Improving Classification by Reducing Lens Aperture

Inna Stainvas
School of Computer Science
Tel-Aviv University
stainvas@post.tau.ac.il

Zeev Zalevsky **David Mendlovic**
Department of Electrical Engineering-Physical Electronics
The Iby and Aladar Fleischman Faculty of Engineering
Tel-Aviv University

Nathan Intrator*
School of Computer Science
Tel-Aviv University
Ramat-Aviv, 69978 ISRAEL.
nin@post.tau.ac.il

September 2000

Abstract

We propose a mechanism to reduce defocus blur by reducing the aperture of the camera lens, and show that it leads to a far more robust recognition. The recognition is demonstrated via a Neural Network architecture which we have previously proposed for blurred face recognition.

Keywords: Classification network; Face Recognition; Network ensembles; Image Blur; Lens Aperture; Artificial Neural Networks; Hybrid Architecture.

*This work was partially supported by the Israeli Ministry of Science, by the Israel Science Foundation – Center of Excellence Program and by the Hermann Minkowski – Minerva Center for Geometry at Tel Aviv University.

1 Introduction

It is well known that blur is particularly harmful for recognition [11, 12, 2, 18] 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. [26, 3, 17]. Under the *restoration* approach, blurred images are restored before recognition [25, 13, 11]. The success of this approach is not obvious since image restoration is an ill-posed inverse problem [4, 7, 20] and restored images may contain artifacts. Under the *direct* recognition approach, blur is addressed explicitly in the recognition model [2, 14, 17]. In this paper, we propose to reduce an aperture of the camera lens in order to reduce the sensitivity to out of focus blur. The lens aperture controls the amount of light that enters the camera, and affects the depth of field and the resolution of the image. Depth of field is the range around a sharply focused object which is in focus. A small lens aperture reduces image resolution but results in greater depth of field (Figures 1, 5) and eventually leads to improved classification results in a greater distance range.

The reduced sensitivity to out-of-focus distance is demonstrated on two face recognition tasks which we have extensively studied in the past [19]. The training of a hybrid neural network architecture is done on faces that are in focus. Testing is done on in and out-of-focus faces.

2 Optical background

Imaging is one of the most basic applications in optics. The imaging may be obtained by using a lens. In order to achieve the imaging, one must fulfil a certain relation between the focal length of the lens f, the distances between the lens and the object d_o and the distance between the lens and the detector on which the image is generate d_i. Such a relation is:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f} \tag{1}$$

If one violates this imaging relation, the obtained image is called to be out of focus. In addition, the relation between the focal length of the lens and its dimensions is an important factor in determining the spatial resolution of the captured image. In systematic formulation, the popular term is the Optical Transfer Function (OTF) which is the spectral response of the imaging system. Thus, having a certain object which is being imaged by a given lens one may compute the obtained image by performing a Fourier transform for the object image multiplying it by the OTF and performing the inverse Fourier transform to return to the image plane. Obviously, violating the imaging relation and having the image out of focus also depends upon the dimensions of the lens and its focal length. It is interesting to note that increasing the ratio between the focal length and the lens aperture decreases the spatial resolution of the imaged picture but it also increases the tolerance to being out of focus. Meaning: getting out of focus distorts the spatial resolution. An increase in the ratio between the focal length and the aperture, increases the region in which a given spatial distortion is obtained in the image plane.

The dependence of the OTF upon all the above mentioned parameters may be formulated as:

$$OTF(f_x, f_y) \quad = \quad \Lambda(\frac{f_x}{2f_0})\Lambda(\frac{f_y}{2f_0}) \times$$

