Inhibition of Occluded Facial Regions for Distance-Based Face Recognition

Daniel López-Sánchez[1], Juan M. Corchado[1,2], Angélica González Arrieta[1]
[1] University of Salamanca, BISITE Research Group
[2] Osaka Institute of Technology
{lope, corchado, angelica}@usal.es

Abstract

This work focuses on the design and validation of a CBR system for efficient face recognition under partial occlusion conditions. The proposed CBR system is based on a classical distance-based classification method, modified to increase its robustness to partial occlusion. This is achieved by using a novel dissimilarity function which discards features coming from occluded facial regions. In addition, we explore the integration of an efficient dimensionality reduction method into the proposed framework to reduce computational cost. We present experimental results showing that the proposed CBR system outperforms classical methods of similar computational requirements in the task of face recognition under partial occlusion.

1 Introduction

This work focuses on the design and implementation of a Case-Based Reasoning (CBR) system for efficient face recognition, with a special focus on robust face recognition under partial occlusion conditions[1]. Although the problem of face recognition has been extensively addressed in the available literature, most state-ot-the-art proposals impose a series of constraints that limit their applicability in real world scenarios where only a limited amount of computational power and training information is available. The CBR method proposed in this paper seeks to cover the full recognition process (i.e., face detection, normalization, and identity prediction). In addition, we focused on methods which are able to work under the constraints of low computational power and little training information. As opposed to other occlusion-robust face recognition systems, the proposed CBR framework does not make any assumption about the nature of occlusion that it will have to face at test time. We also studied the possible integration of an efficient dimensionality reduction method in the proposed framework to reduce the computational cost. The experimental results presented in this paper show that the proposed method outperforms traditional face recognition methods in the task of partially occluded face recognition.

---

[1]This paper is an abridged version of a paper titled "A CBR System for Efficient Face Recognition Under Partial Occlusion", presented at the ICCBR-2017 conference

2 Related Work

Most occlusion-robust face recognition systems available in the literature include a prior step to identification where they determine which parts of the images are affected by occlusion. In particular, some studies have used manually annotated occluded/non-occluded facial image patches to explicitly train a classifier [Min *et al.*, 2011]. However, this approach has the drawback of requiring occluded face images during the training stage. As a consequence, if the nature of occlusion changes, the accuracy of the occlusion detector might be affected. Using color-based segmentation methods to detect occluded facial regions has also been proposed in the literature [Jia and Martinez, 2008]. However, these methods are very sensitive to lighting conditions and make the unsafe assumption that the occlusion is not caused by artefacts with human-skin color. [Ekenel, 2009] holds the idea that most of the accuracy loss registered by face recognition systems when dealing with partially occluded images is due to alignment errors, rather than information corruption by the occlusion. To address this problem, he proposed a method which seeks to minimize the distance between each sample in the training set and a new observation by evaluating a number of different alignment variations. As a consequence, searching for the best alignment variation requires hundreds of comparisons for each training sample. Although this method achieved notable accuracy rates, the computational cost supposes a major drawback.

Nowadays, the state-of-the-art in one of the most widespread face recognition datasets, namely the LFW dataset, is mostly held by deep neural network models[2]. However, such models are not always applicable due to computational constraints. In scenarios with this limitation, LBP descriptor-based methods remain as one of the best options.

3 Proposed Framework

In this section, we describe in detail both the proposed CBR framework and the selected pre-processing steps needed for face recognition. At test time, when the system is presented with an image that contains a human face in it, the following processing stages are executed: (1) A region of interest is determined for the human face in the image; (2) the detected face is aligned; (3) the image is pose-normalized, rotating
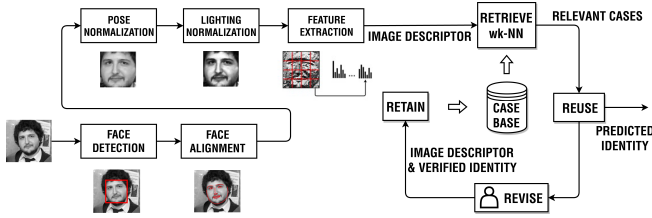
---

[2]http://vis-www.cs.umass.edu/lfw/results.html#UnrestrictedLb

Figure 1: Architecture of the proposed CBR framework and preprocessing steps.

