Robust Face Identification In Video

L. ESSANNOUNI[1], E. IBN ELHAJ[1,2], D. ABOUTAJDINE[1]
[1]GSCM Faculté des Sciences de Rabat, Rabat, Maroc
[2]Institut National des postes et télécommunication, Rabat, Maroc

Summary
In this article, we propose a robust person identification algorithm in video that is able to detect and identify faces. The method we have developed combines approaches for detection and recognition using an efficient correlation technique and auto associative neural network. The experiments confirm the robustness of our approach against occlusion and illumination changes.
Key words
Face identification, occlusion, illumination, video.

Introduction

Important practical applications of automatic face recognition have made it a very popular research area in the last three decades, see [1] [2] [3] [4] for surveys. Most of the methods developed deal with *single-shot* recognition. In controlled imaging conditions (lighting, pose and/or occlusions), many have demonstrated good (nearly perfect) recognition results [4]. On the other hand, single-shot face recognition in uncontrolled, or loosely controlled conditions still poses a significant challenge [4]. In this study special attention is paid to occlusions handling and illumination variation.
The objective of this paper is to annotate video with the identities, location within the frame of specific people. This requires both detection and recognition of the individuals. In the detection stage we use a technique of orientation correlation [5]. Each face is modelled by an auto associative memory. For handling occlusions the certainty of each pixel is introduced [6], in fact the network has an auto associative memory with backward connection. This auto associative network is used to estimate the original pixel values of the input images and to replace the input with the estimated values when the differences are large.
It is possible to deal with the illumination at three different stages: during the pre-processing, the feature extraction or the classification [7][8][9].
The proposed system is robust to illumination variations , in fact in the detection stage, the method of orientation correlation is illumination invariant as the matching is performed on orientation of image intensity gradient [5]. In the recognition stage, illumination changes can be easily compensated by locally normalizing the image [10].
The paper is organized as follows: the proposed approach for recognition in video is described in section 2, the experiments and results are presented in section 3. Finally, section 4 concludes this paper.

2. Proposed approach

The overall identification approach consists of two steps: (i) modelling faces of database and (ii) recognition.

2.1 Neural network for modelling

This method is based on an auto associative neural network. We use an auto associative memory for modelling the faces of the database Figure 1. Each face is modelled by a network.

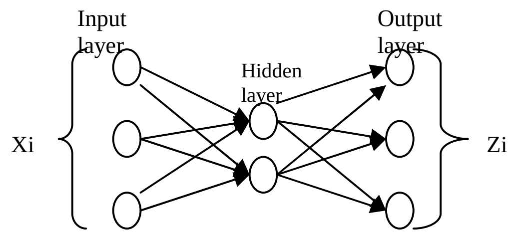


FIG.1. Auto associative memory.

The auto associative memory is implemented by using a Multi-Layer Perceptron (MLP) with the same number of the input and output units. The image of the output have to be similar to the image of input.
Let the training data $x_j = (x_{j1},..., x_{jM})^T \in R^M$. Denote the output of the hidden units and the output of the output units for the input vector as $y_j = (y_{j1},..., y_{jH})^T \in R^H$ and $z_j = (z_{j1},..., z_{jN})^T \in R^M$. Also the weights from the input to the hidden and from the hidden to the output are

denoted as $U = [u_1,...,u_H]$ and $W = [w_1,...,w_M]$. Then the network computes the outputs as :

$$y_{jh} = u_h^T x_j \quad \text{(h = 1,…,H).} \qquad \textbf{(1)}$$

$$z_{jM} = w_M^T y_j \ \text{(m = 1,…,M).}$$

The network can be trained by minimizing a sum-of squares error

$$\varepsilon^2 = \frac{1}{2}\sum_{j=1}^{N}\varepsilon_j^2 = \frac{1}{2}\sum_{j=1}^{N}\left\|x_j - z_j\right\|^2 . \qquad \textbf{(2)}$$

