

Typesetting math: 100%

JavaScript is disabled on your browser. Please enable JavaScript to use all the features on this page. [Skip to main content](#)[Skip to article](#)

ScienceDirect

\* Journals & Books

\* Help

\* Search

Gergo Gyori

IT University of Copenhagen

\* View **PDF**

\* Download full issue

Search ScienceDirect

## Outline

1. Abstract
2. 3. Keywords
4. 1\ Introduction
5. 2\ Appearance-based gender classification
6. 3\ Gaussian process classifiers
7. 4\ The EM?EP algorithm for GPCs
8. 5\ Experimental results
9. 6\ Conclusion
10. Acknowledgments
11. References

Show full outline

## Cited by (56)

## Figures (7)

1. 2. 3. 4. 5. 6.

Show 1 more figure

## Pattern Recognition Letters

Volume 27, Issue 6, 15 April 2006, Pages 618-626

# Appearance-based gender classification with Gaussian processes

Author links open overlay panelHyun-Chul Kim a, Daijin Kim b, Zoubin Ghahramani c, Sung Yang Bang b

Show more

Outline

Add to Mendeley

Share

Cite

<https://doi.org/10.1016/j.patrec.2005.09.027>Get rights and content

## Abstract

This paper concerns the gender classification task of discriminating between images of faces of men and women from face images. In appearance-based approaches, the initial images are preprocessed (e.g. normalized) and input into classifiers. Recently, support vector machines (SVMs) which are popular kernel classifiers have been applied to gender classification and have shown excellent performance. SVMs have difficulty in determining the hyperparameters in kernels (using cross-validation). We propose to use Gaussian process classifiers (GPCs) which are Bayesian kernel classifiers. The main advantage of GPCs over SVMs is that they determine the hyperparameters of the kernel based on Bayesian model selection criterion. The experimental results show that our methods outperformed SVMs with cross-validation in most of data sets. Moreover, the kernel hyperparameters found by GPCs using Bayesian methods can be used to improve SVM performance.

\* Previous article in issue

\* Next article in issue

## ## Keywords

Gender classification

Appearance-based gender classification

Kernel machines

Gaussian process classifiers

Support vector machines

## ## 1\ Introduction

The face is a characteristic feature of human beings which contains identity and emotion. It is possible to identify a person and her/his characteristics such as emotion (or expression) and gender from her/his face. Recognizing human gender is important since lots of social interactions and services depend on the gender. People respond differently according to gender. Human computer interaction system can be more user-friendly and more human-like when it considers the user's gender.

There are two main approaches for gender classification. The first approach is the appearance-based approach which uses a whole face image. Cottrell and Metcalfe (1991) reduced the dimension of whole face images by autoencoder network and classified gender based on the reduced input features. Golomb et al. (1991) used a two-layer neural network (called SexNet) without dimensionality reduction. Tamura et al. (1996) used a neural network and showed that even very low resolution image such as  $8 \times 8$  can be used for gender classification. Gutta et al. (2000) used the mixture of experts with ensembles of radial basis functions (RBF) networks and a decision tree as a gating network. Moghaddam and Yang (2002) showed that support vector machines (SVMs) worked better than other classifiers such as ensemble of RBF networks, classical RBF networks, Fisher linear discriminant, nearest neighbor etc. Jain and Huang (2004) extracted wholistic features by independent component analysis (ICA) and classified it with linear discriminant analysis (LDA). Costen et al. (2004) used the exploratory basis pursuit classification which is a sparse kernel classifier.

The second approach is the geometrical feature based approach. Burton et al. (1993) extracted point-to-point distances from 73 points on face images and used discriminant analysis as a classifier. Brunelli and Poggio (1992) extracted 16 geometric features such as eyebrow thickness and pupil-to-eyebrow distance and used HyperBF networks as a classifier.

As mentioned above, the appearance-based approach with SVM showed excellent performance (Moghaddam and Yang, 2002). In their experiments the Gaussian kernel worked better than linear or polynomial kernels. They did not mention how to set the hyperparameters<sup>1</sup> for Gaussian kernel which have an influence on performance, but just showed the test results with several different hyperparameters. Learning the hyperparameters should be included in the training process. A standard way to determine the hyperparameters is by cross-validation. Alternatively we could use kernel classifiers such as Gaussian process classifiers which automatically incorporate method to determine the hyperparameters. In this paper we propose to use Gaussian process classifiers (GPCs) for appearance-based gender classification.

