IEEE Computer Society Press
Attn: Bob Werner
(Proceedings ICPR 1996)
10662 Los Vaqueros Circle
PO Box 3014
Los Alamitos, CA 90720-1264
USA

Dear Bob Werner,

Enclosed is the final version of our paper entitled, "Information Theory and Face Detection," for the International Conference on Pattern Recognition, August, 1996.

Sincerely,

Michael S. Lew

Department of Computer Science
Leiden University
Post Box 9512
2300 RA Leiden
The Netherlands
mlew@wi.leidenuniv.nl

Information Theory and Face Detection

Michael S. Lew Nies Huijsmans

Department of Computer Science, Leiden University, Netherlands
E-mail: {mlew huijsman}@wi.leidenuniv.nl

Topics: face detection, information theory, maximum information gain

Abstract

Face detection in complex environments is an unsolved problem which has fundamental importance to face recognition, model based video coding, content based image retrieval, and human computer interaction. In this paper we model the face detection problem using information theory, and formulate information based measures for detecting faces by maximizing the feature class separation. The underlying principle is that search through an image can be viewed as a reduction of uncertainty in the classification of the image. The face detection algorithm is empirically compared using multiple test sets, which include four face databases from three universities.

1 Introduction

The information theoretic approach provides a foundation for determining new insights and solutions toward image modeling and analysis problems. In this paper we describe information theoretic solutions toward solving face detection. From the survey of Chellappa, et. al. [2], they conclude that segmentation of face regions from images is an important problem which has received surprisingly little attention.

Face detection in complex environments is an important unsolved problem which has fundamental importance to human computer interaction, model based video coding and face recognition. The face detection problem may be described as follows: Given a test image (any scanned in photograph or frame from a video camera), find the locations and size of every human face within the image. The problem of face detection differs from the problem of face recognition in that face detection has exactly two classifications: face or nonface, whereas face recognition usually has a number of classifications equal to the number of individuals. Face recognition programs usually require face detection before the individual recognition can be performed. The task of face detection is often avoided by manual segmentation of the input image or by making the assumption of a simple or uniform background.

Face detection is also interesting because it could provide valuable insight into the general topic of 3D object recognition since it shares many of the same problems. Some of the most important problems are (1) view dependence: the image of the face will vary with the viewing direction (2) nonrigidity: from the same viewpoint, different facial expressions will result in different images and (3) lighting: with the same viewpoint and the same facial expression, the image can be different due to diverse lighting environments.

There has been considerable recent interest in face detection. Some of the work includes the following: Yuille, et al. [13] used deformable templates to model the eyes, nose and lips. Huang and Tang [4] used the fast Fourier transform on the Laplacian of the Gaussian to perform face detection. Pentland, et al. [7] used eigenvectors to describe entire faces and features such as the eyes and nose. Yang and Huang [12] use a constraint based image pyramid. Sung and Poggio [11] used a neural network to find face and nonface clusters which are described using eigenvectors. Rowley and Kanade [9] compare different strategies in using neural nets for detection of faces.

The goals of this paper are twofold. First we review the fundamental relationships between the established estimation and information theoretic principles, namely, maximum likelihood, Shannon's [10] mutual information, Akaike's information criterion [1], and the Kullback relative information [5]. Second, we apply the Kullback relative information toward optimizing face detection with emphasis on template design.

2 Estimation principles

In this section, we show that the Kullback relative information can be viewed as an underlying basis for the maximum likelihood principle, Shannon's mutual information, and Akaike's information criterion.