Figure 2: Differential distribution of minimum local distances for occluded and non-occluded blocks in the ARFace database.

and scaling the face to a standard size and orientation; (4) the lighting conditions of the image are normalized; (5) a feature extraction method is applied; and finally (6) the proposed retrieval and reuse stages are executed to emit a prediction regarding the identity of the person in the original image.

3.1 Preprocessing

This section describes the successive preprocessing stages executed before the actual retrieval and reuse stages in the proposed CBR framework.

- **Face detection**. For our experiments, we used the HOG face detector provided by the Dlib C++ library[3].

- **Face alignment**. Our framework uses the face alignment algorithm proposed by V. Kazemi in 2014 [Kazemi and Josephine, 2014]. In particular, for the experiments in this paper we used the pre-trained model provided by Dlib C++[3].

- **Pose normalization**. Previously detected facial keypoints are used to normalize the facial image by rotating and cropping it to display the aligned face in its center, in a vertical pose. In addition, the resulting image is resized to a standard size.

- **Light normalization**. Due to its simplicity, efficiency and good performance, Histogram Equalization (HE) was used to normalize the lighting conditions in all experiments in this paper.

3.2 Feature Extraction: Local Binary Patterns

In this paper, we focus on a specific family of feature descriptors known as *Local Binary Patterns* (LBP) [Ojala *et al.*, 2002]. As described in the following sections, the local nature of this descriptor will allow us to maintain features from occluded regions isolated from those extracted from visible parts of the face. The LBP descriptor labels pixels in an image by considering value differences with their neighbors. This label is treated afterwards as a binary number. The notation $LBP_{P,R}$ is usually used to denote the LBP descriptor with P neighbors and a radio of value R. It has been proved that, using the $LBP_{8,1}$ descriptor, almost 90% of the extracted labels are uniform (i.e., its binary representation contains two transitions at most) [Ojala *et al.*, 2002]. For this reason, a variant of LBP was designed where non-uniform patterns are merged together in a single label. This variant of the descriptor is known as uniform LBP ($LBP_{P,R}^u$). Before training a

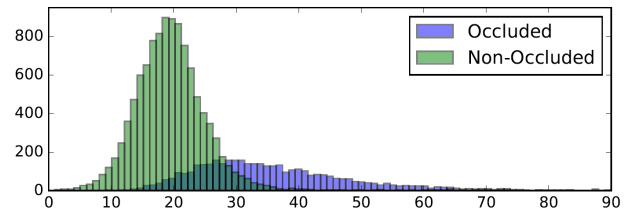classifier, the LBP representation is often refined by dividing the image in a number of blocks (arranged in a grid structure) and counting the number of occurrences of patterns in each block. After this, the corresponding histograms of each block are concatenated to form the final descriptor. This approach in known as *Local binary pattern histograms* (LBPH).

3.3 Identification: Occlusion-Robust Retrieval

First, we introduce a method to detect partial occlusion in LBPH blocks. Conveniently, our method requires no *a priori* knowledge about the nature of occluded blocks. We define the *minimum local distance* for the histogram of an LBP block as the minimum squared Euclidean distance obtained when comparing this histogram with the LBP histograms corresponding to the same facial region in the descriptors stored in the Case-Base. Then, the only assumption made by our method is that minimum local distances of occluded blocks are usually larger than those of unoccluded blocks. To provide insight into the veracity of this assumption, we calculated the distribution of minimum local distances for occluded and unoccluded blocks in the ARFace database[4]; the resulting distributions are shown in figure 2. Although some overlapping exists among the two distributions, it might be possible to define a conservative threshold to discard most occluded blocks. Formally, the Case-Base of our framework is defined as a set of identity label y and LBPH descriptor x pairs:

$$CB = \{(y^{(i)}, x^{(i)}), i = 1, 2, ..., n\} \quad (1)$$

When an unlabeled image I is presented to the system, it is first transformed by the successive preprocessing steps defined in the previous section. Afterwards, the LBPH descriptor $x \in \mathbb{R}^d$ of image I is generated and the retrieval stage begins. Our proposed retrieval stage begins by computing the $n \times d/p$ local distance matrix L, where p is the size of each histogram concatenated to form the LBPH descriptors. Each entry L_{ij} in this matrix corresponds to the local distance between the j-th histogram in x and the j-th histogram in the i-th descriptor in the Case-Base; formally:

$$L_{i,j} = ||(x_{p(j-1)+1}, \cdots, x_{pj}) - (x_{p(j-1)+1}^{(i)}, \cdots, x_{pj}^{(i)})||^2$$
$$\text{for} \quad i = 1, 2, \cdots, n \text{ and } j = 1, 2, \cdots, d/p$$
$$(2)$$

