

# Bayesian object matching based on MCMC sampling and Gabor filters

Jouko Lampinen, Toni Tamminen, Timo Kostiainen and Ilkka Kalliomäki

Laboratory of Computational Engineering  
Helsinki University of Technology  
P.O.Box 9400, FIN-02015 ESPOO, FINLAND

## ABSTRACT

We study an object recognition system where Bayesian inference is used for estimating the probability distribution of matching object locations on an image. The representation of the object contains two parts: the likelihood part that defines the probability of perceiving a given (gray scale) image corresponding to the matched object detail, and the prior part that defines the probability of variation of the object, including elastic distortions and interclass variations.

The application we are studying is related to recognition of faces and recovering the shape of human heads. The prior consists of covariance model for the face, which encodes the correlations of matching point locations over different face images. The approach is closely related to eigenshapes and linear object classes. In the likelihood part, measurement of point correspondence is based on Gabor filter responses. In addition to matching filter amplitudes, we compute also the discrepancy of the filter phases, which increases the spatial accuracy of the matched locations. The similarity is based on measuring the distribution of Gabor filter responses in a given object location, and finding a low dimensional subspace to model the variation. The probability of match is proportional to the angle between the subspace and the vector of Gabor filter responses.

The object matching is carried out as Bayesian inference: the goal is to find the posterior probability distribution of the possible matches given the image and the object prior. We use Markov Chain Monte Carlo (MCMC) methods, such as Gibbs and Metropolis sampling, to estimate the posterior probability of matches. In the paper we present MCMC methods suitable for the matching task and present some preliminary results.

## 1. INTRODUCTION

Recognition of 2D and 3D objects from images is an inherently ill-posed problem, as there are much more unknown attributes in the objects than what can be estimated from the image. First of all, the image contains no direct information of any dimensions in the view direction. Also, for any image there are infinite number of combinations of surface texture and shape that produce exactly the observed image. Further, many objects or classes of objects have considerable internal variations making it difficult to decide when to images display the same object (such as elastic distortions of one person's face, and variations between all objects in the general class 'face'). In scene analysis, where the image can contain several objects, additional tasks are detection and segmentation of objects, and dealing with background clutter, occlusions, etc.

The underdetermined nature of the problem means that there is no unique solution for estimating the object attributes from the image. The standard approach in computer vision is to constrain the search to plausible explanations and to use constrained optimization.

Another approach that has gained increasing interest in this type of problems since the seminal paper by Geman and Geman,<sup>1</sup> is Bayesian inference, which is based on defining probability space for all events (joint probability of all observed and unobserved variables) and then finding out the conditional posterior probability of the variables of interest given the observed variables.<sup>2</sup> In object recognition, for example, we must define the joint probability of perceiving an image (or set of features)  $I$  together with object attributes or classes  $C$ ,  $P(I, C)$ , which is often convenient to construct as  $P(I, C) = P(I|C)P(C)$ , where the first term is the probability of observing the image (or

---

Further author information: (Send correspondence to J.Lampinen)

J. Lampinen: E-mail: Jouko.Lampinen@hut.fi, T. Tamminen: E-mail: Toni.Tamminen@hut.fi  
T. Kostiainen: E-mail: Timo.Kostiainen@hut.fi, I. Kalliomäki: E-mail: Ilkka.Kalliomaki@hut.fi

features)  $I$  if the class of the object were  $C$  (also called the likelihood of class  $C$ ), and  $P(C)$  is the prior probability of class  $C$  before the observation  $I$ . The posterior probability is then, according to the Bayes' rule

$$P(C|I) = \frac{P(I|C)P(C)}{P(I)} \quad (1)$$

where the denominator  $P(I)$  is a normalization term, making  $P(C|I)$  to integrate to unity over  $C$ ,  $P(I) = \int P(I|C)P(C)dC$ , or  $P(I) = \sum_i P(I|C_i)P(C_i)$  in case of discrete  $C$ .

This is typical inverse problem, where the forward problem  $P(I|C)$  explains the observed data if all the unknowns were available, and is usually easier to model than the inverse part  $P(C|I)$ . For example, in tomographic imaging and other similar inverse problems statistical inverse methods based on Bayesian inference have shown increasing success in recent years.<sup>3</sup>

The posterior distribution of the end variables  $C$  contains all information that can be inferred about  $C$  given the observation  $I$  and the model assumptions. The distribution is often summarized by some indicator values or point estimates, such as the MAP (Maximum a Posteriori) estimate  $C_{MAP} = \arg \max_C \{P(C|I)\}$ , the posterior mean  $C_{mean} = E_C \{P(C|I)\}$ , or the posterior median. It should be noted, however, that proper Bayesian analysis requires that the actual posterior distribution is used in any further inference, instead of any estimated values.

