

# SEPARATING USEFUL FROM USELESS IMAGE VARIATION FOR FACE RECOGNITION

Peter Kalocsai

University of California, Los Angeles, USA

## ABSTRACT

For a general purpose face recognition system one of the largest challenge is to separate useful identity related from useless variations in the image data due to nuisance variables such as: orientation, lighting, expression, possible disguise. A recognition system is presented in which the effect of the secondary/nuisance variables is to a large degree accounted for before the matching process even begins. Greatly improved performance is shown on a large database of faces in 42 conditions.

## 1. INTRODUCTION

Compared with other biometrics used in recognition systems such as fingerprint or iris, face recognition lags far behind in performance. The reason for the large discrepancy directly relates to the number of primary and secondary variables that these systems have to deal with. Typically in a face recognition system the input face images can be expressed in the following form:

$$I = f(i, o, l, e, d) \quad (1)$$

where a face image  $I$  is expressed as a function of  $i$ =identity,  $o$ =orientation (including tilt, pose, slant, inclination variations),  $l$ =lighting,  $e$ =expression,  $d$ =disguise (possible glasses, sunglasses, hat, hairstyle change etc.). From all these variables only identity provides useful variation for a face recognition system the rest can be considered useless noise. In a much more controlled environment for fingerprint and iris recognition the effect of useless secondary variables can be significantly reduced or almost eliminated. Face recognition promises the least intrusive way to collect biometric data, but it can only be successful if the impact of the useless secondary variables can be reduced in the recognition process. The purpose of the current study is to provide a way for separating useful variations caused by the single primary variable (identity) from useless variations caused by all the other secondary variables.

## 2. FACE REPRESENTATION

We represent face images as convolution values with a set of multiscale and multiorientation Gabor kernels at multiple characteristic locations on the face (e.g. corner of the mouth, tip of the nose as shown on Figure 1) [2].

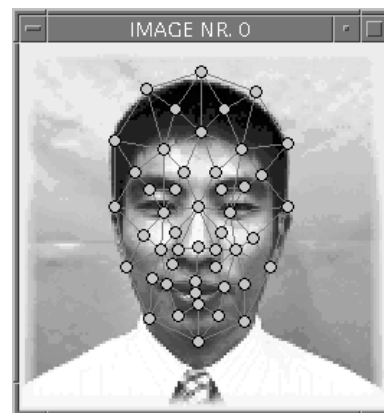


Figure 1. Convolution locations indicated on a neutral frontal image.

A univariate analysis of variance was applied to the Gabor wavelet representation of 1414 Asian face images (101 individuals X 14 conditions) to derive the discriminative power of all wavelets individually for the recognition of these faces.

$$F = \frac{\sum_{i=1}^g (n_i (\bar{x}_i - \bar{x})) / (g-1)}{\sum_{i=1}^g \left( \sum_{j=1}^{n_j} (x_{ij} - \bar{x}_i)^2 \right) / \left( \sum_{i=1}^g (n_i) - g \right)} \quad (2)$$

where  $x$  represents a kernel's activation value,  $\bar{x}_i$  and  $\bar{x}$  are average activation values for individual  $i$  and for all images respectively,  $n_i$  expresses the number of face images in group  $i$  (always 14 in our case) and  $g$  refers to the number of groups (101 in the above example).

After the analysis all 1920 Gabor wavelets (48 locations X 8 orientations X 5 scales) in the representation

were ranked according to their discriminative power. The result of the ANOVA showed that the forehead and the eye regions were quite important. The region between the nose and mouth also seemed to be rather informative (Figure 2). The different frequency channels overall seemed to have equal importance. The range of the F values was about one magnitude. The highest F value was  $F(100,1313) = 25$  and the lowest was  $F(100,1313) = 1.5$ . The reader is referred to [1] for more details on the analysis. To limit the dimensionality only the top 150 best discriminating kernels were used in the further analyses.