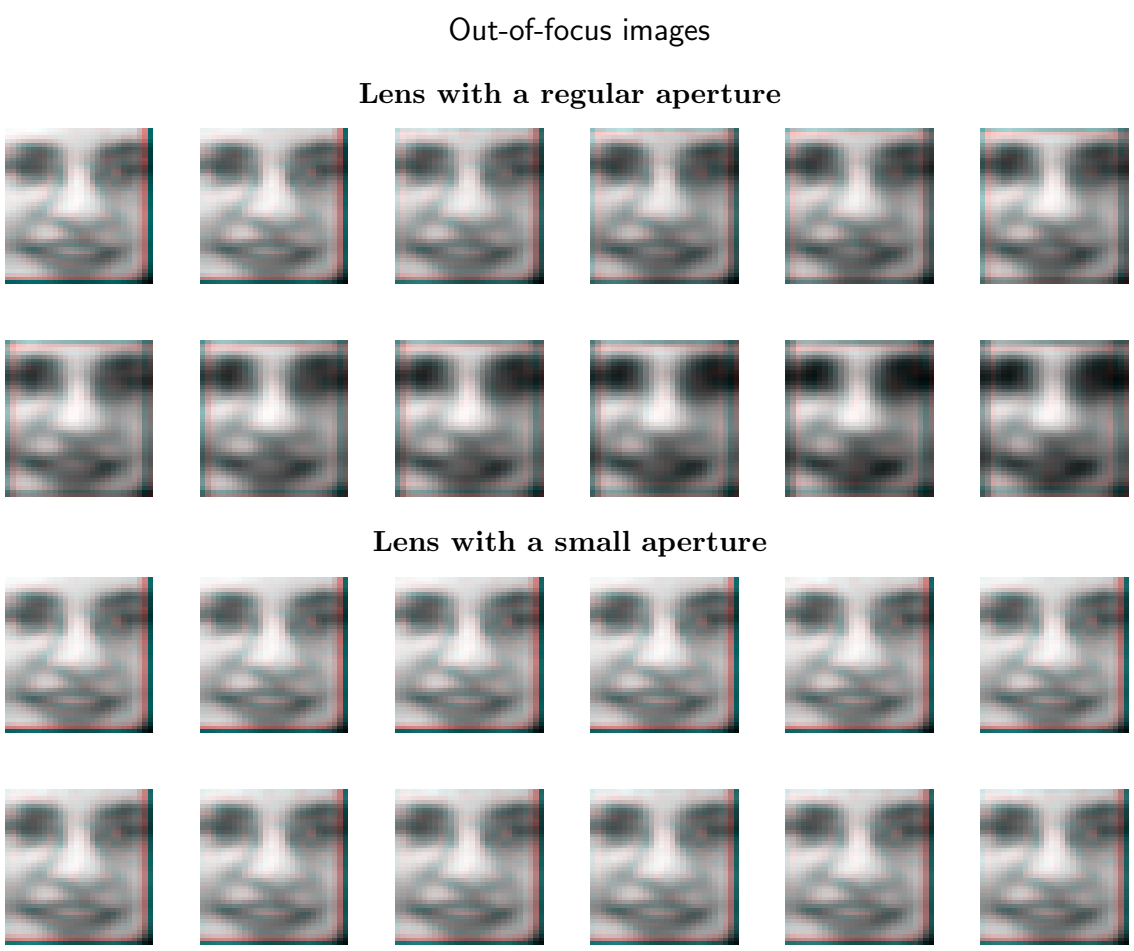

Figure 1: Images taken by two cameras with square shape lenses. The first image in the top and bottom figures correspond to the subject being in focus (with distance of 6m between the subject and the camera). Subsequent images are obtained when the subject gradually moves out of focus in steps of $0.5m$. The regular lens corresponds to a minimal focal number of $f_{\#} = 3$ and the small aperture to a focal number is $f_{\#} = 12$ (an aperture width is given by $D = f/f_{\#}$, where f is the lens focus). A small aperture width is 4 times less than for a regular lens. The camera's focal length is $f = 0.08m$. The interplay between low resolution and reduced out-of-focus blur is evident.

$$\operatorname{sinc}\left[\frac{8w}{\lambda}\left(\frac{f_x}{2f_0}\right)\left(1-\frac{|f_x|}{2f_0}\right)\right]\operatorname{sinc}\left[\frac{8w}{\lambda}\left(\frac{f_y}{2f_0}\right)\left(1-\frac{|f_y|}{2f_0}\right)\right],$$

where

$$w = \frac{\epsilon D^2}{8}, \tag{2}$$

with λ as the wavelength and D as the width of the aperture.

$$\frac{1}{d_o} + \frac{1}{d_i} - \frac{1}{f} = \epsilon \tag{3}$$

$$f_0 = \frac{D}{2\lambda d_i}. \tag{4}$$

Λ denotes the triangular function. The last equation may be approximated as:

$$OTF(f_x, f_y) = \text{sinc}\left[\frac{8w}{\lambda}\left(\frac{f_x}{2f_0}\right)\right] \text{sinc}\left[\frac{8w}{\lambda}\left(\frac{f_y}{2f_0}\right)\right]. \tag{5}$$

Let us now derive the tolerance to getting out of focus. To do so we will assume the geometrical optics approximation and refer to Figure 2. According to this figure and by using geometrical

