Gender Classification in Large Databases

Enrique Ramón-Balmaseda, Javier Lorenzo-Navarro,
and Modesto Castrillón-Santana*

SIANI Universidad de Las Palmas de Gran Canaria Spain
enrique.de101@alu.ulpgc.es, {jlorenzo,mcastrillon}@siani.es

Abstract. In this paper, we address the challenge of gender classification using large databases of images with two goals. The first objective is to evaluate whether the error rate decreases compared to smaller databases. The second goal is to determine if the classifier that provides the best classification rate for one database, improves the classification results for other databases, that is, the cross-database performance.

Keywords: Gender Recognition, Local Binary Pattern, Large Facial Image Databases.

1 Introduction

Children with few years are able to quickly and easily determine the gender of the people who they interact with. For an automatic system, covering a larger population defines a more complex class border, making the process of classification more difficult. Due to the open challenges, gender classification is a current field of research in computer vision, with different application scenarios that include demographics, direct marketing, surveillance and forensics among others.

Automatic gender classification is therefore an active topic as evidenced by recent publications in major journals [2, 10, 11]. Nowadays, state-of-the-art approaches are based on the facial appearance, although some papers considering the analysis of context information are emerging. The interested reader should study the work by Mäkinen et al. [10], a valuable source presenting a comparison of classification results for this problem with automatically detected faces.

Until recently, FERET [12] has been the database mostly used to evaluate different gender classifiers due to the high quality of the images [2, 5, 6]. In this paper, we work with three large databases to evaluate gender classification: MORPH [9], The Images of Groups [7] database and the Labelled Faces in the Wild (LFW) [8] database. The first two databases contain more than 10,000 faces of people and the third one more than 5,000 images. These databases hhave recently attracted researchers due to the challenging variety of identities, ages, ethnicities, poses and the absence of controlled lighting conditions, they are therefore a source for testing algorithms of classification invariant to real world imagery. We will carry out an experimental analysis with normalized face

* Work partially funded by the Spanish Ministry of Science and Innovation funds (TIN 2008-06068), and the Departamento de Informática y Sistemas at ULPGC.

L. Alvarez et al. (Eds.): CIARP 2012, LNCS 7441, pp. 74–81, 2012.

images at different resolutions, and including both the face area and the face area along with the local context. We use the Local Binary Pattern operator (LBP) [13] to extract images features, and Support Vector Machines (SVM) [14] for classification. The paper is organized as follows. In section 2 the LBP descriptor is introduced. In section 3 details about databases, methodology and experimental results and finally, our conclusions in section 4.

2 Local Binary Pattern

LBP [13] is a simple but efficient texture descriptor that labels the pixels of an image by thresholding the neighbourhood of each pixel with the central pixel value considering the result a binary code. Due to its capacity of discrimination and the simplicity of calculation, this texture descriptor has become a popular method that is used in several real world applications. The most important property of this operator is its robustness to monotonic gray-scale changes that may be caused, for example, by variations in lighting.

The original operator LBP, was introduced by Ojala et al. [1]. The operator labels each of the pixels of an image using the 3×3 neighbourhood, comparing each pixel with the central 3×3 window value. The result is considered as a binary number, and a histogram is calculated, and used as a texture descriptor of the image.

The original definition has been extended to a set of arbitrary circular neighbourhoods, and new definitions have been developed. However, the main idea is the same: a binary code that describes the pattern of the local texture is computed thresholding the neighbours by the center pixel gray value. The expression to compute the generalized LBP is the following:

$$LBP_{P,R} = \sum_{p=0}^{p-1} s(g_p - g_c)2^p \tag{1}$$

where, g_c is the gray level of p neighbours $g_p(p = 1, 2, \cdots, p-1)$.The function $s(x)$ is defined as:

$$s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \tag{2}$$

The LBP operator has multiple variants which reflect the attention received by the computer vision community. For the purposes of this work, we focus on one of them, the uniform patterns.

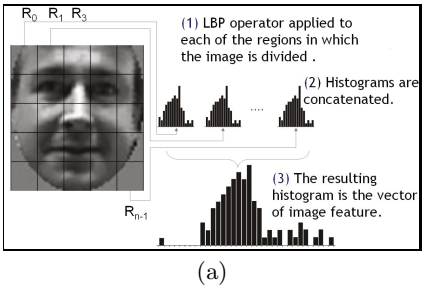
Uniform patterns can be used to reduce the length of the feature vector and to implement a simple descriptor invariant to rotations. This version was inspired by the fact that some binary patterns are more frequent than others. A LBP code is called uniform if the binary pattern contains a maximum of two transitions to bit-level, from 0 to 1 or vice versa, when the bit pattern is circularly travesed. For example, patterns of 00000000 (0 transitions), number 01110000 to (2 transitions) and 11001111 (2 transitions) are uniform, while the patterns of 11001001 (4 transitions) and 01010010 (6 transitions) are not. To get the LBP codification of an

image using uniform patterns, each pixel is assigned the corresponding uniform pattern code, or a unique label for all non-uniform patterns. For example, when it is used (8, R) neighbourhood, there are a total of 256 patterns, 58 of which are uniform and the rest are not uniform, so that there are a total of 59 different labels.