Based on this matrix and the desired *threshold* value for occlusion detection, the retrieve stage computes an occlusion

---

[3] Dlib C++ web page: http://dlib.net

[4] See section 4 for details about the evaluation database.

mask $M \in \{0,1\}^{d/p}$ that determines which of the histograms that conform descriptor x are considered as occluded:

$$M_j = T_h(\, min(\, col_j L \,) \,)$$
$$T_h(x) = \begin{cases} 1 & if \;\; x < threshold \\ 0 & if \;\; x > threshold \end{cases} \qquad (3)$$

Using this occlusion mask, the retrieval stage of the proposed CBR framework finds the k most similar cases to x in the Case-Base, according to the following dissimilarity function:

$$d(x, x^{(i)}) = \sum_{j=1}^{j=d/p} M_j \cdot L_{i,j} \qquad (4)$$

Intuitively, this dissimilarity function corresponds to the squared Euclidean distance between the features in x and $x^{(i)}$ that do not come from occluded facial zones (as predicted in the previous step). In other words, the proposed similarity measure dynamically inhibits the use of corrupted features while retrieving the most relevant cases from the Case-Base.

Afterwards, the reuse stage analyzes the retrieved cases and their labels to emit a prediction regarding the identity associated to the new case. To this extent, we use the weighted voting scheme proposed in [Hechenbichler and Schliep, 2004] with the dissimilarity function of eq. 4.

Conveniently, the computations defined by equations 2 and 3 can be done at the cost of time $\mathcal{O}(n(\frac{d}{p} \cdot p + \frac{d}{p}))$, which simplifies to $\mathcal{O}(nd)$ given that $p > 1$. Finding the most similar cases in the Case-Base according to equation 4 takes $\mathcal{O}(n \cdot \frac{d}{p} + kn)$ time; which can be simplified to $\mathcal{O}(nd)$ by considering hyperparameters p and k as constants. Finally, the remaining computations which correspond to the voting process have a complexity of $\mathcal{O}(k)$. Therefore, the computational complexity of the complete test stage is $\mathcal{O}(nd) + \mathcal{O}(nd) + \mathcal{O}(k)$, which simplifies to $\mathcal{O}(nd)$ given that $n >> k$. Hence, we can conclude that the scalability of the proposed retrieval and reuse stages is equivalent to that of classical nearest neighbour search methods.

In the context of the proposed CBR system, the revision should be carried out by a human expert who determines whether an image has been assigned the correct identity. In particular, retaining revised cases only involves storing their case representation into the Case-Base. In addition, this mechanism can also be applied to provide the CBR system with knowledge of previously unseen individuals, thus extending the number of possible identities predicted by the system.

3.4 Multi-scale Local Binary Pattern Histograms

Several studies have found that higher recognition accuracy rates can achieved by combining LBPH descriptors extracted form the same image at various scales [Chan et al., 2007][Chen et al., 2013]. Unfortunately, the computational costs derived from using such a high-dimensional feature descriptor suppose a serious problem. Apart from that, this image descriptor is perfectly compatible with the proposed method. The only requirement is that histograms corresponding to the same image region are placed next to each other

when forming the final descriptor. Then, selecting the correct value for p, the corresponding histograms for a specific face region will be treated as a single occlusion unit (i.e., a set of features which our method considers as occluded or non-occluded as a whole).

Local Dimensionality Reduction With Random Projection

This subsection tries to address the problem of high-dimensionality of multi-scale LBPH descriptors. In the literature, [Chen et al., 2013] proposed using an efficient dimensionality reduction algorithm to reduce the size of multi-scale LBP descriptors. However, this approach is not directly compatible with the method proposed in the previous section. The reason for this being that we need to keep features from different occlusion units (i.e., facial regions) isolated form each other, so we can later detect and inhibit features coming from occluded facial areas. To overcome this limitation, we propose performing dimensionality reduction at a local level. To this extent, the histograms extracted from a specific facial region (at various levels) are considered as a single occlusion unit. Then, the Random Projection [Achlioptas, 2001] (RP) algorithm is applied locally to each occlusion unit. As opposed to other dimensionality reduction methods, RP generates the projection matrix from a random distribution. As a consequence, the projection matrix is data-independent and cheap to build.