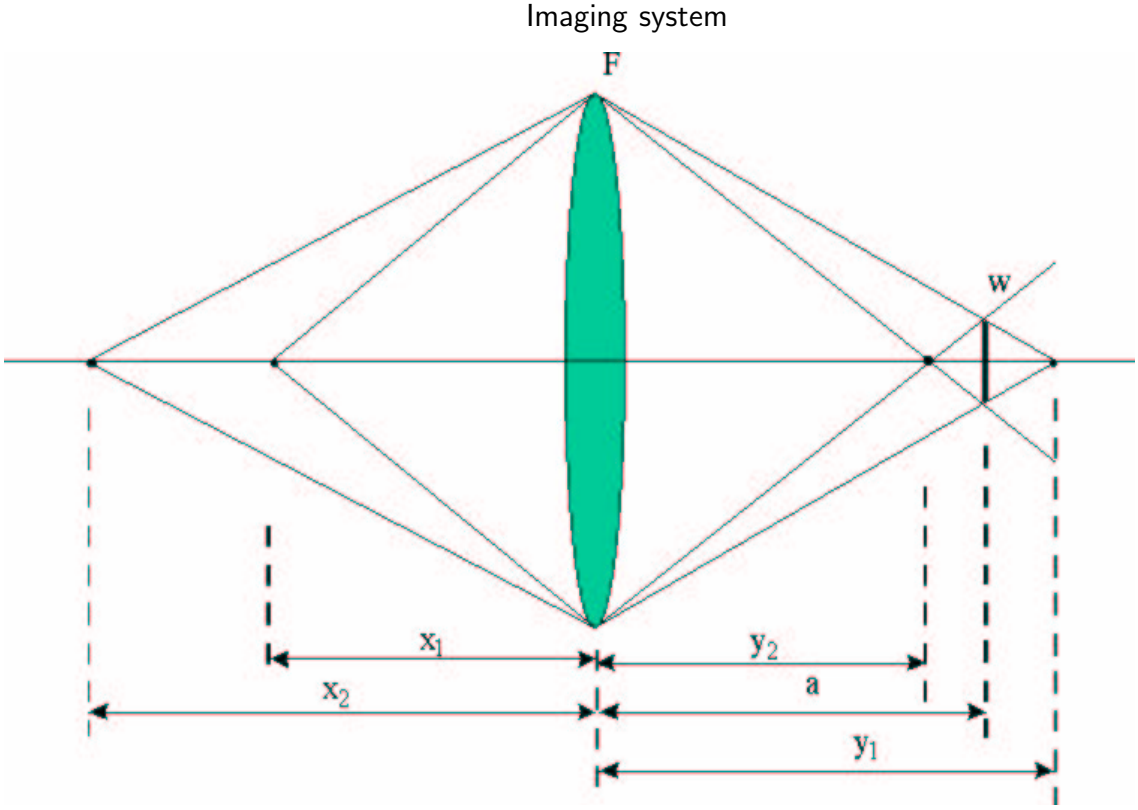

Figure 2: Basic setup for estimating the depth of focus range. w is the maximal bluring spot which is allowed. f is te focal length. D is the aperture. (x_1, y_1) and (x_2, y_2) fulfill the imaging relation.

relations we may define:

$$\begin{aligned} \frac{D}{y_2} &= \frac{w}{a - y_2} \\ \frac{D}{y_1} &= \frac{w}{y_1 - a}. \end{aligned} \tag{6}$$

where f is the focal length of the lens, D is the aperture, a is the distance between the imaging lens and the detection plane and w is the maximal bluring spot. Using the imaging relations yields:

$$\begin{aligned} \frac{1}{x_1} + \frac{1}{y_1} &= \frac{1}{f} \\ \frac{1}{x_2} + \frac{1}{y_2} &= \frac{1}{f}. \end{aligned} \tag{7}$$

Thus, the out of focus range becomes:

$$\Delta = x_1 - x_2 = \frac{2awDf^2}{a^2D^2 - 2aD^2f + d^2f^2 - w^2f^2}, \tag{8}$$

where w is the allowed resolution distortion in the image plane. The last expression may be approximated as:

$$\Delta \approx \frac{2awff_{\#}}{(a-f)^2}, \tag{9}$$

where

$$f_{\#} = \frac{f}{D} \tag{10}$$

As obtained, indeed an increase in the out of focus range is proportional to $f_{\#}$.

This result is the basic motivation for this paper. Increasing the $f_{\#}$ decreases the resolution but it increases the depth of focus range. Thus, training the neural net on images seen through an imaging system with increased $f_{\#}$ will decrease the depth of focus distortions. On the other hand, such increase reduces the spatial resolution and thus damages the training abilities of the network. Optimizing the $f_{\#}$ parameter allows to obtained improved recognition performance.

3 Neural Network Ensembles

For the face classification task, we have used a feed-forward artificial neural network architecture regularized by weight decay [16, 5, 9]. We combine networks to simple regression ensembles [17] which were shown to improve the performance of single experts [24, 10, 15]. The regression ensemble classification rule is based on averaging the real values of the outputs of all the ensemble members and then producing a decision by thresholding. The improvement in regression ensembles depends on the level of independence of the errors made by the experts. This independence reduces the contribution of the variance portion of the error when ensemble average is used [15]. This also gives some hints to which networks to combine. We consider two types of regression ensembles with network outputs averaged over: *(Ens. A)*, this is an ensemble on the initial weights, and *(Ens. B)* which averages of different training cross-folds – a "bagging" type of ensemble [1].

4 Data-set description and implementation details

The widely available facial data-set [23] as well as a face data-set locally collected by the Tel-Aviv University Computer Vision Group [21] were used in our simulations [17]. While there have been

many successful classification approaches with the Turk/Pentland data, we use it in a reduced resolution 32×32 and under considrable blur (the original resolution is 64×64). The data-set contains 27 images of 15 male faces (we took out the single bearded person). For each person, we randomly choose 15 images for training (data D) and 12 images for testing (data T). The 15 training samples were split into five cross-folds (by taking out 3 different images per person).

