Blurred Face Recognition via a Hybrid Network Architecture

Abstract

We introduce a hybrid recognition/reconstruction architecture that is suitable for recognition of images degraded by various forms of blur. This architecture includes an ensemble of feed-forward networks each of which is constrained to reconstruct the inputs in addition to performing classification. The strength of the constraints is controlled by a regularization parameter. Networks are trained on original as well as Gaussian-blurred images, so as to achieve higher robustness to different blur operators.

Face recognition is used to demonstrate the proposed method and results are compared to those of classical unconstrained feed-forward architectures. In addition, the effect of state-of-the-art restoration methods is demonstrated and it is shown that image restoration with the proposed hybrid architecture leads to the best and most robust results under various forms of blur.

1. Introduction

Image blur turns out to degrade recognition more than noise [11, 2, 17, 9]. Approaches to address recognition of blurred images can be divided into three groups: *implicit, restoration and direct*. Under the *implicit* approach, blur is not addressed during training and blurred images are tested as other degraded images. [23, 3, 17]. Under the *restoration* approach, blurred images are restored before recognition [22, 12, 9]. The success of this approach is not obvious since image restoration is an ill-posed inverse problem [4, 6, 18] and restored images contain artifacts. Under the *direct* recognition approach, blur is addressed explicitly in the recognition model [2, 13, 17]. In particular, symmetric blur was studied [2] and poor quality character images [13]. In this paper, we propose a combined approach to address blurred face image recognition via an ensemble of recently introduced [17] hybrid recognition-reconstruction network ensembles which are trained on original and blurred images.

2. Methodology

2.1. Recognition/reconstruction networks and their ensembles

Figure 1 presents the hybrid recognition/reconstruction network architecture. This network attempts to improve the low dimensional representation by minimizing concurrently the mean squared error (MSE) of reconstruction and classification outputs. The relative

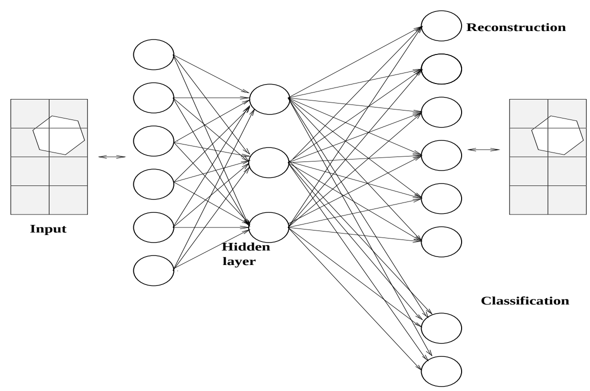

Figure 1. A single hidden layer drives both the classification output layer and the reconstruction.

influence of each of the output errors is defined by a trade-off parameter λ that is unknown *a-priori*.

Under a Bayesian formulation this parameter is a hyper-parameter, therefore, one can integrate the network predictions over it's posterior distribution [10, 16]. In the same way integrating over the *posterior* distribution of the weights can be considered [5]. A rough approximation of such integration leads to combining of suboptimal neural networks to regression ensembles which classify by the Bayesian rule from an average over all ensemble members' outputs. We consider three types of regression ensembles: **U** - *unconstrained ensemble*, corresponding to integration over posterior weights and with fixed $\lambda = 0$; **R** - *reconstruction ensemble*, including networks with different λ parameter and with learning started from fixed initial weights; **J** - *joined ensemble*, combining networks with different λ parameters and different initial weights.

2.2. Training with blurred images

Recognition of blurred images requires a substantial amount of training data processed by different blur operators. Unfortunately, such data is not available, and to solve the problem, *a-priori* information about possible degradation transformations is imposed by adding extra Gaussian blurred images (with standard deviation $\sigma = 2$) to the training set. The network is trained only on original images (method A) or on the additional blurred images while constrained to reconstruct the non-blurred images (method B). This biases the hidden units to become insensitive to *various* blur operation and thus improves the results of the classification system[1]. Three types of ensembles are studied for each of methods A,B: **U**, **R** and **J**. The number of networks in the unconstrained ensemble (**U**) of methods A,B is equal to 6. Reconstruction ensembles (**R**) of methods A,B have been composed from networks with the trade-off parameter λ, with values from [0, 0.3] with steps of 0.05. Ensemble performances are tested on images degraded by different blur operations in conjunction with additive noise and also on restored images.