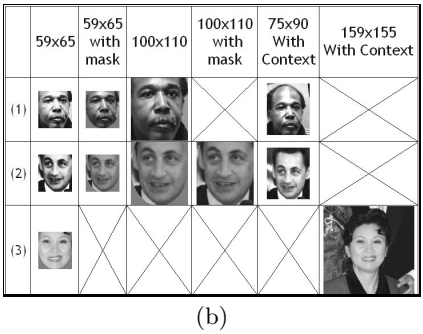
2.1 Face Recognition with LBP

In texture classification, the LBP code occurrences in an image are described in a histogram. The classification is then performed using the simple calculation of histogram similarity. However, for facial recognition this approximation implies the loss of spatial information and, therefore, the location information must be coded somehow into the texture model. One way to achieve this goal is the use of LBP descriptors to construct several local descriptions of the face and combine them into a comprehensive description. Recently, this approach has gained adepts due to the limitations of the holistic representations. These methods based on local features are more robust to variations in the position or the lighting than the holistic methods.

The basic methodology for the extraction of features is proposed in [4]. The algorithm introduces a new approach to facial recognition because it considers not only the face shape but also the texture. In this algorithm the face is divided into small regions where the LBP operator is applied and later concatenated, following a Bag of Words scheme [15], into a single histogram that represents the image of the face. The textures of the facial regions are locally encoded by LBP, while the entire shape of the face is retrieved by the histogram of the face. The underlying idea of using LBP features, is that a face image can be seen as the composition of micro-patterns that are invariant to monotonic transformations of gray-scale. The combination of these micro-patterns generates a comprehensive description of the image of the face, as shown in Figure 1a. This histogram has a description of the face at three different levels:



(a)



(b)

Fig. 1. (a) Face image feature vector computation. The final histogram is obtained by the concatenate of the respective cell histograms (b) Sample images of (1) MORPH (2) LFW and (3) The Image of Groups.

- The labels of the LBP histogram contain information about the patterns at pixel level.
- The labels are summarized in a small region to collect information at the local level.
- Region histograms are concatenated to create a comprehensive description of the face.

3 Datasets, Methodology and Experimental Results

Automatic gender classification would be of interest in scenarios where the number of people would be very high, e.g. direct marketing in a mall. In these scenarios, detected people would rarely be contained in the training set. In general, the works described in the literature make use of small dataset which is not representative of a real environment where the system must deal with thousands of people. Therefore, the challenge in this work is to cope with large databases of images in order to the percentage of correctly classified instances. The objective is therefore to use a large database of images to train, and a different database to evaluate the classification quality. In addition, we can determine if it is possible to keep the performance of state-of -the-art gender classifiers.

3.1 Datasets Description

The databases that have been used are MORPH, LFW (usig only one image per identity) and The Image of Groups. The next table summarizes the databases characteristics Figure 1 (b) shows examples of each database and image resolutions employed.

Table 1. Databases summary. (wm with mask), (wc with context).

Database	Female	Male	Sum	59x65	59x65wm	100x110	100x110wm	75x90wc	159x155wc
MORPH	8.488	46.646	55.134	x	x	x		x	
LFW	1.484	4.261	5.745	x	x	x	x	x	
Image of Groups	14.549	13.671	28.220	x					x

3.2 Methodology

The experimental procedure has been the following:

1. Each sample in the database is divided into a grid of nxn cells, where $n = 3, 5, 7, 9$. A LBP operator is applied to each grid cell: (LBP $\{8, 1\}$ or LBP$^{u2}\{8, 1\}$) and the resulting grid cell histograms are concatenated to obtain a new histogram. The resulting histogram is the corresponding feature vector, as seen in Figure (1).
2. The MORPH database was preprocessed, due to its size and the important unbalance between women and men (15% versus 85%), to randomly generate an additional balanced subset, i.e. with approximately equal number of females and males.