Preprocessing details and earlier results studying effect of background, illumination and comparison with PCA for original resolution are given in [6, 17]. The preprocessing partially removes the variability due to viewpoint, by setting (automatically) the eyes and tip of the mouth to the same position in all images [22]. 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 second data-set was collected by the Computer Vision group at Tel-Aviv University (TAU). It is of high resolution 84×56, and contains images of 37 male and female faces with 10 images per person. We reduced the resolution to 42×28 and split the data into test T (4 images per person) and training D (6 images per person) sets. Cross-validation with three disjoint groups having a size of 2 images per person is considered. Preprocessing was similar to the one described above, except that only the eye locations were fixed [22].

All networks have the same architectural complexity: the number of input units is the same as the number of image pixels (1176 for TAU data and 1024 for Pentland data); there are 15 hidden units with sigmoidal activation function in all networks and a number of output classification units is as the number of classes (37 for TAU data and 15 for Pentland data). A sigmoidal output activation function assures outputs between 0 and 1. The initial weights of the networks are chosen randomly out of a uniform distribution between -0.001 and 0.001. A constant learning rate is set to 0.05 and a weight decay regularization parameter is tuned to 0.05 (Full details of the choice of these parameters and the cross validation approach are given in [19].) Networks are trained 3000 epochs for Pentland data and 5000 epochs for TAU data. Initial weights are resampled 5 times.

5 Experimental design

We train networks with 5 different initial weight conditions per 3 cross-folds for TAU data and per 5 cross-folds for Pentland data. Therefore, 15 networks for TAU data and 25 networks for Pentland are considered in summary. Ensembles of type (A) are obtained by averaging networks trained on the same cross-folds and with different initial weights, i.e. a number of ensembles of type (A) coincides with the number of cross-folds used. Therefore, ensemble (A) is run 3 times for TAU data and 5 times for Pentland data; the number of network members in ensemble A equals 5. Ensembles of type (B) are obtained by averaging networks trained with the same initial weights but on different cross-folds, i.e. a number of ensembles of type (B) coincides with the number of initial weight conditions used. There are 5 ensembles of type B with the number of members equal to 3 for TAU data and 5 ensembles of type B with the number of members equal to 5 for Pentland data (i.e. ensembles B are run 5 times). Mean and variance statistics over ensemble runs are evaluated on the separately kept test set T.

To demonstrate our approach networks are trained twice on the focused images "taken" by two cameras with a lens of a regular and small aperture. Indeed, an original data-set is considered as an "ideal subject images". We filter the original data assigned to training by OTF of square shape lenses in focus (See Figure 3). In the testing stage images assigned for testing are filtered by OTF corresponding to situation when subject gradually moves out of focus. For simulation we have used

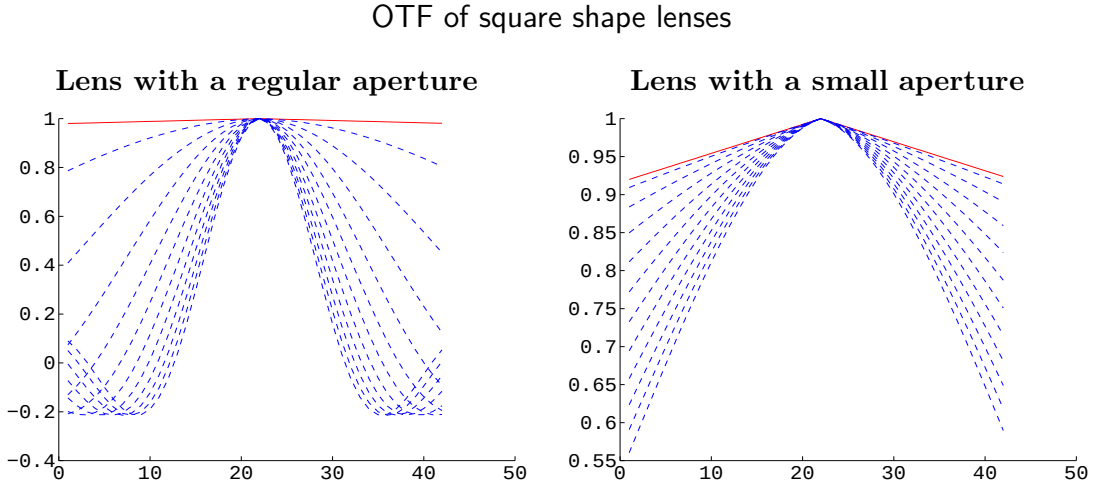

Figure 3: OTF of two cameras with square shape lenses. The OTF in focus are shown by solid lines and OTF in out of focus by dashed lines. Corresponding filtered images are presented in Figure 1 and camera's characteristics are the same as indicated in the caption of Figure 1. Pay attention that though OTF in focus of a small lens introduces much more corruption, its OTF sensitivity to a shift from the focus is much smaller.

a real camera parameters used in our face recognition application (see Table 1)

Imaging system parameters

focus f	focal number for a regular aperture	focal number for a small aperture	distance between subject and lens in focus	a step for shift from focus
$0.08m$	3	12	$6m$	$0.5m$

Table 1: Parameters of the imaging system that are used in our simulations. Focal number is a ratio between the focal length and the aperture: $f_{\#} = f/D$. A small lens has 4 times less aperture than a regular lens.