3. Image degradation

Degradation process is often modeled as a space-invariant blurring with a convolution operator h and corruption with additive noise n : $\mathbf{g} = \mathbf{h} * \mathbf{f} + \mathbf{n}$. Major causes of blurring are misfocus, camera jitter, object motion and atmospheric turbulence. They lead to a low pass operation. Of particular interest is blur with band-pass filter, that is a difference of Gaussians (DOG) filter. A third family of image filters is the high pass filter which leads to image sharpening. This is common in medical imaging, industrial inspection and military applications.

Noise results from image sampling, recording, transmission, etc. In this paper we consider additive *Gaussian white noise* limiting it to be independent on each pixel, with zero mean and some variance σ. Examples of images with different degradations are presented in Figure 2 and main filters that have been used in Table 1. For all filters their point spread functions in polar coordinate systems are given. We also use a root

[1] An alternative training with reconstruction of blurred images to their copy led to slightly inferior results [16] and is not addressed here

Filter type	Blur parameters	Filter - h
Gaussian blur	σ - standard deviation	$h(r,\phi) = C\sigma^2 \exp(-r^2/2\sigma^2)$
Out-of-focus blur	d - blur propagation	$h(r,\phi) = 1/\pi d^2,\ if\ \lvert r\rvert \le d$ 0, otherwise
DOG - difference of Gaussians	$\sigma_1 < \sigma_2$ - standard deviations	$h(r,\phi) = C\sigma_1^2 \exp(-r^2/2\sigma_1^2)-$ $-C\sigma_2^2 \exp(-r^2/2\sigma_2^2)$
Motion blur	d - propagation θ - direction	$h(r,\phi) = 1/d,\ \ if\ r \le d\ \&\ \phi = \theta$ 0, otherwise

Table 1. Filter types used for image degradation.

filter to enhance images. This nonlinear filter acts on image amplitudes in Fourier domain: $\hat{A} = A^{\alpha},\ \alpha < 1$.

4. Image restoration

Image restoration refers to the problem of recovering an image from its blurred and noisy version using some a priori knowledge of the degradation phenomenon and the image nature. It is well-known that restoration problem is an ill-posed problem [4, 6, 18], i.e. a small noise in the observed image results in an unbounded perturbation in the solution. This instability is often addressed by a regularization approach [20, 7, 15, 14, 21] that includes restricting the set of admissible solutions and introducing some a priori knowledge about the image and the degradation model. In the following sections, we briefly review the restoration techniques we have used in this work.

4.1. MSE minimization and regularization

Assuming the blur operator H is known, a natural criterion for estimating an original pixel image f from an observed pixel image g in the absence of any knowledge about noise, is to minimize the difference between the observed image and a blurred version of the restored image:

$$\min_{\mathbf{f}} \mathbf{M}(\mathbf{f}) = \min_{\mathbf{f}} \parallel \mathbf{g} - \mathbf{H}\mathbf{f} \parallel^2. \qquad (1)$$

Often, gradient or conjugate gradient descent methods are used for $\mathbf{M}(\mathbf{f})$ minimization [7, 15]. An application of the gradient method to the minimization problem (1) produces the following iterative method:

$$\mathbf{f_{k+1}} = \mathbf{f_k} + \beta(\mathbf{H^t g} - \mathbf{H^t H f_k}). \qquad (2)$$

When the blur matrix H is nonsingular and β is sufficiently small, the iterative method converges to the $\hat{f} = H^{-1}g$. This solution is known as the *inverse filter* method. In the frequency domain, it corresponds to the following estimation of the ideal image frequency response: $\hat{F}(u,v) = G(u,v)/H(u,v)$. Motion or defocusing blur leads to a singular matrix $\mathbf{H}$. In this case, the above optimization method yields an iterative method that converges to a least square solution $f^+ = H^+ g$ of (1) [7, 6], where $\mathbf{H}^+$ is the generalized inverse of matrix $\mathbf{H}$. In the presence of noise, the iterative algorithm converges to $f^+ + H^+ n$ and thus contains noise filtered by the pseudo-inverse matrix. Often, $\mathbf{H}$ is a low-pass filter, therefore, the noise is amplified and the obtained solution may be very far from the desired one.

To overcome this sensitivity to noise, some a priori information about the noise or the ideal image is often introduced as a quantitative constraint that replaces an ill-posed problem by a well-posed one. This regularization method [20, 15] has a functional minimization form: $\mathbf{L}(\mathbf{f}) = \| \mathbf{Hf} - \mathbf{g} \|^2 + \alpha \| \mathbf{Cf} \|^2$, where the regularization operator $\mathbf{C}$ is chosen to suppress the energy of the restored image in the high frequencies. This is equivalent to smoothing in the spatial domain. Since the $\mathbf{H}$ filter is often a low pass filter, the regularization operator $\mathbf{C}$ is taken to be a Laplacian in order to recover the smooth original image. A regularization parameter α may be known a priori or estimated, but theoretically it is inversely proportional to the signal to noise ratio (SNR).