The main theoretical result behind RP is the Johnson-Lindenstrauss (JL) lemma. This result guarantees that a set of points in a high dimensional space can be projected to a Euclidean space of much lower dimension while approximately preserving pairwise distances between points [Dasgupta and Gupta, 2003]. Formally, given $0 < \epsilon < 1$, a matrix X with n samples from $\mathbb{R}^p$, and $k > 4 \cdot ln(n)/(\epsilon^2/2 - \epsilon^3/3)$ a linear function $f : \mathbb{R}^p \to \mathbb{R}^k$ exists such that:

$$(1 - \epsilon)||u - v||^2 \leq ||f(u) - f(v)||^2 \leq (1 + \epsilon)||u - v||^2$$
$$\forall u, v \in X$$
$$(5)$$

In particular, the map $f : \mathbb{R}^p \to \mathbb{R}^k$ can be performed by multiplying data samples by a random projection matrix R whose elements are drawn from $\mathcal{N}(0, 1)$:

$$f(x) = \frac{1}{k} \, xR \qquad (6)$$

As previously said, in order to apply RP in the context of the proposed method, me must first ensure that histograms coming from the same face region are placed contiguously in the final descriptor (see Fig. 3). Afterwards, we can apply the RP method locally to each occlusion unit. Formally, let $x \in \mathbb{R}^d$ be a multi-scale LBPH descriptor where each occlusion unit consists of p features. The reduced version of descriptor x is computed as follows:

$$x' = f(x_{1:p}) \, || \, f(x_{p+1:2p}) \, || \, \cdots \, || \, f(x_{d-p+1:d}) \qquad (7)$$

where $||$ denotes vector concatenation. Note that, thanks to the JL-lemma, for a sufficiently large k value the result of applying the proposed retrieval and reuse stages over reduced descriptors is approximately the same as doing it over the
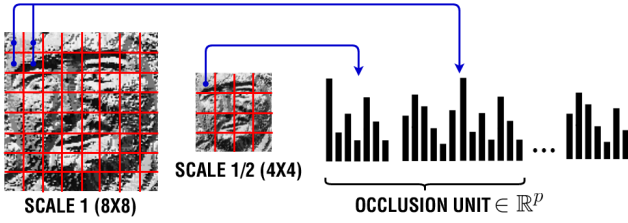
Figure 3: Features are grouped together in the final descriptor according to the facial region they come from.

original high-dimensional descriptors. To prove this, it suffices to consider the different computations carried out by the proposed retrieval and reuse stages. First, the local distance matrix is computed according to eq. 2. If we reduce both the new case x and the descriptors $x^{(i)}$ in the Case-Base as described in eq. 7, and set hyperparameter p to the new size of occlusion units (i.e., $p = k$), equation 2 can be rewritten as follows:

$$L'_{i,j} = ||(x'_{k(j-1)+1}, \cdots, x'_{kj}) - (x'^{(i)}_{k(j-1)+1}, \cdots, x'^{(i)}_{kj})||^2$$
$$\text{for} \quad i = 1, 2, \cdots, n \text{ and } j = 1, 2, \cdots, d'/k \tag{8}$$

where $x' \in \mathbb{R}^{d'}$. Then, applying the JL-lemma, for a sufficiently large k value we can ensure that:

$$(1 - \epsilon)\, L_{i,j} \leq L'_{i,j} \leq (1 + \epsilon)\, L_{i,j}$$
$$\text{for} \quad i = 1, 2, \cdots, n \text{ and } j = 1, 2, \cdots, d'/k \tag{9}$$

In other words, the distortion induced in matrix L' with respect to L is bounded. The following steps of the proposed method are based on L'. Therefore, if the difference between L' and L is small enough, the proposed retrieval and reuse stages will provide the same results when executed over the reduced descriptors. Section 4 reports on several experiments where the descriptors were reduced with this approach.