6 Results

Figure 4 and a corresponding Table 2 present results on the TAU data-set. These include misclassification error for ensembles A and B for cameras with a regular and small aperture lenses. They demonstrate that using a small aperture lens leads to robust classification results, i.e. though the out focus parameter grows, changes in degraded images have a smaller effect on classification.

The sensitivity of our network ensembles to a lens with a regular aperture is very strong: the error grows from 14.9% to 22.1% for Ens. A and from 10.3% to 17.0% for Ens. B. When objects

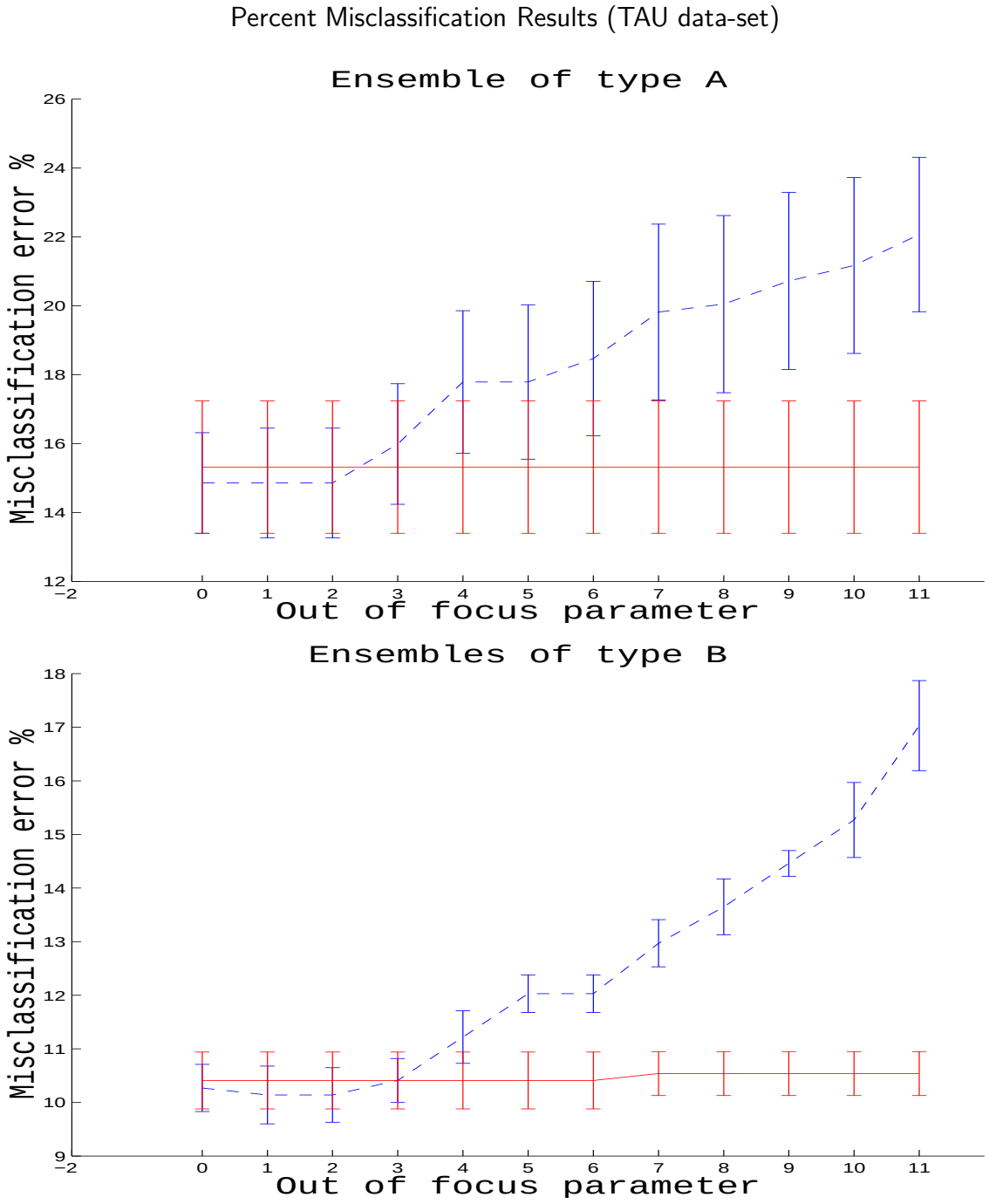

Figure 4: Percent Misclassification Results for Ensembles of types A and B. Dashed blue lines correspond to a camera with the regular lens and red for a camera with a small aperture. Out of focus parameter is a number of step taken from the focus. In focus this parameter is equal to zero.