Table 2. Comparative MORPH all images or balanced set of females and males

Instances		Grid		
Learning	Test	3x3	5x5	7x7
14.244	3.560	94,39%	95,42%	96,27%
44.105	11.025	92,13%	94,27%	94,58%

3. A Support Vector Machine classifier with linear kernel is computed for each database.
4. The database that provides the best classification results is selected to analyze other versions of the LBP operator in order to determine which LBP configuration yields better results. If any LBP operator reports better classification results in the selected database, the process is repeated in the rest of the databases to confirm this fact.
5. Once we have the database and the LBP configuration providing the best classification results, the whole database is used to train the classifier and then the resulting classifier is tested with the other two databases. With this last test the aim is at comparing cross database classification results with those results achieved training and testing with the same database.

3.3 Experimental Results

First, the MORPH database was preprocessed as mentioned above to consider the original database or a balanced version of it. In this case we used the standard image size of 59x65 pixels, $\text{LBP}^{u2}\{8,1\}$ and SVM classifier, with the setup described in the preceding section.

The best results are obtained with the balanced subset with (8.488 females and 9,326 males), that reported an accuracy of 96,27% versus 94,58% with the original unbalanced database. In the rest of the work we make use only this balanced subset for this database. To build the balanced set, all female samples are used, while male samples are selected randomly. Identity intersection of the sets of training and test, in the worst cases, it is of one or two instances, and then we can see that the set of training examples and test are disjoint.

Then we apply the operator LBP^{u2} to generate a feature vector of lower dimension for each image, reducing the computational cost. With this experiment we select the database wih best accuracy.

The best results have been obtained with the MORPH database, achieving an accuracy of $97,64\%$ with a 9×9 grid applied on the images with context. The LFW database [6] reaches a $90,83\%$ with the LBP images computed using 120×105 pixels. In this paper, for this database, we reached $90,60\%$ for 75×90 pixel images. This difference is likely due to the lower image resolution, the reduced number of training samples, additionaly the absence of images of the same identity in both training and test sets, and the unbalanced number of female and male samples. In [6] the classification rates achieved for The Image of Group's database is a $86,34\%$. As shown in Table 3, we achieved $82,65\%$ using $\text{LBP}^{u2}\{8,1\}$, which significantly reduces the dimension of the feature vector, and uses normalized images of greater resolution. In [5] we can find the best

Table 3. Results apply LBP$^{u2}\{8,1\}$ to three databases

Database	Size	Instances		Grid			
		Learning	Test	3x3	5x5	7x7	9x9
MORPH	59x65	14.244	3.560	94,39%	95,42%	96,27%	96,18%
	59x65 with mask	14.244	3.560	94,34%	95,51%	96,49%	96,38%
	100x110	14.244	3.560	95,09%	96,49%	97,19%	97,06%
	75x90	14.244	3.560	95,23%	96,07%	97,06%	**97,64%**
LFW	59x65	4.596	1.149	87,12%	84,94%	88,34%	87,64%
	59x65 with mask	4.596	1.149	87,12%	85,64%	87,90%	87,90%
	100x110	4.596	1.149	87,82%	87,82%	87,47%	90,08%
	100x110	4.596	1.149	87,95%	86,95%	88,25%	89,82%
	75x90	4.596	1.149	86,86%	87,82%	89,56%	**90,60%**
The Image of Groups	59x65	22.529	5.632	80,43%	**82,65%**	82,34%	81,53%
	159x155	22.529	5.632	55,38%	54,27%	55,52%	n.a

Table 4. Results of applying different operator LBP to the MORPH database

LBP Operator	Instances		Grid			
	Learning	Test	3x3	5x5	7x7	9x9
$LBP\{8,1\}$	14.244	3.560	93,97%	95,93%	97,05%	**97,53%**
$LBP^{u2}\{8,1\}$	14.244	3.560	95,23%	96,07%	97,06%	**97,64%**
$LBP\{8,2\}$	14.244	3.560	94,28%	97,03%	97,22%	**97,25%**
$LBP^{u2}\{8,2\}$	14.244	3.560	95,32%	96,49%	96,52%	**96,69%**
$LBP\{4,1\}$	14.244	3.560	92,54%	95,63%	96,55%	**96,78%**
$LBP^{u2}\{4,1\}$	14.244	3.560	92,60%	95,81%	96,24%	**96,80%**
$LBP\{4,1\}$ concatenated with global histogram	14.244	3.560	92,74%	95,65%	96,63%	**96,75%**
$LBP^{u2}\{4,1\}$ concatenated with global histogram	14.244	3.560	92,65%	96,68%	96,16%	**96,77%**

Table 5. Cross-database results using MORPH database like training set and the other two databases like test set

