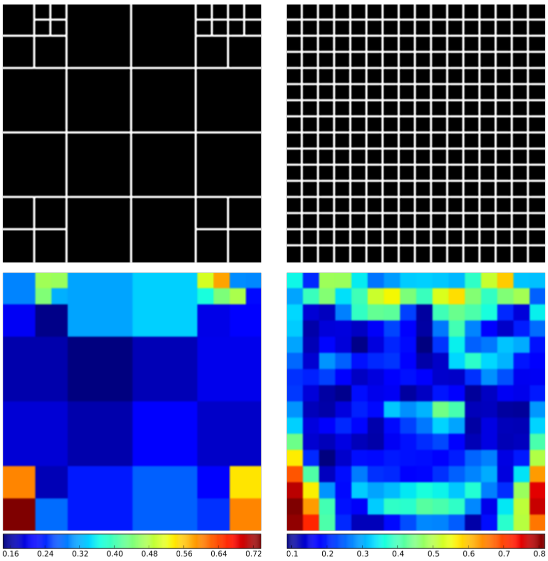


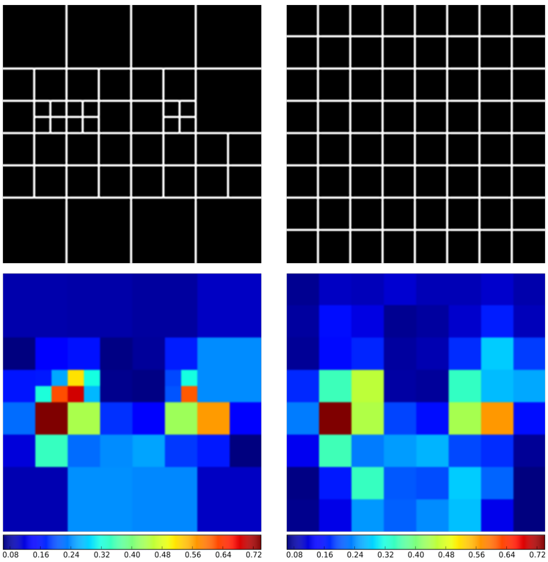
Figure 7: Expression classification rates and their corresponding number of regions for both approaches.

The spatial maps that provided the best classification rates are illustrated in Figures 8 and 9, which correspond to the 0.90 and 0.75 threshold shown on the previous Figures 6 and 7, respectively.



(a) 37 regions (b) 256 regions

Figure 8: Spatial maps for gender analysis. The images above highlights the segmentation maps and below are their corresponding spatial weights: a) Quadtree with the best accuracy; b) Uniform grid with the best accuracy.



(a) 46 regions (b) 64 regions

Figure 9: Spatial maps for facial expression analysis. The images above highlights the segmentation maps and below are their corresponding spatial weights: a) Quadtree with the best accuracy; b) Uniform grid with the best accuracy.

The best classification rates achieved in experiments with the above maps and FEI Database are shown in Table 1. Table 2 highlights the results obtained with the FERET Database to validate the proposed spatial mapping approach.

As shown in the previous tables, the quadtree top-

Table 1: Classification rates for FEI Face Database.

Map	**Gender**		**Expression**	
	Rate	**Size**	**Rate**	**Size**
QT Top-Down	93.5%	37	95.0%	46
Weighted Grid	91.2%	256	95.2%	64
Default Grid	88.5%	256	96.0%	256

Table 2: Classification rates for FERET Face Database.

Map	**Gender**		**Expression**	
	Rate	**Size**	**Rate**	**Size**
QT Top-Down	77.2%	37	85.7%	46
Weighted Grid	79.2%	256	86.7%	64
Default Grid	76.5%	256	84.7%	256

down approach proposed in this study achieves similar classification results to static grids, using though much less regions. The distinct classification rates observed between the databases have been probably given by the differences in the original sample quality. The better resolution of FEI face images provided more texture details to the LBP feature descriptors.

7 CONCLUSION

In this paper we have compared two quadtree image segmentation methods to find the most important physiognomical regions, in according to statistical spatial maps generated for specific binary classification tasks. One of them performs the top-down quadtree decomposition, with only recursive split function, and the other approach executes the merge step after satisfying the split criteria (bottom-up). The experiments were limited to binary selection tasks due to the t-Student test restrictions.

Analyzing the preliminary results we can see that the top-down quadtree segmentation has obtained similar classification results to the uniform grid layout using, however, much less regions, based on the same task driven statistical spatial maps. In our opinion, these differences are very relevant in terms of storage and computational time, mainly in real time applications and mobile platforms, because the segmentation with less regions implies on fewer histograms to store and compare during the classification process.

Therefore, as future work, we intend to explore the proposed quadtree decomposition for other facial classification problems, e. g. disgusting vs. angry or fear vs. surprise, and for face recognition as well. Besides, we believe that another physiognomical relevance prior information could be employed as decomposition criteria, such as the human perception maps provided by eye tracking devices.