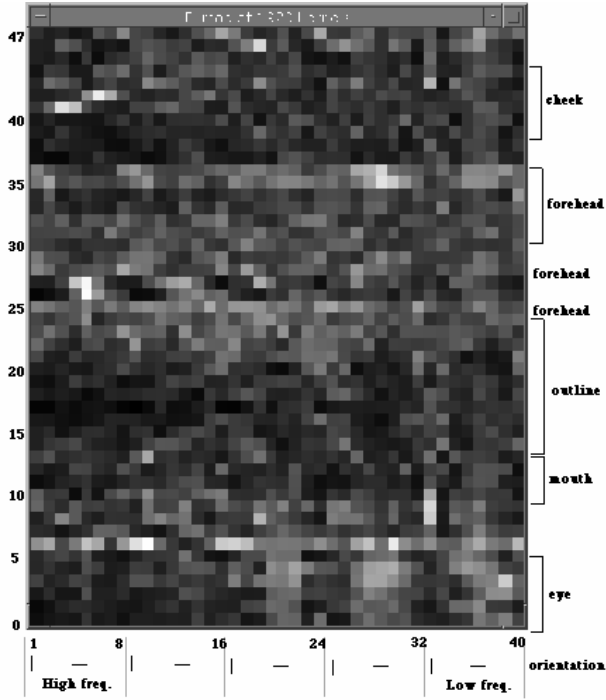


Figure 2. Gray level coded significance map (F map) of all 1920 kernels in the representation. Whiter shading indicates higher discriminative power. The horizontal axis represents kernel size and orientation and the vertical one indicates the location of the kernel on the face.

### 3. SEPARATING IMAGE VARIATION

As noted earlier the great difficulty in trying to recognize faces in many different conditions is that most of the variance in the data could usually be attributed to the conditions themselves and possibly only a relatively small portion of it represents individual variations. If there was a way to eliminate or at least limit the effect of conditional changes even before the matching process starts that could significantly increase the chances of correct recognition. In this scenario when a test image is presented to the system first the conditions under which this image was

taken would be detected. This is achieved by modeling the possible changes that could be introduced to faces in various conditions (e.g., it would be modeled how the representation would change if the face would be rotated from frontal to 30 degrees to the left). For purposes of matching the size of the face database could still be kept minimal with only one frontal image of every person being stored with neutral expression and normal lighting. The critical part is the modeling of how the representation of a face would change in various conditions. Once the density of a conditional change is correctly estimated then the rest of the variation would account for changes in identity. In other words by accounting for the useless secondary variables we are left with the only relevant one primary variable. To estimate these densities first based on the top 150 best discriminating kernels in the representation (out of 1920) the representative difference space is calculated for every individual in the training data in each condition:

$$\Delta_i^c = |f_i - y_i^c| \quad (3)$$

where  $f_i$  is the frontal image for individual  $i$  and  $y_i^c$  is another image of that individual in condition  $c$ . The difference space covariance matrices are calculated for all conditions:

$$S^c = \frac{1}{n} \text{Sum}_{i=1}^n ((\Delta_i^c - \bar{\Delta}^c)(\Delta_i^c - \bar{\Delta}^c)') \quad (4)$$

When a test image is presented first the difference vectors are calculated between this image and all frontal gallery images in the database:

$$\Gamma_i^t = |f_i - x^t| \quad (5)$$

Then the closest  $c$  condition is derived for our test image via finding the best match between  $\Gamma^t$  and  $\Delta^c$ :

$$c = \arg \min_{c \in C} |\Gamma^t - \Delta^c| \quad (6)$$

With this the first part of the recognition task is completed by accounting for the nuisance/secondary variable(s). Now we can focus on the recognition task itself based on the primary variable: identity. Once the appropriate  $c$  condition is chosen for our  $\Gamma^t$  differences then the Mahalanobis distance is applied to calculate the distance between  $\Gamma^t$  and the  $\Delta^c$  difference vectors:

$$d_{ij}^2 = (\Gamma_i^t - \Delta_j^c)'(S^c)^{-1}(\Gamma_i^t - \Delta_j^c) \quad (7)$$