From the viewpoints of the maximum likelihood estimation, this is equivalent to considering a probabilistic model of errors ε_j^2 as Gaussian with zero mean. Here we use the evaluation criterion:

$$l = -\varepsilon^2 = -\frac{1}{2}\sum_{j=1}^{N}\varepsilon_j^2 . \qquad \textbf{(3)}$$

By taking the partial derivatives of this evaluation criterion, we can obtain the learning rule as

$$\Delta w_{mh} = \alpha\sum_{j=1}^{N}(x_{jm} - z_{jm})y_{jh} . \qquad \textbf{(4)}$$

$$\Delta u_{hn} = \alpha\sum_{j=1}^{N}\sum_{m=1}^{M}\ (x_{jm} - z_{jm})w_{mh}x_{jn,} .$$

Where α is the learning rate.

2.2 Recognition

The proposed recognition system consists in detecting and identification of faces. This system includes on four main steps:

Selection candidate face region: the image of face is searched in a frame of sequence using orientation correlation. This method is initially proposed for translational image registration. We have extended this method for template matching. Orientation correlation works by correlating orientation images. Each *pixel* in orientation image is a complex number that represents the orientation of intensity gradient. This representation is invariant to illumination changes. Angles of gradient orientation are matched. Andrews robust kernel [11] function is applied to angle differences. Through the use of correlation the method is exhaustive. The method is fast as the correlation can be computed using Fast Fourier Transforms.

Let be an image *f* of sequence and g the image of face, indexed using *(x, y)*, let f_d, g_d the orientation of images *f* and *g*.

$$f_d(x,y) = \operatorname{sgn}\left(\frac{\partial f(x,y)}{\partial x} + i\,\frac{\partial f(x,y)}{\partial y}\right) \qquad \textbf{(5)}$$

$$where\ \operatorname{sgn}(x) = \begin{cases} 0 & if\ \ x = 0 \\ \dfrac{x}{|x|} & otherwise \end{cases}$$

g_d is constructed in the same fashion as f_d.

Orientation images are matched using correlation. Let $F_D(k,l)$ and $G_D(k,l)$ the FFT respectively of f_d and g_d. and IFFT() the Inverse Fast Fourier Transform function, the orientation correlation matching surface is

$$R\{IFFT\left(F_D(k,l)G_D^*(k,l)\right)\ \} \qquad \textbf{(6)}$$

Normalisation : Each sub image is normalised and transformed on vector with size 625 pixel. Illumination changes are compensated by locally normalizing the image [10].

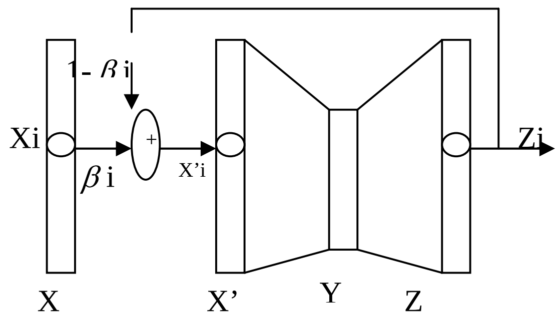


Fig. 2 Modified auto associative memory

Reconstruction:

We use the modified auto associative memory for improving the robustness of the auto associative memory to occlusions. In fact a certainty measure of each pixel is introduced by evaluating the difference between the pixel values of the input image and the retrieved image Figure 2. The pixel values are modified by using the certainty of each pixel and this process of the retrieval and the modification is repeated several times. The concrete process is summarized as follows;

- **STEP 0:** Initialize the iteration parameter *t* as *t* = 0 and assign the input image *x* to the input vector of the auto associative memory *x'*(0).
- **STEP 1:** Recall the output *z*(*t*) of the auto-associative memory from the input *x'*(*t*).