Given a set of estimates *v'* of the vector of parameters v of a probability distribution with density function $p(x \mid v)$, we use in maximum likelihood estimation the estimate which maximizes the expected log likelihood

$$E[\log p(X|v')] = E\left[\int p(x|v)\log p(x|v')dx\right] \quad (1)$$

where X is a random variable with distribution $p(x|v)$. It can be shown that (1) is equivalent to maximizing

$$E\left[\log\frac{p(X|v')}{p(X|v)}\right] = E\left[\int p(x|v)\log\left(\frac{p(x|v')}{p(x|v)}\right)dx\right] \quad (2)$$

which by definition is the Kullback relative information between $p(x \mid v')$ and $p(x \mid v)$ which is usually written as

$$J(q_1,q_0) = \int\left[q_1(y)\log\frac{q_1(y)}{q_0(y|v')}\right]dy \quad (3)$$

Shannon's [10] model of an information system is completely defined by the source alphabet, $\mathbf{A}=[a_1,...,a_J]$, the source symbol probabilities, $\mathbf{z}=[P(a_1),...,P(a_J)]$, the channel matrix $\mathbf{Q}$, the user ensemble, $\mathbf{B}=[b_1,...,b_J]$, and the user symbol probabilities, $\mathbf{v}=[P(b_1),...,P(b_k)]$ such that

$$\mathbf{v} = \mathbf{Q}\mathbf{z}. \quad (4)$$

or

$$\mathbf{Q} = \begin{bmatrix} P(b_1|a_1) & P(b_1|a_2) & \cdots & P(b_1|a_J) \\ P(b_2|a_1) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ P(b_K|a_1) & \cdots & \cdots & P(b_K|a_J) \end{bmatrix}$$

as shown in Figure 1.

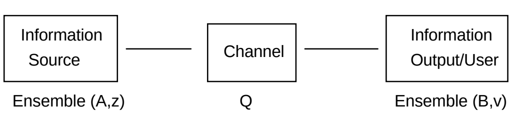


Figure 1. An information system.

The average information of the source, $H(\mathbf{z})$, is the mathematical expectation of the source information

$$H(\mathbf{z}) = -\sum_{j=1}^{J} P(a_j)\log P(a_j) \quad (5)$$

In Figure 2, we show the entropy image of the eyes/nose templates for the front view. Each pixel is treated as an independent information channel.

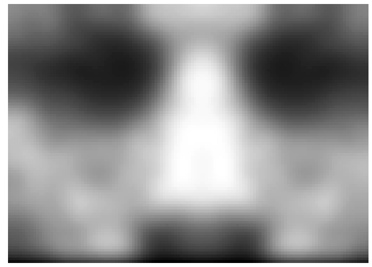


Figure 2. The entropy image of the eyes/nose templates. Average information is 5.16 bits/pixel. Whiter pixels have greater entropy than darker pixels.

The mutual information of **z** and **v** is

$$I(\mathbf{z},\mathbf{v}) = H(\mathbf{z}) - H(\mathbf{z}|\mathbf{v}) \quad (6)$$

or

$$I(\mathbf{z},\mathbf{v}) = \sum_{j=1}^{J}\sum_{k=1}^{K} P(a_j,b_k)\log\frac{P(a_j,b_k)}{P(a_j)P(b_k)} \quad (7)$$

Thus, the mutual information is the average information received upon observing a single output of the information channel.

If we rewrite the Kullback relative information in discrete form, we find that

$$J(q_1,q_0) = \sum_y q_1(y)\log\frac{q_1(y)}{q_0(y|v')} \quad (8)$$

and that the mutual information is a special case of J, when $q_1=P_{ab}$, and $q_0=M_{ab}$ where P_{ab} is the joint distribution of a and b, and M_{ab} is the product of the marginals.

By taking log $q_0(v' \mid v')$ as the estimate of its mean and then removing the bias, we arrive at [8]

$$N \log r + 2k \quad (9)$$

which is Akaike's information criterion.