GPCs are a Bayesian kernel classifier derived from Gaussian process priors over functions which were developed originally for regression (O'Hagan, 1978, Neal, 1997, Williams and Barber, 1998, Gibbs and MacKay, 2000). In classification, the target values are discrete class labels. To use Gaussian processes for binary classification, the Gaussian process regression model can be modified so that the sign of the continuous latent function it outputs determines the class label. Observing the class label at some data point

constrains the function value to be positive or negative at that point, but leaves it otherwise unknown. To compute predictive quantities of interest we therefore need to integrate over the possible unknown values of this function at the data points.

Exact evaluation of this integral is computationally intractable. However, several successful methods have been proposed for approximately integrating over the latent function values, such as the Laplace approximation (Williams and Barber, 1998), Markov Chain Monte Carlo (Neal, 1997), and variational approximations (Gibbs and MacKay, 2000). Opper and Winther (2000) used the TAP2 approach originally proposed in statistical physics of disordered systems to integrate over the latent values. The TAP approach for this model is equivalent to the more general expectation propagation (EP) algorithm for approximate inference (Minka, 2001). The expectation maximization?expectation propagation (EM?EP) algorithm has been proposed to learn the hyperparameters based on EP (Kim and Ghahramani, 2003). GPCs with the hyperparameters obtained by the EM?EP algorithm have shown better performance than SVMs which had the hyperparameters set by cross-validation, on most of data sets tested. In many cases the hyperparameters determined by the EM?EP algorithm were more suitable for SVMs than the ones determined by cross-validation technique. In this paper we use the EM?EP algorithm to learn Gaussian process classifiers for gender classification. We expect that GPCs with the EM?EP algorithm work better than SVMs with the cross-validation and provide better hyperparameters for the kernels of SVMs.

The paper is organized as follows. Section 2 introduces appearance-based gender classification. In Section 3, we introduce Gaussian process classification. In Section 4, we describe the EP method and the EM?EP algorithm for Gaussian process classification. In Section 5, we show experimental results on the PF01 database and compared with other classification methods including SVMs. In Section 6, we draw conclusions and remark on future work.

## ## 2\ Appearance-based gender classification

The appearance-based approach to gender classification discriminates between male and female classes from face images without first explicitly extracting any geometrical features. A typical way to do this is to train a classifier with training images and to classify new images by the trained classifier. Face images should be well-aligned so that facial features are in the same positions. Since gender classification is a two-class classification problem, any kind of binary classifier can be deployed.

Fig. 1 shows the process of appearance-based gender classification. Assume that a classifier has been already trained with some images in advance. The whole process of gender classification can be explained by the following. First, images are captured. Then, the captured images are preprocessed by face detection and facial feature extraction algorithms and cropped by an appropriate cropping technique. The preprocessed face images can include a whole outline of faces with hair or can include only inner face parts with only facial features. Then, the preprocessed image (pixel-level features) is applied to the classifier and the classifier determines the gender of the input image.

### 1. Download: Download full-size image

Fig. 1. The process of appearance-based gender classification.

The appearance-based approach has two main advantages. First, it preserves appearance of face images which can be considered to be naive features. It is difficult to determine what kind of geometrical features we should use and to tell the meaning of those features. In contrast to this, appearance-based

approach is more natural since it uses face images themselves. The benefit in being natural is that it could make it easier to do what the natural being does. Second, it does not need to extract facial features or points very accurately. To get good geometrical features, we need to know quite accurate facial feature or point locations which requires accurate facial feature extraction. In contrast to this, we need to know relatively small number of facial features for alignment in the appearance-based approach. The disadvantages of the appearance-based approach is that it has more features than the geometrical feature based approach and that it does not provide a good explanation why a facial image is classified as a male or female. We follow the above process for appearance-based gender classification and use Gaussian process classifiers.