– **STEP 2:** Compute the pixel-wise differences $\varepsilon_i^{\,2}(t) = (x_i(t) - z_i(t))^2$ between the input image *x* and the retrieved image *z*(*t*). Compute the "certainty" $\beta_i(t)$ of each pixel by using these differences as

$$\beta_i(t) = \exp(-\frac{\varepsilon_i^2(t)}{2\sigma^2(t)}), \qquad (7)$$

Where $\sigma(t)$ is the robust estimation of the standard deviations of the differences $\varepsilon_i(t)$ and is obtained by

$$\sigma(t) = 1.4826(1 + \frac{5}{N-1}) \underset{i}{med} \sqrt{\varepsilon_i^2(t).} \qquad (8)$$

Here med(*x*) denotes the median of the *x*.

– **STEP 3:** Compute the new input *x'*(*t* + 1) of the auto associative memory by using the "certainty" of each pixel as

$$x_i'(t+1) = \beta_i(t)x_i(t) + (1 - \beta_i(t))z_i(t). \qquad (9)$$

Set the iteration parameter as $t \leftarrow t+1$ and go to STEP1 to repeat the modification process until the number of iterations is less than the specified value.

Comparison: For classification we compute the difference between the reconstructed sub image and the original one, the difference must be less than the fixed threshold.

3. Results

The image of face is modelled by training the auto associative network, the number of neurons in hidden layer is 17.
Simulations are performed to evaluate the performance of the proposed system. We use some frames which are part from the standard sequences Mother-daughter and Carphone with QCIF format. Each face of database is modelled using the auto associative memory. For recognition we give in input, a frame of video sequence and a face, we search the regions where appears the image of face in frame, using the correlation orientation technique. To allow occlusion in the detection stage, the correlation threshold had to be set at lower values. In all experiments the correlation threshold was set to 0.2. this choice derived empirically from the experiments as shown in figure 3 . We return the position of face in the whole image. The sub images extracted are normalized and reconstructed by the modified auto associative memory. Finally we compare the reconstructed image with the original face, the difference value determines the presence of face.
The system detects and identifies correctly faces, note the absence of false positive due to the comparison process as shown in Figure 6.
The Table1 summarizes the recognition results performed on the Mother-daughter sequence, varying the comparison threshold and the correlation threshold fixed at 0.2
.

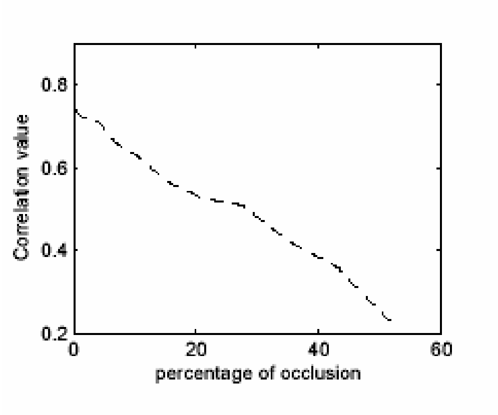


Fig.3 Relation between detection threshold and percentage occlusion

(a) Face

(b) 5%occluded face (c) 20%occluded face

Fig .4 Examples of rectangular occlusion in frame

To confirm the robustness of our approach against occlusion, we performed experiment on the sequence Mother-Daugther. We applied partial occlusion of 0-60% in the face region shown in Figure 4(a). In order to construct the occlusion area, we selected white occluding squared windows within the face region, and computed the window size related to the total number of pixels attributed to a given face. This assures that the face is actually occluded according to a given occlusion rate (some examples of occluded face image are shown in figure 4(b)-(c)).

We apply our algorithm for the recognition of the face shown in Figure 4(a). Figure 5 shows the mean absolute difference (MAD) between the reconstructed detected face and the original face. It is clear from this figure that the results given with the simple auto-associative memory (SAM) are improved by the modified auto associative memory (MAM). Since the (MAD) comparison threshold is set to 5. The system can recognize the face even if it is 20-30% occluded.