3 Information theoretic template design

The Kullback relative information measures the class separation between q_0 and q_1. In face detection we define the most informative pixels (MIP) as the ones which maximize the relative information or gives the maximum class separation. Since the number of priors and the dimensionality of the estimates to the distributions increases for every additional pixel, we apply the Markov condition which assumes that the distribution of states for a pixel is only dependent on its neighbors. This results in a first order Markov random field approximation to $q_0(y \mid \{v_1', v_2', ..., v_n'\})$ where $\{v_1', v_2', ..., v_n'\}$ represents the set of known pixels.

3.1 Markov random fields

A subset of Markov random fields are known as Gibbsian random fields [14], which have the conditional probabilities of the form

$$P(s_{k,l} \mid S_{k,l}) = \frac{1}{Z_s} \exp\left[-\frac{1}{T} \sum_j F_j(C_j(k,l)) \right] \tag{10}$$

For our purpose, we use the four-neighbor structure for the clique, $C_j(k,l)$, and the term in the exponent as

$$-\frac{1}{T} s_{k,l} \left[\alpha + \beta_1 (s_{k-1,l} + s_{k+1,l}) + \beta_2 (s_{k,l-1} + s_{k,l+1}) \right] \tag{11}$$

where α, β_1, and β_2 are parameters of the distribution, which depend on the pixel location. By using the most informative pixels, we can maximize equation (3), J, while reducing computational complexity.

3.2 Face detection

In the face detection problem domain our goal is to detect every human face in an image while minimizing the number of false alarms. The algorithm is briefly described below:

(1) For a given number of pixels, N_p, find the set of most informative pixels (MIP) using the Kullback relative information measure.
(2) Use the MIP in obtaining linear features for classification and representation using the method of Fukunaga and Koontz [3]
(3) For every interesting image scale and every 23x32 window, use the distance from feature space metric, DFFS (for more information see Pentland, et al. [7]. When the DFFS to the face cluster is lower than the DFFS to the nonface cluster, then it is assumed that a face is within the window.

Our training database consisted of 9 views of 100 individuals as shown in Figure 3.

Figure 3. Nine views of one individual for the training set.

Face databases from MIT, CMU, and Leiden University were used for testing.

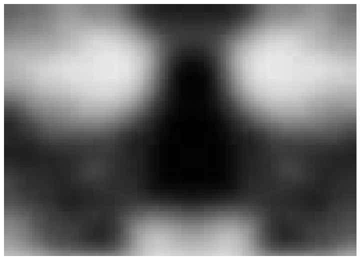
The application of the Kullback relative information to face detection yields the eye/nose template relative information image and the 256 most informative pixels in Figure 4(a), and an example is shown in Figure 4(b).

It is interesting to note that the MIP distribution generally avoids the nose area. This agrees with psychological evidence on human face detection in which several studies showed that the nose plays an insignificant role in face perception and retention [2].

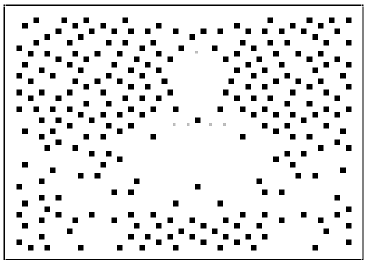
We applied the face detection algorithm to four datasets: the Leiden 19th century portrait database, a Carnegie Mellon University (CMU) database, a database from the Massachusetts Institute of Technology(MIT) and the CMU website database.

The 19th century database is composed of 494 images containing 574 faces. These are portrait images from scanned photos of a wide range of people from the 19th century. There is significant noise in the form of film discoloring, general mishandling, and loss of contrast due to film degradation.

The CMU database consists of 42 scanned photographs with 169 faces. These photographs originated from television broadcasts, newspapers, and magazines. The interesting question created arising from their database is whether hand drawn faces (i.e. a smiley face) should be recognized as a face. Although our face detector was trained only on human faces, we used their ground truth, which assumes that hand drawings are faces. Thus, we would not expect our face detector to perform well on their database.



(a)



(b)

Figure 4. The relative information image (a) of the face class (no priors) with respect to the nonface class, and the 256 most informative pixels (b).