Although regularization of the MSE criterion with smoothness constraint $\| \mathbf{Cf} \|$ is the basis for most of the work in image restoration, it often leads to unacceptable ringing artifacts around sharp intensity transitions. This effect is due to image blurring around lines and edges. The ways to address this problem include adaptive image regularization [7] and considering total variation (TV) regularization [14]. These methods are generalized to both unknown filter and image in [21, 1] and refered to as blind deconvolution.

5. Data set description and network details

We have used a data-set locally collected by the Tel-Aviv University Computer Vision Group [19]. The data-set contains images of 37 male and female faces with 10 pictures for each person in high resolution 84×56. We split the data to 6 training images and 4 testing images for each person and used a normalization preprocessing. This preprocessing partially removes the variability due to viewpoint, by setting (automatically) the eyes to the same position in all images [19]. Further preprocessing evaluates the difference between each image and an average over all the training set, leading to the so called "caricature" images [8].

The number of hidden units was set to 10 and a learning rate constant was set to 0.01. The initial weights were generated from uniform distribution in the interval [0 0.001]. The number of epochs was about 5000–10000.

6. Results

Table 2 presents classification errors (in percent) of degraded and restored images. Figure 2 shows corresponding degradation and restoration results on one image from TAU data set. Results of other restoration methods and other types of noise appear in [16].

Types of corruption	Training scheme A			Training scheme B		
	U	**R**	**J**	**U**	**R**	**J**
(a) "Clean data"	12.8	12.8	13.5	9.5	10.8	8.8
(b) Out-of-focus blur: $d = 5$	20.9	20.9	17.6	16.2	10.8	11.5
(c) DOG filter: $\sigma_1 = 1$ & $\sigma_2 = 2$	31.8	26.4	23.6	23.0	26.4	20.9
(d) Root filter: $\alpha = 0.6$	16.9	17.6	12.8	10.1	10.8	8.1
(e) Gaussian blur: $\sigma = 2$	19.6	16.2	16.9	14.2	11.5	10.8
Root filter on **(e)**: **(f)** $\alpha = 0.6$:	12.8	13.5	12.8	14.9	12.8	10.8
$\alpha = 0.8$:	13.5	14.2	14.2	12.2	12.2	8.8
(g) Pruned 7×7 pixels Gaussian filter & Gaussian noise, SNR=100	18.9	16.2	18.2	10.8	10.1	10.8
(h) Blind deconvolution of **(g)**	12.8	14.2	12.8	10.1	10.1	8.8
(i) Motion blur: $d = 5$ pixels ($\theta = 0$) & Gaussian noise with SNR=10	21.6	23.6	19.6	16.9	15.5	12.2
(j) smoothing & blind deconvolution of **(i)** [21]	14.9	14.9	14.2	9.5	10.8	9.5
(k) Motion blur $d = 7$ pixels ($\theta = 0$) SNR=inf (no noise)	27.0	29.1	23.6	20.3	23.0	16.2
(l) Blind deconvolution of **(k)** [21]	13.5	15.5	12.8	10.8	11.5	9.5

Table 2. Ensembles of classification schemes (A,B) versus blur and its restoration.

We note that ensembles of method B are less sensitive to noise and blur than ensembles of method A and the joined ensemble of method B has the best classification performance. Restoration preprocessing further improves recognition.

7. Summary

We have proposed a hybrid recognition/reconstruction network architecture for addressing object recognition in blurred-images. This approach includes the following methods: (i) Expansion of the training data with Gaussian blurred images