Figure 7 shows the results of our method on some frames of carphone video sequence having some artificial occlusions.

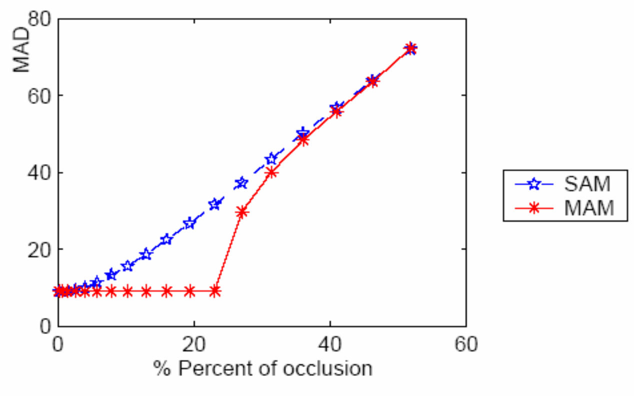


Fig. 5 The MAD between the reconstructed face and the original one using the auto-associative memory and the modified one.

The figure 8 illustrates the application of our approach on noised Mother-daughter sequence. The system still detects and identifies correctly the face.

Finally we test our system in case of illumination variations conditions. The system developed gives very reliable results in case of illumination changes. Figure 9 shows some images taken from Carphone video sequence under illumination variations.

The experiments are run on 2.8 GHz-Pentium 4 processor with 256 MHz memory. The processing time to detect and identify face region from an input frame is about 0.4 sec.

4. Conclusion

In this paper we have proposed a new identification system for video sequences. This system uses auto associative memory for modelling faces, this auto associative memory is efficient to the recognition of occluded and noised images. On the other hand the orientation correlation technique used for detection is fast, exhaustive, statistically robust and invariant to illumination. Experiments have confirmed the robustness of the system against occlusions noise and illumination variations.

References

[1]. W. A. Barrett. A survey of face recognition algorithms and testing results. *Systems and Computers*, *1*, 1998, 301–305.

[2]. Wen-Yi Zhao, Rama Chellappa, P.J. Jonathon Phillips, & Azriel Rosenfeld, Face Recognition: A Literature Survey, *ACM Computing Survey*, *35(4)*, 2003, 399-458.

[3]. T. Fromherz, P. Stucki, & M. Bichsel. A survey of face recognition. *MML Technical Report.*, *(97.01*), 1997.

[4]. W. Zhao, R. Chellappa, A. Rosenfeld, & P. J. Phillips. Face recognition: *A literature survey*. *UMD CFAR Technical Report CAR-TR-948*, 2000.

[5]. A.J. Fitch, A. Kadyrov,W.J. Christmas, J. Kittler , Orientation correlation, British Machine Vision Conference, 2002.

[6]. T. Kurita, T. Takahashi, & Y. Ikeda., Neural network classifier for occluded images, Proceedings of international conference on Pattern Recognition, vol III, 2002, 45-48.

[7]. Y. Adini, Y. Moses & S. Ullman, "Face recognition: the problem of compensating for changes in illumination direction," IEEE Trans. on PAMI, 19(7)., 1997, 721–732.

[8]. M. Savvides, & V. Kumar, "Illumination normalization using logarithm transforms for face authentication," 4th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA), Springer, 2003.

[9]. R. Gross, &V. Brajovic, "An image preprocessing algorithm for illumination invariant face recognition," 4th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA), Springer, 2003.

[10]. F. Jurie, & M. Dhome, Real Time Robust Template Matching, British Machine Vision Conference, 2002

[11]. Peter J. Huber. Robust Statistics. John Wiley & Sons, 1981.