The MIT database is composed of 23 images of group team photos, images of friends, and coworkers with a total of either 149 or 155 faces depending on how the face count is performed. Specifically, MIT labelled 149 faces whereas CMU labelled 155 faces. In order to make benchmarking stable, we always test our algorithm on the ground truth as defined by the creator of the dataset.

The CMU website database is a subset of the images at the CMU WWW face detection website, which allows images to be submitted from any website in the world. It is composed of 71 images and 270 faces, which include scanned photos from magazines, newspapers, personal collections, and TV shows. In Table 1, we show results for each of the international test sets, and in operating characteristic in Figure 5.

Table 1. Results for specific test sets.

	MIP Face Detector Detection % / # False Alarms
Leiden Portrait Database	97.4 / 46
Test Set A - CMU	88.3 / 508
Test Set B - MIT	94.1 / 64
Website Database	84.9 / 13

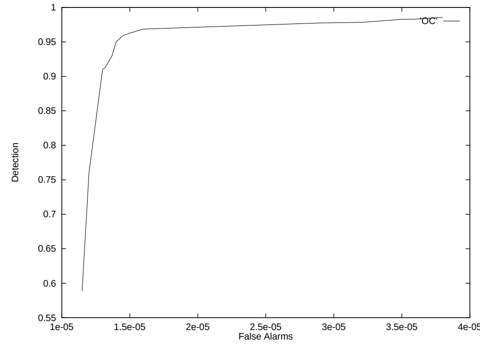
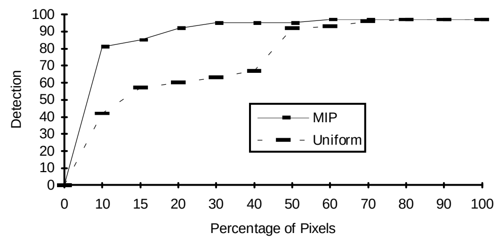
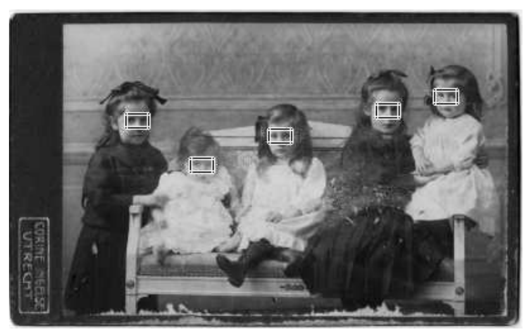


Figure 5. The operating characteristic for the face detector.

The detection rates for the MIP and the uniform distribution are shown in Figure 6 for .01% false alarms. For the same percentage of pixels, the Kullback distribution had greater detection rates relative to the uniform distribution.



(a)



(b)

Figure 6. MIP and uniform distribution (w/o MIP) detection rates with respect to percentage of pixels (a), and an example of face detector output (b).

In Figure 7 and 8, we show the face detector output for images from the CMU database and the CMU website. Note in Figure 7 that the false alarm is roughly similar to a face.

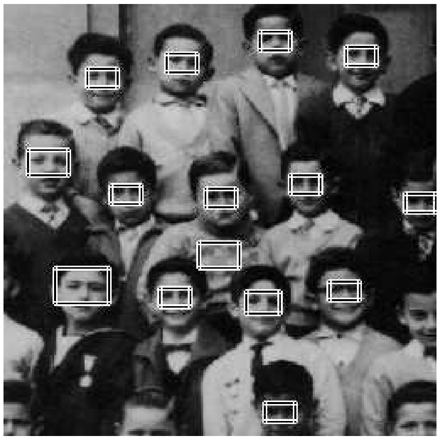


Figure 7. Image 182 from the CMU website

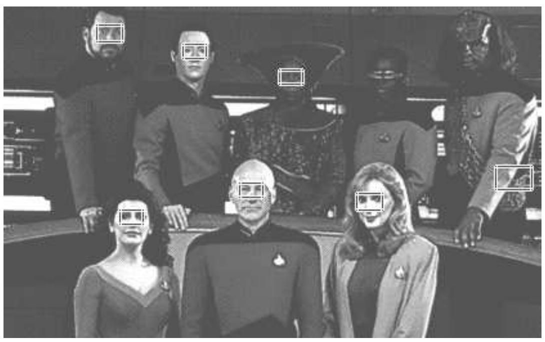


Figure 8. Image "next" from the CMU database.