4 Experimental Results

We evaluated the proposed system over a database of facial images with different types of occlusion and using different image descriptors. In particular, the evaluation dataset is the *ARFace database* [Martinez, 1998]. This dataset contains about 4,000 color images corresponding to 126 individuals (70 men and 56 women). The images display a frontal view of individuals' faces with different facial expressions, illumination conditions and partial occlusions. We used the images in the ARFace dataset to create several subsets for our experiments. In particular, we arranged a training set, a validation set, and several test sets with different characteristics:

- *Training set*: one image per individual (neutral, uniform lighting, first session).
- *Validation set*: almost one image per individual[5] (neutral, uniform lighting, second session).

---

[5]Second session images are not available for all individuals in the dataset.

- *Light test set*: almost four images per individual (neutral, illumination left/right, first and second sessions).
- *Glasses test set*: almost two images per individual (sunglasses, uniform lighting, first and second session).
- *Scarf test set*: almost two images per individual (scarf, uniform lighting, first and second session).

We evaluate the proposed method against other common classification methods used in the field of face recognition, namely Logistic Regression (LR), Support Vector Machine (SVM) and Naive Bayes (NB). In the case of the proposed method, p determines the size of the occlusion unit, and is fully determined by the parametrization of the LBP descriptor. The remaining hyperparameter is the *threshold* for occlusion detection. This hyper-parameter can be selected using a validation set of images without occlusion, even if such validation set contains less than one image per individual. In particular, the *threshold* is set to the minimum possible value which does not produce misclassification of non-occluded blocks in the validation set as occluded (see [López-Sánchez *et al.*, 2017] for more details).

The evaluation protocol for all our experiments has been the following: (1) the classifier under evaluation is trained over the training set; (2) the validation set is used to perform hyperparameter selection; (3) the classifier, parametrized as determined in the previous step, is re-trained over the union of the training set and the validation set; (4) the trained classifier is evaluated over the different test datasets available.

Table 1 presents the results obtained by using the automatic face detection and alignment methods explained in section 3.1. Similarly, Table 2 compiles the results obtained by using the manual face alignment annotations provided by the AR-face database. For single scale descriptors, we used $LBP_{8,2}^u$ histograms over a 8×8 grid, thus obtaining a descriptor of $3,776$ dimensions. In the case of multi-scale descriptors, we used $LBP_{8,2}^u$ histograms over 12×12 and 6×6 grids. The resulting descriptor dimension was therefore $10,620$. Finally, for our experiments with local RP (see Sec. 3.4), each 295-dimensional occlusion unit in the high-dimensional multi-scale descriptor was reduced to 150 features. Therefore, the complete descriptor ended up having a dimension of $5,400$ (i.e., approximately half the original dimension).

5 Discussion and Future Work

This work proposed a novel CBR framework for occlusion-robust face detection. The retrieval and reuse stages of the system use a modified version of the weighted k-Nearest Neighbour [Hechenbichler and Schliep, 2004] algorithm to dynamically inhibit features from occluded face regions. This is achieved by using a novel similarity function which discards local distances imputable to occluded facial regions. As opposed to recent deep learning-based methods, the proposed system can operate in domains where only a small amount of training information is available, and does not require any specialized computing hardware to run. In addition, we proved that the Random Projection algorithm can be applied in a local manner to reduce the dimension of multi-level LBPH descriptors, while ensuring that the proposed retrieval

Features	Classifier	Light	Scarf	Glasses
$LBP^u_{8,2}$	wkNN	81.4%	39.4%	30.0%
$LBP^u_{8,2}$	our CBR p=59, thr=27	96.2%	83.6%	50.2%
$LBP^u_{8,2}$	SVM (poly)	78.1%	36.9%	25.1%
$LBP^u_{8,2}$	LR	84.8%	45.0%	23.4%
$LBP^u_{8,2}$	NB	82.5%	43.7%	20.1%
m.s. $LBP^u_{8,2}$	wkNN	98.8%	73.3%	34.9%
m.s. $LBP^u_{8,2}$	our CBR p=295, thr=17	99.2%	89.2%	50.6%
m.s. $LBP^u_{8,2}$	SVM (poly)	88.1%	59.6%	27.9%
m.s. $LBP^u_{8,2}$	LR	96.2%	75.1%	28.3%
m.s. $LBP^u_{8,2}$	NB	86.29%	72.1%	37.03%
m.s. $LBP^u_{8,2}$ + RP	wkNN	98.5%	66.5%	31.2%
m.s. $LBP^u_{8,2}$ + local RP	our CBR p=150, thr=100	98.8%	90.5%	51.0%
m.s. $LBP^u_{8,2}$ + RP	SVM (poly)	85.5%	55.3%	20.9%
m.s. $LBP^u_{8,2}$ + RP	LR	93.3%	69.0%	25.5%
m.s. $LBP^u_{8,2}$ + RP	NB	84.0%	54.5%	27.1%

Table 1: Experimental results with automatic face alignment.