ACKNOWLEDGEMENTS

The authors would like to thank the financial support provided by FAPESP (2012/22377-6), CNPq (309532/2014-0) and the Royal Society (NA140147).

REFERENCES

Ahonen, T., Hadid, A., and Pietikäinen, M. (2006). Face description with local binary patterns: application to face recognition. *IEEE Trans. Pattern Analisys and Machine Intelligence*, 28:2037–2041.

Amaral, V., Giraldi, G. A., and Thomaz, C. E. (2013). LBP estatístico aplicado ao reconhecimento de expressões faciais. In *Proceedings of the X Encontro Nacional de Inteligência Artificial e Computacional, ENIAC'13*, Fortaleza, Ceara, Brasil.

Amaral, V., Giraldi, G. A., and Thomaz, C. E. (2014). Segmentação espacial não uniforme aplicada ao reconhecimento de gênero e expressoes faciais. In *Proceedings of the XI Encontro Nacional de Inteligência Artificial e Computacional, ENIAC'14*, São Carlos, São Paulo, Brazil.

Amaral, V., Giraldi, G. A., and Thomaz, C. E. (2015). Statistical and cognitive spatial mapping applied to face analysis. In *Proceedings of the 28th SIBGRAPI, Conference on Graphics, Patterns and Images - Workshop of Works in Progress*, Salvador, Bahia, Brazil.

Amaral, V. and Thomaz, C. E. (2013). Um estudo sobre o detalhamento espacial de descritores locais aplicados ao reconhecimento de gênero e expressões faciais. In *Anais do 3 Simpósio de Pesquisa do Grande ABC*, São Bernardo do Campo, São Paulo, Brazil.

Blais, C., Roy, C., Fiset, D., Arguin, M., and Gosselin, F. (2012). The eyes are not the window to basic emotions. *Neuropsychologia*, 50(12):2830–2838.

Conde-Marquez, G. R., Escalante, H. J., and Sucar, E. (2011). Simplified quadtree image segmentation for image annotation. In Sucar, E. and Escalante, H. J., editors, *Proceedings of the 2010 Automatic Image Annotation and Retrieval Workshop (2010)*, volume 719, pages 24–34. CEUR-Workshop Proceedings.

Fu, G., Zhao, H., Li, C., and Shi, L. (2013). Segmentation for high-resolution optical remote sensing imagery using improved quadtree and region adjacency graph technique. *Remote Sensing*, 5(7):3259.

Muhsin, Z. F., Rehman, A., Altameem, A., Saba, T., and Uddin, M. (2014). Improved quadtree image segmentation approach to region information. *The Imaging Science Journal*, 62(1):56–62.

Ojala, T., Pietikäinen, M., and Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59.

Ojala, T., Pietikäinen, M., and Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 24(7):971–987.

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, Washington, DC, USA. IEEE Computer Society.

Pietikäinen, M., Zhao, G., Hadid, A., and Ahonen, T. (2011). *Computer Vision Using Local Binary Patterns*. Number 40 in Computational Imaging and Vision. Springer.

Samet, H. (1984). The quadtree and related hierarchical data structures. *ACM Comput. Surv.*, 16(2):187–260.

Santarcangelo, V., Farinella, G. M., and Battiato, S. (2015). Gender recognition: Methods, datasets and results. In *IEEE International Conference on Multimedia Expo Workshops (ICMEW 2015)*, pages 1–6.

Scholefield, A. and Dragotti, P. L. (2014). Quadtree structured image approximation for denoising and interpolation. *IEEE Transactions on IMage Processing*, 23:1226–1239.

Shan, C. (2012). Learning local binary patterns for gender classification on real-world face images. *Pattern Recognition Letters*, 33(4):431–437.

Shan, C., Gong, S., and McOwan, P. W. (2005). Robust facial expression recognition using local binary patterns. In *ICIP 2005. IEEE International Conference on Image Processing, 2005.* IEEE.

Shan, C., Gong, S., and McOwan, P. W. (2009). Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, 27(6):803–816.

Thomaz, C. E. and Giraldi, G. A. (2010). A new ranking method for principal components analysis and its application to face image analysis. *Image and Vision Computing*, 28:902–913.

Torrisi, A., Farinella, G. M., Puglisi, G., and Battiato, S. (2015). Selecting discriminative clbp patterns for age estimation. In *IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, pages 1–6.

Usó, A. M. (2003). A quadtree-based unsupervised segmentation algorithm for fruit visual inspection. In López, F. J. P., Campilho, A. C., de la Blanca, N. P., and Sanfeliu, A., editors, *IbPRIA*, volume 2652 of *Lecture Notes in Computer Science*, pages 510–517. Springer.

Zhao, W., Chellappa, R., Phillips, P. J., and Rosenfeld, A. (2003). Face recognition: a literature survey. *ACM Computing Surveys*, 35(4):399–458.