Percent Misclassification Results (TAU data set)

Out of focus parameter												
Step $0.5m$	0	1	2	3	4	5	6	7	8	9	10	11
Ens. A corresponding to a regular lens aperture												
means	14.9	14.9	14.9	16.0	17.8	17.8	18.5	19.8	20.1	20.7	21.2	22.1
std	1.5	1.6	1.6	1.8	2.1	2.2	2.2	2.5	2.6	2.6	2.5	2.2
Ens. A corresponding to a small lens aperture												
means	15.3	15.3	15.3	15.3	15.3	15.3	15.3	15.3	15.3	15.3	15.3	15.3
std	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9
Ens. B corresponding to a regular lens aperture												
means	10.3	10.1	10.1	10.4	11.2	12.0	12.0	13.0	13.7	14.5	15.3	17.0
std	0.4	0.5	0.5	0.4	0.5	0.3	0.3	0.4	0.5	0.2	0.7	0.8
Ens. B corresponding to a small lens aperture												
means	10.4	10.4	10.4	10.4	10.4	10.4	10.4	10.5	10.5	10.5	10.5	10.5
std	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.4	0.4	0.4	0.4	0.4

Table 2: Out-of-focus parameter stands for a number of steps of a length of 0.5m taken by subject when he/she moves out-of-focus. See also corresponding Figures 4–1.

are in focus, their images for a lens with a small aperture has lower resolution than for a regular one (compare OTF in focus Figure 3). This explains why ensembles corresponding to the camera with the regular lens have smaller mean classification errors than ensembles corresponding to the camera with the small lens (Table 2). However, analysis of variances demonstrate that this superiority is insignificant. In addition, it is easy to see that ensembles of type B lead to better classification results and this is despite a smaller number of ensemble members. However, one has to take into account that network members of ensemble B see the whole training data by blocks/cross-folds, while networks of ensemble A see only a particular cross-fold for which they were trained on.

Similar results for Pentland data-set are presented in Table 3 and Figure 6. This data is easier to classifity since faces were normalized by both eyes and mouth; and there are less classes. Starting misclassification percent errors turns out to be small, about 1.7%. Neural networks ensembles A–B for this data-set are less sensitive to out-of-focus shifting. Nevertheless, the same qualitative result is found, i.e. using of a lens with a smaller aperture improves classification and ensembles B are superior to ensembles A. In contrast with TAU data, error means for ensembles A and B for the lens with the small aperture are less than error means for ensemble counterparts for the regular lens when images are in focus. This is despite the lower image resolution of focused images for lenses with smaller apertures. However, analysis of variances shows that this difference is insignificant. Out-of-focused images for one subject of Pentland data are presented in Figure 5.

7 Conclusions

In this paper we have demonstrated that a trade off axists between depth of focus and spatial resolution. A decrease in lens aperture increases the depth of focus but also distorts the image resolution. It was evident that decreasing the aperture of the lens up to a certain level indeed produced a bit larger classification error for in focus images, however, it maintained significantly smaller error for much larger depth of focus ranges.