Next we sum up all the distances from the estimated  $\Delta^c$  target vectors:

$$D_i = \text{Sum}(d_{ij}^2) \quad (8)$$

The minimum of these summed difference values is the found match, the only same person difference in the database, since all the other differences are higher due to the additional variation caused by differences in identity:

$$p = \arg \min_{i \in P}(D_i) \quad (9)$$

where  $P$  is the total set of images in the database and  $p$  is the found match in that database. Note that if the minimum of the  $D_i$  values is not sufficiently small, it is above a predetermined threshold value, then there is no found match for the test image. Moghaddam and Pentland describes a similar method of density estimation except that they had to deal with reducing the dimensionality of their PCA representation [3,4]. Since our analysis was based on the limited set of kernels chosen by ANOVA we escaped this problem. In addition, Moghaddam and Pentland formulates a Bayesian rule for the posterior probability that a given difference vector represents an intrapersonal difference as opposed to an extrapersonal one:

$$\Pr = P(\Omega_I | \Delta) = \frac{P(\Delta | \Omega_I)P(\Omega_I)}{P(\Delta | \Omega_I)P(\Omega_I) + P(\Delta | \Omega_E)P(\Omega_E)} \quad (10)$$

where  $\Omega_I$  and  $\Omega_E$  represent intrapersonal and extrapersonal variations respectively. In this formulation two images are determined to belong to the same individual if  $P(\Omega_I | \Delta) > P(\Omega_E | \Delta)$ , or equivalently, if  $\Pr \langle I_t, I_g \rangle > 1/2$  where  $I_t$  is a test image and  $I_g$  is an image from the gallery. This formulation is somewhat similar to ours although on a completely different representation and without clearly accounting for different conditional variations.

Condition	Sep. Meth.	Baseline	MIT/Eig
Tilt Down 45	<b>71.8</b>	6.4	5.5
Tilt Down 30	<b>76.4</b>	26.4	22.7

Tilt Down 20	<b>90</b>	57.3	40.9
Tilt Down 10	<b>95.4</b>	<b>95.4</b>	70
Tilt Up 45	<b>81.8</b>	9.1	5.5
Tilt Up 30	<b>88.2</b>	28.2	14.5
Tilt Up 20	<b>90</b>	46.4	27.3
Tilt Up 10	<b>95.5</b>	79.1	65.5
Pose Left 90	<b>68.8</b>	0.9	0.9
Pose Left 45	<b>80.9</b>	10.9	2.7
Pose Left 30	<b>80.9</b>	28.2	7.2
Pose Left 20	<b>88.1</b>	59.6	28.1
Pose Left 10	<b>95.4</b>	91.7	64.4
Pose Right 90	<b>66.1</b>	3.7	0.0
Pose Right 45	<b>72.7</b>	9.1	6.4
Pose Right 30	<b>79.1</b>	28.2	8.2
Pose Right 20	<b>90</b>	57.3	29.1
Pose Right 10	<b>95.5</b>	89.1	63.6
Incl. Right Up	<b>85.6</b>	20.2	10.4
Incl. Right Down	<b>75</b>	20.2	5.8
Incl. Left Up	<b>81.9</b>	10.5	14.3
Incl. Left Down	<b>69.5</b>	13.3	11.4
Slant Right 15	<b>81</b>	22.8	22.4
Slant Right 10	<b>86.1</b>	48.1	38.3
Slant Right 5	<b>91.8</b>	73.1	59.8
Slant Left 15	<b>81.9</b>	31.4	26.2
Slant Left 10	<b>87.1</b>	53.7	37.4
Slant Left 5	<b>94.4</b>	74.1	59.8
Size 120%	97.2	<b>99.1</b>	70.1
Size 64%	93.3	<b>98.3</b>	95
Size 80%	97.2	<b>100</b>	89.5
Illum. Bottom	<b>85.2</b>	64.8	39.8
Illum. Side	<b>90.7</b>	87.9	38
Illum. Dark2	98.2	<b>100</b>	82.7
Illum. Dark1	97.2	<b>100</b>	90.8
Glasses Wearing	<b>100</b>	89.7	90.9
Glasses Taking off	85.4	<b>97.6</b>	72.7
Hair Change style	<b>98.9</b>	93.2	46.1
Cap Wearing	<b>84.4</b>	77.8	15.5
Expr. Open mouth	<b>97.3</b>	87.3	59.1
Expr. Close one eye	97.2	<b>100</b>	78.3
Expr. Close eyes	99.1	<b>100</b>	71.8
<b>Average</b>	<b>87.2</b>	<b>56.9</b>	<b>40.3</b>

Table 1. Performance comparison of the Separation Method, the Baseline Gabor filter and the PCA based model. Best performers are indicated in bold.