### ## 3\ Gaussian process classifiers

Let us assume that we have a data set  $\mathcal{D}$  of data points  $\mathbf{x}_i$  with binary class labels  $y_i \in \{1, -1\}$ :  $\mathcal{D} = \{(\mathbf{x}_i, y_i) \mid i = 1, 2, \dots, n\}$ ,  $\mathcal{X} = \{\mathbf{x}_i \mid i = 1, 2, \dots, n\}$ ,  $\mathcal{Y} = \{y_i \mid i = 1, 2, \dots, n\}$ . Given this data set, we wish to find the correct class label for a new data point. We do this by computing the class probability. We assume that the class label is obtained by transforming some real valued latent variable, which is the value of some latent function  $f(\cdot)$  evaluated at  $\cdot$ . We put a Gaussian process prior on this function, meaning that any number of points evaluated from the function have a multivariate Gaussian density (see Williams and Rasmussen (1995) for a review of GPs). Assume that this GP prior is parameterized by  $\theta$  which we will call the hyperparameters. We can write the probability of interest given  $\theta$  as (1) This is the probability of the class label at a new data point given data  $\mathcal{D}$  and hyperparameters  $\theta$ . The second part of Eq. (1) is obtained by further integration over  $\mathbf{f} = [f_1, f_2, \dots, f_n]$ , the values of the latent function at the data points. (2) where and (3) The first term in Eq. (3) is the likelihood: the probability for each observed class given the latent function value, while the second term is the GP prior over functions evaluated at the data. Writing the dependence of  $\mathbf{f}$  on  $\mathbf{x}$  implicitly, the GP prior over functions can be written (4) where the mean  $\mathbf{m}$  is usually assumed to be the zero vector  $\mathbf{0}$  and each term of a covariance matrix  $C_{ij}$  is a function of  $\mathbf{x}_i$  and  $\mathbf{x}_j$ , i.e.  $C(\mathbf{x}_i, \mathbf{x}_j)$ . One form for the likelihood term  $p(y_i \mid f_i, \theta)$ , which relates  $f(\mathbf{x}_i)$  monotonically to probability of  $y_i = +1$ , is (5) Other possible forms for the likelihood are a sigmoid function  $1/(1 + \exp(-y_i f(\mathbf{x}_i)))$ , a step function  $H(y_i f(\mathbf{x}_i))$ , and a step function with a labelling error  $\frac{1}{2} + (1 - 2H(y_i f(\mathbf{x}_i)))$ .

Since  $p(\mathbf{f} \mid \mathcal{D}, \theta)$  in Eq. (3) is intractable due to the non-linearity in the likelihood terms, we use an approximate method. Laplace approximation, variational methods and Markov Chain Monte Carlo method were used in (Williams and Barber, 1998, Gibbs and MacKay, 2000, Neal, 1997), respectively. Expectation propagation, which is described in the next section, was used in (Opper and Winther, 2000, Minka, 2001).

### ## 4\ The EM/EP algorithm for GPCs

#### ### 4.1. Expectation propagation for GPCs

The expectation-propagation (EP) algorithm is an approximate Bayesian inference method (Minka, 2001). We review EP in its general form before describing its application to GPCs.

Consider a Bayesian inference problem where the posterior over some parameter  $\theta$  is proportional to the prior times likelihood terms for an i.i.d. data

set(6) We approximate this by (7) where each term (and therefore  $q_i$ ) is assumed to be in the exponential family. EP successively solves the following optimization problem for each  $i$  (8) where KL is the Kullback-Leibler divergence and (9) Since  $q_i$  is in the exponential family, this minimization is solved by matching moments of the approximated distribution. EP iterates over  $i$  until convergence. The algorithm is not guaranteed to converge although it did in practice in all our examples and has worked well for many other authors. Assumed density filtering (ADF) is a special online form of EP where only one pass through the data is performed ( $i = 1, \dots, n$ ).

We describe EP for GPC referring to (Minka, 2001, Opper and Winther, 2000). The latent function  $f$  plays the role of the parameter  $\theta$  above. The form of the likelihood we use in the GPC is (10) where  $H(x) = 1$  if  $x > 0$ , and otherwise 0. The hyperparameter,  $\lambda$  in Eq. (10) models labeling error outliers. The EP algorithm approximates the posterior  $p(f | D) = p(f) p(D | f) / p(D)$  as a Gaussian having the form  $\mathcal{N}(f | \mu, \Sigma)$ , where the GP prior has covariance matrix  $C$  with elements  $C_{ij}$  defined by the covariance function (11) where  $x_{im}$  is the  $m$ -th element of  $x_i$ , and if  $x_{im}$  is continuous; if  $x_i$  is discrete, where  $\delta$  is 1 if and 0 if  $i \neq j$ . The hyperparameter  $v_0$  specifies the overall vertical scale of variation of the latent values,  $v_1$  the overall bias of the latent values from zero mean,  $v_2$  the latent noise variance, and  $l_m$  the (inverse) lengthscale for feature dimension  $m$ . The erf likelihood term in Eq. (5) is equivalent to using the threshold function in Eq. (10) with  $\lambda = 0$  and non-zero latent noise  $v_2$ . EP tries to approximate  $p(f | D)$ , where  $(0, C)$ .  $p(y_i | f) = t_i(f)$  is approximated by  $\mathcal{N}(f | \mu_i, \Sigma_i)$ . From this initial setting, we can derive EP for GPC by applying the general idea described above. The resulting EP procedure is virtually identical to the one derived in (Minka, 2001). We define the following notation:  $\mathbf{z}_i$ , where  $\mathbf{z}$  and  $i$  are quantities obtained from a whole set except for  $x_i$ . The EP algorithm is as follows which we repeat for completeness: please refer to Minka (2001) for the details of the derivation. After the initialization  $v_i = 1$ ,  $m_i = 0$ ,  $s_i = 1$ ,  $h_i = 0$ ,  $\mu_i = C_{ii}$ , the following process is performed until all  $(m_i, v_i, s_i)$  converge.