(a)Frame57 (b)Frame 58 (c)Frame59

(i)

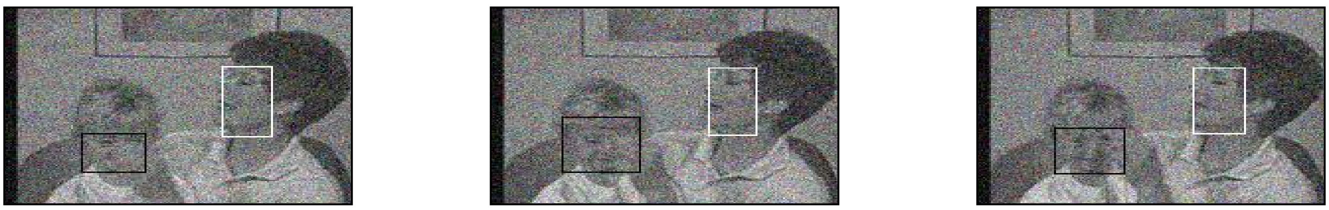
(a)Frame 57 (b)Frame 58 (c)Frame 59

(ii)

Fig.6 (i) Recognition of face using orientation correlation. (ii)Recognition of face using orientation correlation and neural network.

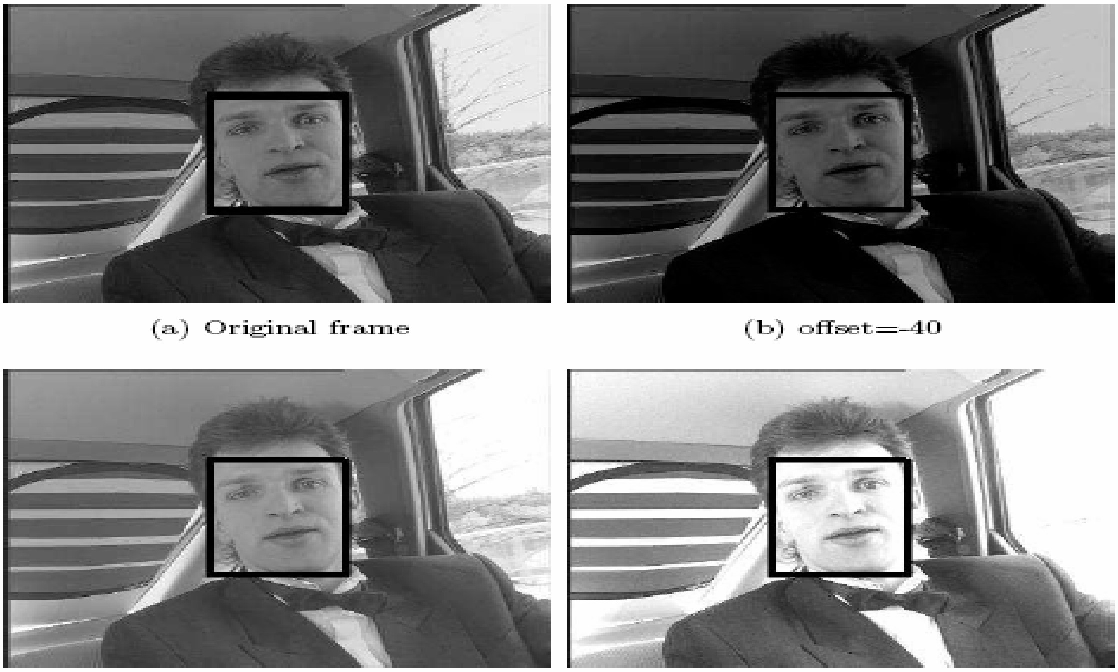
(a)Frame 1 (b)Frame 2 (c)Frame 3

Fig.7 Recognition of the face in presence of occlusions.



(a)Frame 60 (b)Frame 61 (c)Frame 62

Fig.8 Recognition of faces in moving noised sequence.



(a) Original frame (b) offset=-40

(c) offset=30 (d) scale =2

Fig.9 Recognition of face with illumination changes.