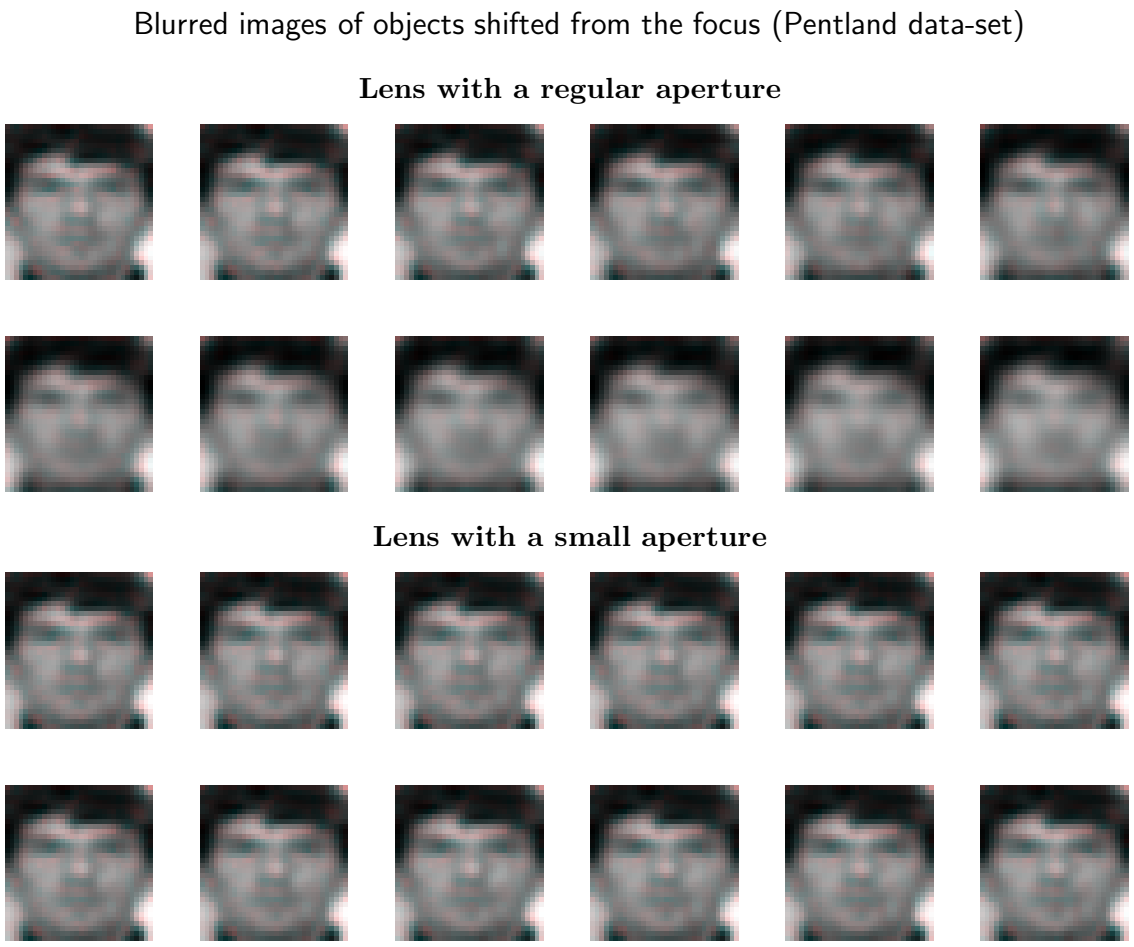

Figure 5: Images taken by two cameras with square shape lenses. The first image on the top and bottom figures correspond to the subject being in focus (a distance between subject and camera equals $6m$). Subsequent images are obtained when subject gradually moves out of focus with a step $0.5m$. A regular lens corresponds to a minimal focal number equal $f_{\#} = 3$ and a small aperture to a focal number equal $f_{\#} = 12$ (an aperture width is given by $D = f/f_{\#}$, where f is a lens focus). A small aperture width is 4 times less than for a regular lens. Cameras focus is equal $f = 0.08m$. Pay attention that though images in focus taken by a small lens have lower resolution, they contain less blur when subject moves out of focus.

References

[1] L. Breiman. Bagging predictors. *Machine Learning*, 24:123–140, 1996.