Loop  $i = 1, 2, \dots, n$ :

\* (1)

Remove the approximate density (for  $i$ -th data point) from the posterior to get an "old" posterior:  $p_{old}$ .

\* (2)

Recompute part of the new posterior:  $p_{new}$ , where  $\text{erf}(z)$  is a cumulative normal density function.

\* (3)

Get a new  $\mu_i$ .

\* (4)

Now that  $v_i$  is updated, finish recomputing the new posterior:  $A = (C^{-1} + \lambda^{-1} \Sigma_i^{-1})^{-1}$ ; For all  $i$ ,  $\mu_i = A \Sigma_i^{-1} \mu_i$ .

Our approximated posterior over the latent values is: (12) where (or).

Classification of a new data point can be done according to  $\mu_i$ .

The approximate evidence can be obtained as (13) where  $\mathbf{z}$ . The approximate evidence in Eq. (13) can be used to evaluate the feasibility of kernels or their hyperparameters to the data. But, it is tricky to get a hyperparameter updating rule from Eq. (13). In the following section, we derive the algorithm to find the hyperparameters automatically based not in Eq. (13) but a variational lower bound of the evidence.

### 4.2. The EM/EP algorithm

EP for GPCs propose a method to estimate latent values but not hyperparameters. We put  $\mu = \mu \text{ cov } \Sigma$ , and  $\mu \text{ cov } = \{\mu_0, \mu_1, \mu_2\} \times \{l_p \times p = 1, 2, \dots, d\}$  for the hyperparameters. Here we present the EM<sup>2</sup>EP algorithm based on EP to estimate both latent values and hyperparameters (Kim and Ghahramani, 2003). We tackle the problem of learning the classifier hyperparameters as one of optimizing hyperparameters for Gaussian process regression with hidden target values. This idea makes it possible to apply an approximate EM (expectation maximization) algorithm. In the E-step, we infer the approximate (Gaussian) density for latent function values  $q(f)$  using EP. In the M-step, using  $q(f)$  obtained in the E-step, we maximize the variational lower bound of  $p(Y | X, \mu, \Sigma)$ . The E-step and M-step are alternated until convergence.

**E-step:** EP iterations are performed given the hyperparameters.  $p(f | D)$  is approximated as a Gaussian density  $q(f)$  given by Eq. (12).

**M-step:** Given  $q(f)$  obtained from the E-step, find the hyperparameters which maximize the variational lower bound of  $p(Y | X, \mu, \Sigma) = \int p(Y | f, X, \mu, \Sigma) p(f | X, \mu, \Sigma) d f$ . Since the above integral is intractable, we take a variational lower bound  $F$  as follows:(14)

Using the E-step result Eq. (12) and the definition of  $F$ , we obtain the following gradient update rule with respect to the covariance hyperparameters(15)

(See Kim and Ghahramani, 2003 for the derivation of the M-step.)

We found that in practice EM<sup>2</sup>EP always converged and the local maxima were good solutions. EM<sup>2</sup>EP has a complexity of  $O(n^3)$  due to the matrix inversion in EP.

## 5. Experimental results

We performed experiments on appearance-based gender classification with Gaussian processes using the database PF01 (Postech Faces 2001) (Kim et al., 2001) and Aleix database (Martinez and Benavente, 1998). The database PF01 has color face images of 103 Asian people, 53 men and 50 women, where for each person there are 17 images under various conditions (one normal, four illumination-varying ones, eight pose-varying ones, four expression-varying ones). The Aleix database has over 4000 color images of 126 people's faces (70 men and 56 women), where images are frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). We performed gender classification on four partial data sets which are only normal face images (103 images, Faceset PF-I) in PF01, normal and expression-varying face images ( $5 \times 103 = 515$  images, Faceset PF-II) in PF01, only normal face images (126 images, Faceset AL-I) in PF01, and normal and expression-varying face images ( $4 \times 126 = 504$  images, Faceset AL-II) in PF01. Fig. 2, Fig. 3 show the normal and expression-varying images of three men and three women in the database PF01 and the Aleix database, respectively. For each partial data set, we preprocessed face images in two ways. The first form of preprocessing downsampled and cropped face images including hairs and contour of faces and the second form of preprocessing further cropped the face images to exclude hair and background. Fig. 4 shows the example of a normalized image ( $256 \times 256$ ) and a cropped face image ( $28 \times 23$ , cropped type A) and a more cropped face image ( $20 \times 16$ , cropped type B). All images are aligned so that eyes are placed in the same positions, which can be done with eye detection algorithms in practice. If images are not aligned well, the appearance-based method would not work well.