4 Conclusions

In this paper, we first reviewed the mathematical relationship between the Kullback relative information and the other well known estimation criterions. Second, using the Kullback relative information, a method was described for determining the most informative pixels for template matching. These pixels have the property that they maximize the class separation, which results in lower classification errors. We proposed a view-based face detection method using the most informative pixels, and extensively tested the algorithm on 4 international databases from 3 universities.

Future work will be directed toward detecting side views and using a modular approach toward feature based face detection where each feature is represented by a Kullback distribution.

Acknowledgments

Funding was provided by the Advanced School for Computing and Imaging, Delft, The Netherlands. We would like to thank Kah Kay Sung and Tomaso Poggio; and Henry Rowley, Shumeet Baluja, and Takeo Kanade for their websites and databases.

WWW & demo sites

You can download the MIP face detector from
http://www.wi.leidenuniv.nl/home/mlew/lim.html
The Leiden 19th century portrait database is at
http://ind156b.wi.leidenuniv.nl:8086/intro.html

References

[1] Akaike, H., "Information Theory and an Extension of the Maximum Likelihood Principle," *2nd Int. Symp. on Information Theory*, Armenia, USSR, 1971.

[2] Chellappa, R., C. L. Wilson, S. Sirohey, "Human and Machine Recognition of Faces: A Survey", *Proc. of the IEEE*, vol. 83, no. 5, May 1995.

[3] Fukunaga, F., and W. Koontz, "Applications of the Karhunen-Loeve Expansion to Feature Selection and Ordering," *IEEE Trans. Computers*, vol. C-19, pp. 917-923, 1970

[4] Huang, T. S., and L. Tang, "Face Recognition in Computers," *Beckman Inst. Tech. Rep.*, Sept., 1994.

[5] Kullback, S., "Information Theory and Statistics," Wiley, New York, 1959.

[6] Levine, M., Vision in Man and Machine, McGraw-Hill, New York, 1985.

[7] Pentland, A., B. Moghaddam, and T. Starner. "View-Based and Modular Eigenspaces for Face Recognition," *IEEE Conf. on Comp. Vis. and Patt. Recog.*, pp. 84-91, 1994.

[8] Rissanen, J., "Modeling By Shortest Data Description," *Automatica*, vol. 14, pp. 465-471, 1978.

[9] Rowley, H, and T. Kanade, "Human Face Detection in Visual Scenes," CMU *Tech. Rep. CMU-CS-95-158R*, November, 1995.

[10] Shannon, C, "A Mathematical Theory of Communication," *Bell System Tech. Jnl.*, vol. XXVII, no. 3, pp. 379-423, 1948.

[11] Sung, K. K., and T. Poggio, "Learning Human Face Detection in Cluttered Scenes," *6th Int. Conf. on Comp. Anal. of Images and Patterns*, Prague, pp. 432-439.

[12] Yang, G., and T. S. Huang, "Human Face Detection in a Complex Background," *Patt.Recog.*, 27(1):53-63, 1994.

[13] Yuille, A., P. Hallinan, and D. Cohen, "Feature Extraction from Faces using Deformable Templates," *Int. Jnl. of Comp. Vis.*, 8(2):99-111, 1992.

[14] Geman, S., and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans. Patt. Anal. and Mach. Intell.*, vol. 6, no. 6, July, 1984.