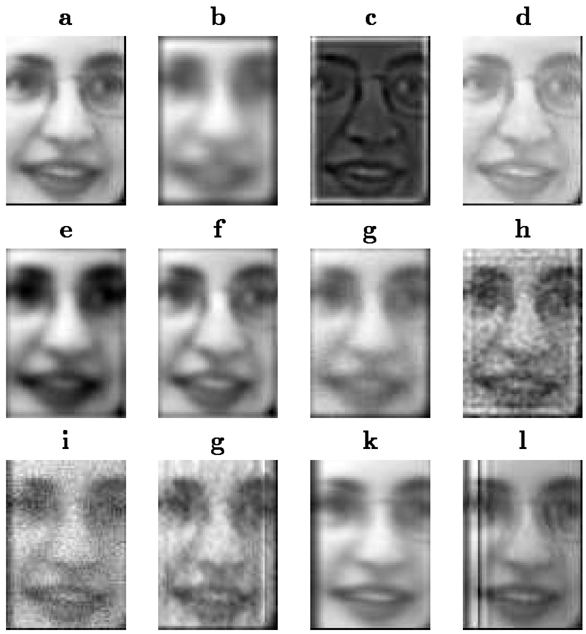

Figure 2. Corresponding degraded images description is given in Table 2

(*direct* approach); (ii) Constraining reconstruction of blurred images to their unblurred version using the hybrid network (*direct* approach); (iii) Use of state-of-the-art restoration methods as preprocessing to the degraded images (*restoration* approach)

We have demonstrated that the combined approach leads to improvement in recognition of images degraded by a wide range of image blur and noise.

References

[1] T. F. Chan and C. K. Wong. Total variation blind deconvolution. Technical report, UCLA, 1997.

[2] J. Flusser and T. Suk. Degraded image analysis: An invariant approach. *pami*, 20(6):590–603, June 1998.

[3] B. Gidas and A. Zelic. Object recognition via hierarchical syntactic models. In *13th International Conference on Digital Signal Processing Proceedings*, volume 1, pages 315–18, New York, NY, USA, 1997.

[4] R. C. Gonzalez and P. Wintz. *Digital Image Processing.* Addison-Wesley Publishing Company, 1993.

[5] G. E. Hinton and D. Camp. Keeping neural networks simple by minimizing the description length of the weights. *Sixth ACM Conference on Computational Learning Theory*, July 1993.

[6] A. K. Jain. *Fundamentals of Digital Image Processing.* Prentice Hall, London, 1989.

[7] A. K. Katsaggelos. Iterative image restoration algorithms. *Optical Engineering*, 28(7):735–748, 1989.

[8] M. Kirby and L. Sirovich. Application of the Karhunen-Loève procedure for characterization of human faces. *PAMI*, 12(1):103–108, 1990.

[9] D. C. Lai, J. Potenza, and K. Verfaille. Evaluation of image restoration filters for machine classification. *Optical-Engineering*, 23(6):794–800, Nov.-Dec 1984.

[10] D. J. C. MacKay. A practical Bayesian framework for backprop networks. *Neural Computation*, 4:448–472, 1992.

[11] R. Melamed, Y. Yitzhaky, N. S. Kopeika, and S. R. Rotman. Experimental comparison of three target acquisition models. *Optical-Engineering*, 37(7):1902–13, July 1998.

[12] H. Ogawa, E. Kawada, and N. Ueda. Application of image processing equipment with multiprocessors to line-drawing recognition. In *Proceedings of the SPIE*, volume 845, pages 97–103, 1987.

[13] S. Omachi and H. Aso. Recognition of printed characters by considering quality. *Transactions of the Information Processing Society of Japan*, 38(12):2490–8, December 1997.

[14] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[15] M. I. Sezan and A. M. Tekalp. Survey of recent developments in digital image restoration. *Optical Engineering*, 29(5):393–404, 1990.

[16] I. Stainvas. *Trade-off between recognition and reconstruction: Application of Neural Networks to Robotic Vision.* PhD thesis, Engineering Faculty, Tel-Aviv University, 1999.

[17] I. Stainvas, N. Intrator, and A. Moshaiov. Improving recognition via reconstruction. *Submitted to Neural Networks*, pages 1–15, 1999. http:/www.math.tau.ac.il/cur.ps.Z.

[18] H. Stark. *Image recovery: Theory and application.* Academic press, San Diego, 1987.

[19] A. Tankus, Y. Yeshurun, and N. Intrator. Face detection by direct convexity estimation. *Pattern Recognition Letters*, 18(9):913–922, 1997.

[20] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems.* V. H. Winston and Sons, Washington, 1977.

[21] Y.-L. You and M. Kaveh. A regularization approach to joint blur identification and image restoration. *IEEE Transactions on Image Processing*, 5(3):416–427, March 1996.

[22] S. S. Young. Real time vehicle identification by signal processing. In *Proceedings of the 5th International Conference on Signal Processing*, volume 1, pages 148–53, Waltham, MA, USA, 1994.

[23] W. Zhao, R. Chellappa, and A. Krishnaawammy. Discriminant analysis of principal components for face recognition. In *International Conference on Automatic Face and Gesture Recognition*, pages 336–341, Nara, Japan, April 14–16 1998.