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

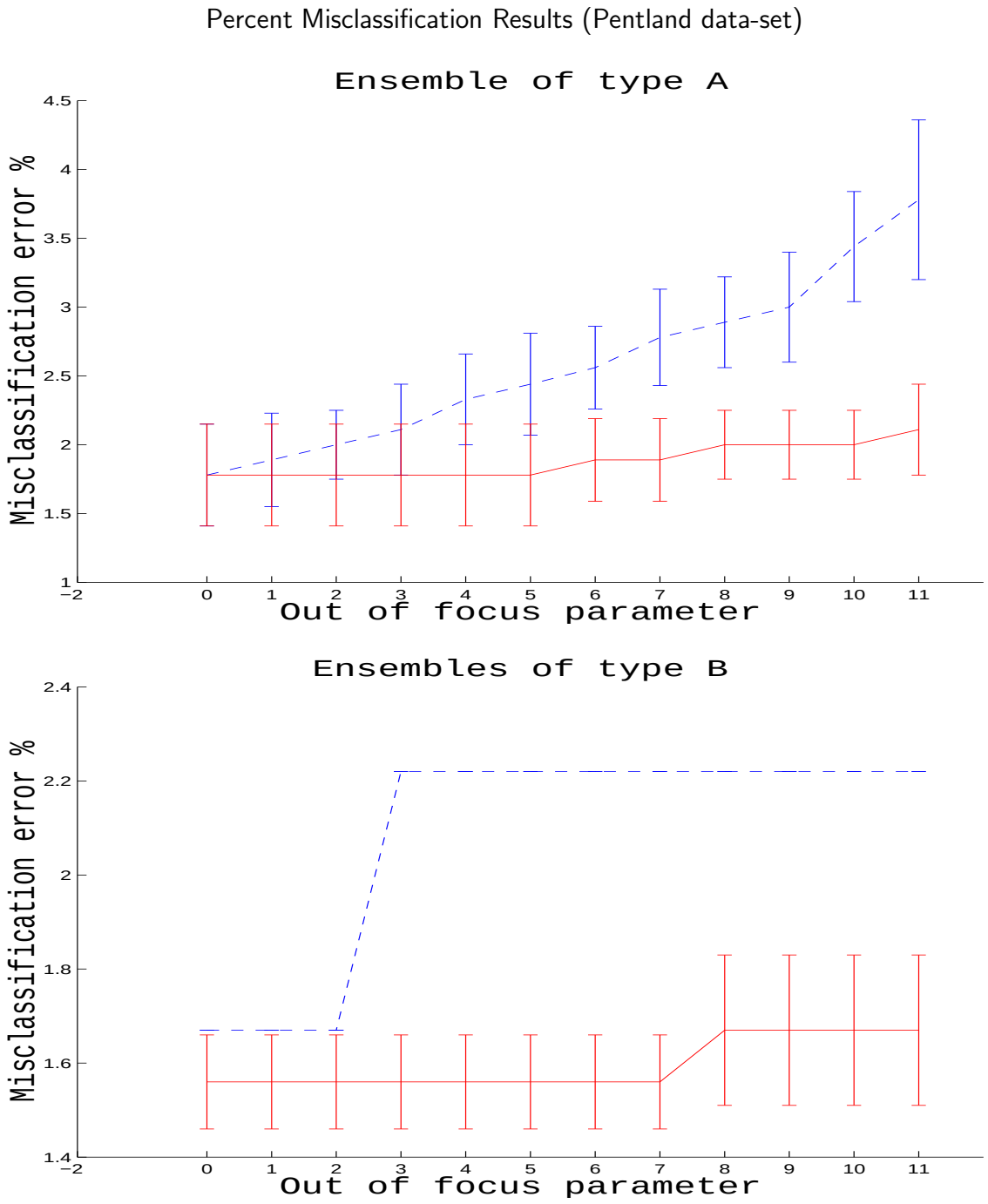

Figure 6: Percent Misclassification Results for Ensembles of types A and B. Dashed blue lines correspond to a camera with the regular lens and red for a camera with a small aperture. Out of focus parameter is a number of step taken from the focus. In focus this parameter is equal to zero.

Percent Misclassification Results (Pentland data-set)

Out of focus parameter												
Step $0.5m$	0	1	2	3	4	5	6	7	8	9	10	11
Ens. A corresponding to a standard lens aperture												
means	1.8	1.9	2.0	2.1	2.3	2.4	2.6	2.8	2.9	3.0	3.4	3.8
std	0.4	0.3	0.2	0.3	0.3	0.4	0.3	0.3	0.3	0.4	0.4	0.6
Ens. A corresponding to a small lens aperture												
means	1.8	1.8	1.8	1.8	1.8	1.8	1.9	1.9	2.0	2.0	2.0	2.1
std	0.4	0.4	0.4	0.4	0.4	0.4	0.3	0.3	0.2	0.2	0.2	0.3
Ens. B corresponding to a standard lens aperture												
means	1.7	1.7	1.7	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2
std	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Ens. B corresponding to a small lens aperture												
means	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.6	1.7	1.7	1.7	1.7
std	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2

Table 3: Out-of-focus parameter stands for a number of steps of a length of 0.5m taken by subject when he/she moves out-of-focus.

[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. Learning distributed representations of concepts. In *Proceedings of the 8th Annual Conference of the Cognitive Science Society*, pages 1–12. Hillsdale: Erlbaum, 1986.

[6] N. Intrator, D. Reisfeld, and Y. Yeshurun. Face recognition using a hybrid supervised/unsupervised neural network. *Pattern Recognition Letters*, 17:67–76, 1996.

[7] A. K. Jain. *Fundamentals of Digital Image Processing.* Prentice Hall, London, 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] A. Krogh and J. A. Hertz. A simple weight decay can improve generalization. In J.E. Moody, S.J Hanson, and R.P. Lippmann, editors, *Advances in Neural Information Processing Systems*, volume 4, pages 950–957. Morgan Kaufmann, San Mateo, CA, 1992.

[10] A. Krogh and J. Vedelsby. Neural network ensembles, cross validation, and active learning. In G. Tesauro, D. Touretzky, and T. Leen, editors, *Advances in Neural Information Processing Systems*, volume 7, pages 231–238, Cambridge, MA, 1995. MIT Press.

[11] 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.

[12] 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.

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

[14] 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.

[15] Y. Raviv and N. Intrator. Bootstrapping with noise: An effective regularization technique. *Connection Science, Special issue on Combining Estimators*, 8:356–372, 1996.

[16] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. In D. E. Rumelhart and J. L. McClelland, editors, *Parallel Distributed Processing*, volume 1, pages 318–362. MIT Press, Cambridge, MA, 1986.

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

[18] I. Stainvas and N. Intrator. Blurred face recognition via a hybrid network architecture. In *Proceedings Int. Conf. on Pattern Recognition*, 2000.

[19] I. Stainvas, N. Intrator, and A. Moshaiov. Improving recognition via reconstruction, 1999. Submitted (ftp://cns.brown.edu/nin/papers/tradenips.ps.Z).

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

[21] A. Tankus. *Automatic face detection and recognition*. Master thesis, Tel-Aviv University, 1996.

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

[23] M. Turk and A. Pentland. Experiments with eigenfaces. *Looking At People Workshop, IJCAI'93*, pages 1–6, 1993.

[24] D. H. Wolpert. Stacked generalization. *Neural Networks*, 5:241–259, 1992.

[25] Sung Soh 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.

[26] 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.