1. Download: Download full-size image

Fig. 2. Some images in the database PF01.

1. Download: [Download full-size image](#)

Fig. 3. Some images in the Aleix database.

1. Download: [Download full-size image](#)

Fig. 4. Preprocessed images: (a) Normalized image, (b) downsampled and cropped image ( $28 \times 23$ ), (c) downsampled and more cropped image ( $20 \times 16$ ).

We have eight different data sets: data set P-I (Faceset PF-I, cropped type A), data set P-II (Faceset PF-I, cropped type B), data set P-III (Faceset PF-II, cropped type A), and data set P-IV (Faceset PF-II, cropped type B), data set A-I (Faceset AL-I, cropped type A), data set A-II (Faceset AL-I, cropped type B), data set A-III (Faceset AL-II, cropped type A), and data set A-IV (Faceset AL-II, cropped type B). Data set P-I, P-II, P-III and P-IV are the data sets which include normal faces, more cropped normal faces, expression-varying faces, and more cropped expression-varying faces from the database PF01, respectively. Data set A-I, A-II, A-III and A-IV are the data sets which include normal faces, more cropped normal faces, expression-varying faces, and more cropped expression-varying faces from the Aleix database, respectively. On these data sets, we applied many different classifiers including one nearest neighbor (1-NN), linear discriminant analysis (LDA), SVM with cross-validation (SVM-CV), SVM with EM?EP hyperparameters (SVM-EP), and GPC with the EM?EP algorithm (GPC-EP). Fig. 5, Fig. 6 show the classification error rates of these methods over four different data sets in the database PF01 and Aleix database. Each data set was divided into 10 folds. Each fold was subsequently used as a test set, while the other nine folds were used as a training set. Before GPC or SVM are applied, all feature values are normalized based on the training set so that their means are zero and their variances are one. The points  $\bar{x}$  in Fig. 5, Fig. 6 are means of 10 trials and error bars are from standard deviations of the mean estimators. In Fig. 7, we show misclassified images. In Fig. 7(a), we can know that long hair is not always a key feature of women. The classification rates of data sets of cropped type A are not always better than ones of cropped type B.

1. Download: [Download full-size image](#)

Fig. 5. Classification error rates of various methods for four kinds of gender classification data sets (PF?01 DB): (a) data set P-I, (b) data set P-II, (c) data set P-III, (d) data set P-IV.

1. Download: [Download full-size image](#)

Fig. 6. Classification error rates of various methods for four kinds of gender classification data sets (Aleix DB): (a) data set A-I, (b) data set A-II, (c) data set A-III, (d) data set A-IV.

1. Download: [Download full-size image](#)

Fig. 7. Misclassified facial images: (a) data set P-I, (b) data set P-II.

GPC-EP used a single lengthscale hyperparameter (i.e.  $\_l\_m\_ = \_l\_$ ) for all feature dimensions<sup>4</sup>. In all GPC models the hyperparameter  $\_?\_$  was not updated but fixed to zero. In SVM-EP the kernel (i.e. covariance function) had the same hyperparameters as the corresponding GPC-EP that were trained using EM?EP except for the latent noise variance  $\_v\_2$  which was omitted because it caused degradation in SVM performance<sup>5</sup>. Instead, the penalty parameter  $\_C\_$  allowing training errors (i.e. penalizing the SVM slack variables) was selected by 5-fold cross-validation.<sup>6</sup> In SVM-CV we applied SVMs with a Gaussian kernel with a single lengthscale hyperparameter (without  $\_v\_0$ ,  $\_v\_1$  and  $\_v\_2$ ) selected by five-fold cross-validation.<sup>7</sup> We also had to determine the penalty parameter  $\_C\_$ , so we performed a 2-level grid search over a 2-dimensional parameter space ( $\_C\_$ ,  $\_l\_$ )<sup>8</sup>.

In the data set P-I, P-II, P-IV, and A-IV, GPC-EP is the best, in the data set P-III, A-I, and A-II, SVM-EP is the best, and in the data set A-III SVM-CV is

the best. In all the data sets except for one, GPC-EP or SVM-EP is the best. Also, in all the data sets except for one, SVM-EP is better than SVM-CV. Therefore, for the data sets tested the hyperparameters found by the EM<sup>2</sup>EP algorithm seem to be also more suitable hyperparameters for SVMs than the ones obtained by cross-validation. This shows that the EM<sup>2</sup>EP algorithm finds suitable hyperparameters successfully and those hyperparameters are also suitable for SVMs. This result is consistent with the result on the benchmark data sets in (Kim and Ghahramani, 2003).