Database test	Type-Size	LBP operator	Instances		Grid			
			Learning	Test	3x3	5x5	7x7	9x9
LFW	59x65	$LBP\{8,1\}$	14.244	1.149	59,23%	66,33%	69,68%	61,28%
		$LBP^{u2}\{8,1\}$	14.244	1.149	57,82%	**75,10%**	63,51%	50,47%
	59x65 with mask	$LBP\{8,1\}$	14.244	1.149	67,36%	72,16%	68,42%	61,28%
		$LBP^{u2}\{8,1\}$	14.244	1.149	67,53%	71,90%	69,00%	69,27%
The Image of Group	59x65	$LBP\{8,1\}$	14.244	5.632	65,54%	**76,74%**	68,15%	68,33%
		$LBP^{u2}\{8,1\}$	14.244	5.632	64,53%	53,37%	53,87%	52,80%
	159x155	$LBP\{8,1\}$	14.244	5.632	46,60%	51,23%	52,79%	51,99%
		$LBP^{u2}\{8,1\}$	14.244	5.632	53,61%	53,45%	53,25%	52,15%

results obtained with the database MORPH of 88% of accuracy in the gender classification. We have reached $97,64\%$, using a set of images with greater size for both training and test.

Other LBP operators have been applied to the MORPH database in order to test if better results can be obtained. In particular, we have used LPB$\{8,2\}$, LBP$^{u2}\{8,2\}$,the uniform, LBP$\{4,1\}$, LBP$^{u2}\{4,1\}$ and LBP$\{4,1\}$ plus its corresponding while image histogram. The best results, see Table 4, are obtained with a different operator the uniform $LPB\{8,1\}$, so the previous experiment was repeated to see if it improved with other databases.

The best results are achieved with the operator LBP$\{8,1\}$ and LBP$^{u2}\{8,1\}$, there is no additional tests in other databases.

Finally, we proceeded with the cross-database test, i.e. the training is performed with the set of images from a database, and evaluated with the data of other database to check the "quality" of the classifier. Note that the best results we got with images that contains part of the context is similar. However, the dimension of this type of images is not the same in the other databases, so we used the images without context for both training and test.

As we can see in Table 5 it does not exist any accuracy that improves the results presented in Table 3, the classifier performing the best for one databse is not able to do so with other databases.

4 Conclusions

We have achieved a very high recognition rate with the MORPH database, reaching $97,64\%$ accuracy. With the other databases under consideration, similar classification rates were not achieved, so the success of the method is not database independent. Indeed characteristics such as dimensions of the images, lighting, normalization procedure, total number of images, affect among others.

In addition, using the database with the best results as training set results, there has not been improvement of the accuracy rate of classification in the other databases, which confirms the assertion that there is a high dependency on the database.

Therefore, there are still open challenges in gender classification based on the facial pattern. The results achieved suggest that the problem is not completely solved and these results demonstrate that there are still areas where improvement must be done with new methods and procedures to get a database independent classifier. Thus a system trained under certain conditions may be used with images obtained and completely different from the training set and obtain similar classification rates.

References

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

[2] Bekios-Calfa, J., Buenaposada, J.M., Baumela, L.: Revisiting linear discriminant techniques in gender recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(4), 858–864 (2011)

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

[4] Ahonen, T., Hadid, A., Pietikäinen, M.: Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(12) (December 2006)

[5] Chu, W.-S., Huang, C.-R., Chen, C.-S.: Identifying gender from unaligned facial images by set classification. In: International Conference on Pattern Recognition (ICPR), Istanbul, Turkey (2010)

[6] Dago-Casas, P., González-Jiménez, D., Long-Yu, L., Alba-Castro, J.L.: Single and cross database benchmarks for gender classification under unconstrained settings. In: Proc. First IEEE International Workshop on Benchmarking Facial Image Analysis Technologies (2011)
[7] Gallagher, A., Chen, T.: Understanding images of groups of people. In: Proc. CVPR (2009)
[8] Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07–49, University of Massachusetts (October 2007)
[9] Ricanek Jr., K., Tesafaye, T.: MORPH: A longitudinal image database of normal adult age-progression. In: IEEE 7th International Conference on Automatic Face and Gesture Recognition, Southampton, UK, pp. 341–345 (April 2006)
[10] Mäkinen, E., Raisamo, R.: Evaluation of gender classification methods with automatically detected and aligned faces. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(3), 541–547 (2008)
[11] Moghaddam, B., Yang, M.-H.: Learning gender with support faces. IEEE Trans. on Pattern Analysis and Machine Intelligence 24(5), 707–711 (2002)
[12] Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.J.: The FERET database and evaluation procedure for facerecognition algorithms. Image and Vision Computing 16(5), 295–306 (1998)
[13] Pietikäinen, M., Hadid, A., Zhao, G., Ahonen, T.: Computer Vision Using Local Binary Patterns. Springer (2011)
[14] Burges, C.J.C.: A tutorial on support vector machines for pattern recognition (1998)
[15] Csurka, G., Dance, C.R., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: Workshop on Statistical Learning in Computer Vision, ECCV, pp. 1–22 (2004)