## 4. RESULTS

The performance of three recognition systems: our separation method based one, that of the baseline Gabor filter based model [2], and that of a PCA based model using eigenfaces was compared (Table 1). The results indicate that out of the 42 tested conditions on faces in the testing set performance improved by using the variance separation method in 33 conditions; it was equal to the original baseline performance in 1 conditions; and it was worse in 8 conditions. Performance of the separation method was better than that of PCA in all but one condition. Overall our method's performance was more than twice as good as that of the PCA based eigenface method developed at MIT. Although for some of the conditions were the baseline method performed very well (100% or close) the separation method was not able to improve or even achieve the same results, in most cases very significant improvement was demonstrated. All across the 42 conditions recognition performance has improved by over 30% on average compared with the baseline Gabor filter based model. Also note that this improvement over the baseline model was achieved with over a magnitude decrease in data (150 kernel values were used instead of 1920). These results demonstrate that if the condition of a test face image (orientation, expression, illumination) compared to a normal, frontal view can be estimated, then the variance separation method can achieve significant improvements in performance.

The two histograms on Figure 3 show that if the condition of a test face is correctly estimated then the difference of a test face from its frontal match in the database produces a much tighter and closer to 0 distribution (left side of Figure 3) on the 150 kernel activation values than the difference between that face and all other individuals in the database (right side of Figure 3). Essentially what happened by taking care of the nuisance/secondary variables was that the variance on the primary variable became much more visible and useful for recognition. Differences due to identity are much more emphasized now. It is also worth mentioning that out of the 150 kernels in the analysis there was not one for which the difference between different person images on average was less than that between two pictures of the same person. Altogether for all 150 kernels the total deviation of same person scores was 198, whereas the same number for different person deviations was 898. This means that on average the different person score was close to four times larger than the same person score. These findings provide further evidence for the usefulness of the variation separation method.

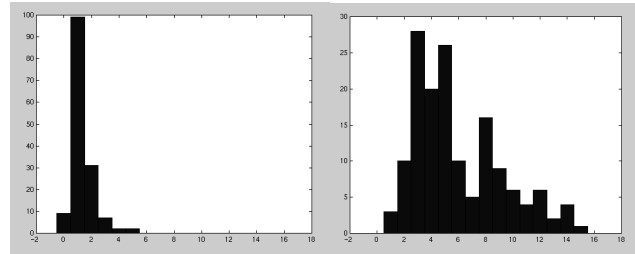


Figure 3. Left: Average deviation of all 150 kernel activation values between a test and database image when they are of the same person (all values are close to 0). Right: The same as above when the images are of two different people (values are much more spread out). Values are presented after variation caused by secondary/nuisance variables is accounted for.

## 5. CONCLUSIONS

Face recognition can only be successful if the image variation caused by the nuisance/secondary variables (orientation, illumination, expression, disguise) can be separated from the only useful variation for recognition: that of identity. We described a two step procedure for the recognition of faces in which first the conditions under which pictures were taken are estimated. Once the found conditional variation caused by nuisance variables is accounted for a much better recognition performance can be achieved. One future direction is building a better model for the secondary variables which would include their possible interactions.

## 6. REFERENCES

- [1] P. Kalocsai, C. von der Malsburg, and J. Horn, "Face recognition by statistical analysis of feature detectors," *Image and Vision Computing*, 18, Elsevier, pp. 273-278, 2000.
- [2] M. Lades, J. C. Vortbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Würtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Transactions on Computers*, 42, pp. 300-311, 1993.
- [3] B. Moghaddam, W. Wahid, and A. Pentland, "Beyond Eigenfaces: Probabilistic Matching for Face Recognition," In *Proc. of the Third IEEE Int. Conf. on Automatic Face and Gesture Recognition*, Nara, Japan, pp. 30-35, 1998.
- [4] B. Moghaddam, and A. Pentland, "Probabilistic Visual Learning for Object Detection," In *Proc. of the 5<sup>th</sup> International Conference on Computer Vision*, Cambridge, MA, 1995.