## ## 6\ Conclusion

We have proposed the appearance-based gender classification method with Gaussian processes. GPCs incorporate the Bayesian model selection framework to determine the kernel hyperparameters, which is an important advantage over SVMs. In the experiments the hyperparameters obtained by GPC with the EM<sup>2</sup>EP algorithm were even more suitable for SVMs than the ones obtained by cross-validation. In most of the data sets, EM<sup>2</sup>EP algorithms worked better than SVMs with cross-validation and provided kernel hyperparameters to make SVMs work better.

We used Gaussian kernels in this paper. Gaussian kernels do not seem to be ideal for image data since they do not capture correlations between pixels. If we invent more proper kernels for face images, we might improve the performance. It would also be interesting to perform experiments on a larger face data set.

## ## Acknowledgments

Hyun-Chul Kim, Daijin Kim and Sung-Yang Bang would like to thank the Ministry of Education of Korea for its financial support toward the Division of Mechanical and Industrial Engineering, and the Division of Electrical and Computer Engineering at POSTECH through BK21 program.

## Recommended articles

## ## References

1. Brunelli and Poggio, 1992  
Brunelli, R., Poggio, T., 1992. Hyperbf networks for gender classification.  
In: DARPA Image Understanding Workshop.  
Google Scholar
2. Burton et al., 1993  
A. Burton, V. Bruce, N. Dench  
What's the difference between men and women? Evidence from facial measurements  
Perception, 22 (1993), pp. 153-176  
CrossrefView in ScopusGoogle Scholar
3. Costen et al., 2004  
Costen, N., Brown, M., Akamatsu, S., 2004. Sparse models for gender classification. In: Proceedings of the 6th International Conference on Automatic Face and Gesture Recognition.  
Google Scholar
4. Cottrell and Metcalfe, 1991  
G. Cottrell, J. Metcalfe  
EMPATH: Face, emotion, and gender recognition using holons  
Advances in Neural Information Processing Systems 3, vol. 3, MIT Press (1991)  
Google Scholar
5. Gibbs and MacKay, 2000  
M. Gibbs, D.J.C. MacKay  
Variational Gaussian process classifiers  
IEEE Trans. NN, 11 (6) (2000), p. 1458  
View in ScopusGoogle Scholar
6. Golomb et al., 1991



B. Golomb, D. Lawrence, T. Sejnowski  
SEXNET: A neural network identifies sex from human faces  
Advances in neural information processing systems 3, vol. 3, MIT Press (1991)  
Google Scholar

7. Gutta et al., 2000  
S. Gutta, J.R.J. Huang, P. Jonathon, H. Wechsler  
Mixture of experts for classification of gender, ethnic origin, and pose of human faces  
IEEE Trans. Neural Networks, 11 (4) (2000), pp. 948-960  
View in ScopusGoogle Scholar

8. Jain and Huang, 2004  
Jain, A., Huang, J., 2004. Integrating independent components and linear discriminant analysis. In: Proceedings of the 6th International Conference on Automatic Face and Gesture Recognition.  
Google Scholar

9. Kim and Ghahramani, 2003  
Kim, H.-C., Ghahramani, Z., 2003. The EM<sup>2</sup>EP algorithm for Gaussian process classification. In: Proceedings of the Workshop on Probabilistic Graphical Models for Classification (ECML), pp. 37-48.  
Google Scholar

10. Kim et al., 2001  
Kim, H.-C., Sung, J.-W., Je, H.-M., Kim, S.-K., Jun, B.-J., Kim, D., Bang, S.-Y., 2001. Asian face image database PF01. Technical Report, Intelligent Multimedia Lab, Department of CSE, POSTECH.  
Google Scholar

11. Martinez and Benavente, 1998  
Martinez, A., Benavente, R., 1998. The ar face database. CVC Technical Report #24.  
Google Scholar

12. Minka, 2001  
Minka, T., 2001. A family of algorithms for approximate Bayesian inference. Ph.D. Thesis, MIT.  
Google Scholar

13. Moghaddam and Yang, 2002  
B. Moghaddam, M. Yang  
Learning gender with support faces  
IEEE Trans. Pattern Anal. Mach. Intell., 24 (5) (2002), pp. 707-711  
View in ScopusGoogle Scholar