Features	Classifier	Light	Scarf	Glasses
$LBP^u_{8,2}$	wkNN	95.5%	76.5%	69.5%
$LBP^u_{8,2}$	our CBR p=59, thr=30	99.5%	91.5%	83.5%
$LBP^u_{8,2}$	SVM (poly)	96.5%	75.0%	61.0%
$LBP^u_{8,2}$	LR	98.5%	81.0%	68.0%
$LBP^u_{8,2}$	NB	94.0%	76.5%	69.0%
m.s. $LBP^u_{8,2}$	wkNN	100%	92.0%	86.0%
m.s. $LBP^u_{8,2}$	our CBR p=295, thr=111	99.5%	97.0%	92.0%
m.s. $LBP^u_{8,2}$	SVM (poly)	100.0%	92.0%	84.5%
m.s. $LBP^u_{8,2}$	LR	100.0%	93.0%	89.5%
m.s. $LBP^u_{8,2}$	NB	95.5%	93.5%	89.0%
m.s. $LBP^u_{8,2}$ + RP	wkNN	100%	92.0%	86.0%
m.s. $LBP^u_{8,2}$ + local RP	our CBR p=150, thr=111	99.5%	97.0%	92.0%
m.s. $LBP^u_{8,2}$ + RP	SVM (poly)	100.0%	92.0%	84.5%
m.s. $LBP^u_{8,2}$ + RP	LR	100.0%	93.0%	89.5%
m.s. $LBP^u_{8,2}$ + RP	NB	95.5%	93.5%	89.0%

Table 2: Experimental results with manual face alignment.

and reuse stages will perform approximately as well as they do over the original high-dimensional descriptors. Experimental results carried out over the ARFace database show that, in most cases, the proposed method outperforms classic classification algorithms when using LBPH features to identify facial images with partial occlusion. As for future work, we propose investigating how automatic face alignment methods can be made more robust to partial facial occlusion.

References

[Achlioptas, 2001] Dimitris Achlioptas. Database-friendly random projections. In *Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 274–281. ACM, 2001.

[Chan *et al.*, 2007] Chi-Ho Chan, Josef Kittler, and Kieron Messer. *Multi-scale local binary pattern histograms for face recognition*. Springer, 2007.

[Chen *et al.*, 2013] Dong Chen, Xudong Cao, Fang Wen, and Jian Sun. Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3025–3032, 2013.

[Dasgupta and Gupta, 2003] Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of johnson and lindenstrauss. *Random Structures & Algorithms*, 22(1):60–65, 2003.

[Ekenel, 2009] Hazım Kemal Ekenel. *A robust face recognition algorithm for real-world applications*. PhD thesis, Karlsruhe, Univ., Diss., 2009, 2009.

[Hechenbichler and Schliep, 2004] Klaus Hechenbichler and Klaus Schliep. Weighted k-nearest-neighbor techniques and ordinal classification. Technical report, Discussion paper//Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München, 2004.

[Jia and Martinez, 2008] Hongjun Jia and Aleix M Martinez. Face recognition with occlusions in the training and testing sets. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–6. IEEE, 2008.

[Kazemi and Josephine, 2014] Vahid Kazemi and Sullivan Josephine. One millisecond face alignment with an ensemble of regression trees. In *27th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, United States, 23 June 2014 through 28 June 2014*, pages 1867–1874. IEEE Computer Society, 2014.

[López-Sánchez *et al.*, 2017] Daniel López-Sánchez, Juan M Corchado, and Angélica González Arrieta. A cbr system for efficient face recognition under partial occlusion. In *International Conference on Case-Based Reasoning*, pages 170–184. Springer, 2017.

[Martinez, 1998] Aleix M Martinez. The ar face database. *CVC Technical Report*, 24, 1998.

[Min *et al.*, 2011] Rui Min, Abdenour Hadid, and Jean-Luc Dugelay. Improving the recognition of faces occluded by facial accessories. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 442–447. IEEE, 2011.

[Ojala *et al.*, 2002] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, 2002.