14. Neal, 1997  
Neal, R., 1997. Monte Carlo implementation of Gaussian process models for Bayesian regression and classification. Technical Report CRG-TR-97-2, Department of Computer Science, University of Toronto.  
Google Scholar

15. O'Hagan, 1978  
A. O'Hagan  
On curve fitting and optimal design for regression  
J. Royal Stat. Soc. B, 40 (1978), pp. 1-32  
CrossrefGoogle Scholar

16. Opper and Winther, 2000  
M. Opper, O. Winther  
Gaussian processes for classification: Mean field algorithms  
Neural Comput., 12 (2000), pp. 2655-2684  
View in ScopusGoogle Scholar

17. Tamura et al., 1996

S. Tamura, Kawai, H. Mitsumoto

Male/female identification from 8 × 6 very low resolution face images by neural networks

Pattern Recogn., 29 (2) (1996), pp. 331-335

[View PDF](#)[View article](#)[View in Scopus](#)[Google Scholar](#)

18. Williams and Barber, 1998

C.K.I. Williams, D. Barber

Bayesian classification with Gaussian processes

IEEE Trans. PAMI, 20 (1998), pp. 1342-1351

[View in Scopus](#)[Google Scholar](#)

19. Williams and Rasmussen, 1995

C.K.I. Williams, C.E. Rasmussen

Gaussian processes for regression

NIPS 8, vol. 8, MIT Press (1995)

[Google Scholar](#)

## Cited by (56)

\* ### Soft biometrics for surveillance: An overview

2013, Handbook of Statistics

Citation Excerpt :

Some feature selection algorithms (Khan et al., 2005) have also been used to select gender specific features. Most gender classifiers are based on Neural Network (Golomb et al., 1990; Khan et al., 2005), Adaboost (Baluja and Rowley, 2007; Yang and Ai, 2007; Sun et al., 2006; Shan, 2010), Gaussian Process (GP) classifier (Kim et al., 2006), and SVM (Moghaddam and Yang, 2002; Guo et al., 2009; Chen and Ross, 2011). A systematic overview of methods for gender classification from face images in the visible and thermal spectra can be found in Makinen and Raisamo (2008) and Chen and Ross (2011), respectively.

[Show abstract](#)

Biometrics is the science of automatically recognizing people based on physical or behavioral characteristics such as face, fingerprint, iris, hand, voice, gait, and signature. More recently, the use of soft biometric traits has been proposed to improve the performance of traditional biometric systems and allow identification based on human descriptions. Soft biometric traits include characteristics such as height, weight, body geometry, scars, marks, and tattoos (SMT), gender, etc. These traits offer several advantages over traditional biometric techniques. Soft biometric traits can be typically described using human understandable labels and measurements, allowing for retrieval and recognition solely based on verbal descriptions. Unlike many primary biometric traits, soft biometrics can be obtained at a distance without subject cooperation and from low quality video footage, making them ideal for use in surveillance applications. This chapter will introduce the current state of the art in the emerging field of soft biometrics.

\* ### What else does your biometric data reveal? A survey on soft biometrics

2016, IEEE Transactions on Information Forensics and Security

\* ### Recognizing human gender in computer vision: A survey

2012, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)

\* ### Gaussian processes for object categorization

2010, International Journal of Computer Vision

\* ### Gender classification using local directional pattern (LDP)

2010, Proceedings - International Conference on Pattern Recognition

\* ### Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion

2009, IEEE Transactions on Pattern Analysis and Machine Intelligence

[View all citing articles on Scopus](#)

1

Hyperparameters control properties of the kernel and the amount of classification noise.

2

TAP is an abbreviation of its developers' names such as Thouless, Anderson and Palmer.

3

$\text{diag}(\_v\_1, ?, \_v\_n)$  means a diagonal matrix whose diagonal elements are  $\_v\_1, ?, \_v\_n$ . Similarly for  $\text{diag}(\_v\_1, \_v\_2, \dots, \_v\_n)$ .

4

The initial values of hyperparameters for the first fold were as follows:  $\_v\_1 = 0.001$ ,  $\_v\_2 = 1/(2 \times d)$ ,  $\_v\_3 = m$ , and those for subsequent folds are the results for the former fold.

5

For SVMs, we used the MATLAB Support Vector Machine Toolbox available from <http://theoval.sys.uea.ac.uk/~gcc/svm/toolbox> with modified kernel functions.

6

Firstly, we did a coarse grid search over  $\{C | \log_{10} C = 0, 0.5, 1, 1.5, 2, 2.5, 3\}$  to obtain  $\_C\_1$ . Then did a finer grid search over  $\{C | \log_{10} C = -0.4 + \log C_1, -0.3 + \log C_1, \dots, 0.4 + \log C_1\}$ .

7

Similarly to the selection of  $\_C\_$ , we did a 2-level grid search over  $\{l | \log_{10} l = -3, -2.5, -2, -1.5, -1, -0.5, 0\}$  and  $\{l | \log_{10} l = -0.4 + \log l_1, -0.3 + \log l_1, \dots, 0.4 + \log l_1\}$ .

8

The same grids as above for parameters  $\_C\_$ ,  $\_l\_$  were used.

View Abstract

Copyright © 2005 Elsevier B.V. All rights reserved.

## Recommended articles

\* ### Bromelain from *Ananas comosus* stem attenuates oxidative toxicity and testicular dysfunction caused by aluminum in rats

Journal of Trace Elements in Medicine and Biology, Volume 62, 2020, Article 126631

Ali B. Jebur, ?, Wenyi Kang

View PDF

\* ### An on-line HPLC system for detection of antioxidant compounds in some plant extracts by comparing three different methods

Journal of Chromatography B, Volume 1052, 2017, pp. 66-72

Nesibe Arslan Burnaz, ?, Zeynep Akar

View PDF

\* ### Marine oxygenation, lithistid sponges, and the early history of Paleozoic skeletal reefs

Earth-Science Reviews, Volume 181, 2018, pp. 98-121

Jeong-Hyun Lee, Robert Riding

View PDF

\* ### Mitochondrial targeted rhodium(III) complexes: Synthesis, characterized and antitumor mechanism investigation

Journal of Inorganic Biochemistry, Volume 218, 2021, Article 111400

Yan-Bo Peng, ?, Ping Zhao

View PDF

\* ### Bayesian optimization of multiclass SVM for efficient diagnosis of erythemato-squamous diseases

Biomedical Signal Processing and Control, Volume 71, Part B, 2022, Article 103223

Alaa M. Elsayad, ?, Mujahed Al-Dhaifallah

View PDF

\* #### Thirty years of clear-cutting maintain diversity and functional composition of woody-encroached Neotropical savannas

Forest Ecology and Management, Volume 494, 2021, Article 119356

André L. Giles, ?, Osmar Cavassan

[View PDF](#)

[Show 3 more articles](#)

## Article Metrics

Citations

\* Citation Indexes: 56

Captures

\* Readers: 32

[View details](#)

\* [About ScienceDirect](#)

\* [Remote access](#)

\* [Shopping cart](#)

\* [Advertise](#)

\* [Contact and support](#)

\* [Terms and conditions](#)

\* [Privacy policy](#)

Cookies are used by this site. [Cookie Settings](#)

All content on this site: Copyright © 2024 Elsevier B.V., its licensors, and contributors. All rights are reserved, including those for text and data mining, AI training, and similar technologies. For all open access content, the Creative Commons licensing terms apply.

## Cookie Preference Center

We use cookies which are necessary to make our site work. We may also use additional cookies to analyse, improve and personalise our content and your digital experience. For more information, see our [Cookie Policy](#) and the list of [Google Ad-Tech Vendors](#).

You may choose not to allow some types of cookies. However, blocking some types may impact your experience of our site and the services we are able to offer. See the different category headings below to find out more or change your settings.

Allow all

#### Manage Consent Preferences

##### Strictly Necessary Cookies

Always active

These cookies are necessary for the website to function and cannot be switched off in our systems. They are usually only set in response to actions made by you which amount to a request for services, such as setting your privacy preferences, logging in or filling in forms. You can set your browser to block or alert you about these cookies, but some parts of the site will not then work. These cookies do not store any personally identifiable information.

[Cookie Details List?](#)

##### Functional Cookies

Functional Cookies

These cookies enable the website to provide enhanced functionality and personalisation. They may be set by us or by third party providers whose services we have added to our pages. If you do not allow these cookies then some or all of these services may not function properly.

[Cookie Details List?](#)

##### Performance Cookies

Performance Cookies

These cookies allow us to count visits and traffic sources so we can measure

and improve the performance of our site. They help us to know which pages are the most and least popular and see how visitors move around the site.

Cookie Details List?

#### Targeting Cookies

Targeting Cookies

These cookies may be set through our site by our advertising partners. They may be used by those companies to build a profile of your interests and show you relevant adverts on other sites. If you do not allow these cookies, you will experience less targeted advertising.

Cookie Details List?

Back Button

### Cookie List

Search Icon

Filter Icon

Clear

checkbox label label

Apply Cancel

Consent Leg.Interest

checkbox label label

checkbox label label

checkbox label label

Confirm my choices