In practice the problem is more complicated than Eq. 1, as it is neither needed nor practical to try to infer all the unknown attributes. Note that  $C$  contains all attributes of the likelihood model  $P(I|C)$ , that is, all the attributes that are needed to explain the perceived image or features. In 3D shape recognition, for example,  $C$  might consist of complete 3D model of the scene, all surface textures and surface reflectance properties, lighting conditions, viewing angle and other camera parameters, etc., so that the probability of observed image  $I$  could be computed by first rendering the model to produce image  $I_C$ , and then computing the probability of observing  $I$  if the actual image were  $I_C$  from the image acquisition model defining the measurement error probabilities.

Let  $S$  denote the variables that we attempt to infer,  $V$  all the other unknown variables needed in the likelihood model, and  $H$  all the attributes, distributions and models that are assumed to be known exactly, such are Gaussian measurement distribution, or the used rendering and lighting models. The desired posterior distribution of  $S$  is then

$$P(S|I, H) = \int P(S, V|I, H)dV = \frac{\int P(I|S, V, H)P(S, V|H)dV}{P(I|H)} \quad (2)$$

where the normalizing factor is again  $P(I|H) = \int P(I|S, V, H)P(S, V|H)dSV$ .

In practice the forward model (from  $S, V, H$  to  $I$ ) cannot explain very accurately all the features in the image, so that the residual model explaining the variation of the observed image (or features) from the expected image (or features), given  $S, V, H$ , must be sufficiently loose. Typically, even in the most simple analysis, the parameters of the noise models need to be included in  $V$  and be integrated out. Otherwise the model builder has to guess the suitable noise variances etc., which is difficult as the residual model must explain the actual observation errors and also the variations due to too simple forward model in the likelihood term, which are probably not Gaussian and thus the distribution and scale of the residuals may be difficult to guess in advance. Fixing the hyperparameter values may tune the system for the specific conditions and produce poor results in other situations.

In the following we describe a object matching system based on Bayesian inference. The goal of inference is to find the posterior probability distribution of the possible matches given the image and the object prior. We use Markov Chain Monte Carlo (MCMC) methods, such as Gibbs and Metropolis sampling, to estimate the posterior probability of matches. The studied problem is the matching of human faces. The prior on the face shape (i.e., prior on the locations of the corresponding points) is built up in bootstrap fashion: initially the prior is very generic (assuming no correlations of dislocations of nearby points), requiring careful fitting. From some number of fitted images we estimate the spatial covariance structure, which is used as the prior in the next stage. Refining the prior makes the fitting task simpler as the extra prior knowledge reduces the search space to be covered.

## 2. MARKOV CHAIN MONTE CARLO METHODS

The integrations needed in the Bayesian analysis can be computed in closed form basically for linear models and Gaussian residual models. Thus in practical problems, either the distributions and models must be approximated by

sufficiently simple forms to allow closed form solutions, or the integrals must be approximated by numerical methods, such as Markov Chain Monte Carlo (MCMC) methods.<sup>4,5</sup> Monte Carlo integration is a stochastic technique for the approximation of integrals by drawing random samples from a distribution and evaluating the integrand in those sample points. The most important methods used in this work are Metropolis and Gibbs sampling algorithms.

The Metropolis algorithm produces a random sequence which converges to a given target distribution  $p(\theta|X)$ . The Metropolis algorithm samples candidate points  $\theta^*$  from a so called jumping distribution  $J_t(\theta^*|\theta^{t-1})$ , which is required to be symmetrical, that is,  $J_t(\theta_a|\theta_b) = J_t(\theta_b|\theta_a)$  for any  $\theta_a, \theta_b$ , and  $t$ . The sample  $\theta^*$  is accepted with probability  $r = p(\theta^*|X)/p(\theta^{t-1}|X)$ . The jumping distribution  $J_t$  has to be able to eventually reach all states with a finite probability.<sup>2</sup>

The Gibbs sampler picks samples from conditional distributions of the target distribution  $p(\theta|X)$ . Each iteration cycles through all the components of  $\theta$  in random order. At each step in the cycle, the new state  $\theta^t$  is the previous state  $\theta^{t-1}$  with the component  $\theta_{\{j\}}^{t-1}$  substituted with a sample from the conditional distribution  $p(\theta_{\{j\}}|\theta_{\{1\dots d\setminus j\}}^{t-1}, X)$ . So the components of  $\theta$  are updated one at a time. Since the samples are taken directly from the conditional posterior density, all samples are always accepted. In order for Gibbs sampling to work, sampling from all conditional distributions of  $p(\theta|X)$  must be possible.

### 3. GABOR FILTER BASED POINT CORRESPONDENCE

We use Gabor filter based method for measuring the degree of similarity of two image locations. The Gabor filters are optimally localized band pass filters, in terms of the width of the pass band and the extent of the impulse response,<sup>6</sup> making them suitable for extracting frequency contents from as small area as possible. The Gabor filters are commonly used feature extraction method in pattern recognition, for several reasons:

- biological motivation: the simple cells in the primary visual cortex have receptive fields that resemble the Gabor filter impulse responses,<sup>7</sup>
- mathematical motivation: the Gabor filters form a complete basis for representing the images and they are optimal for measuring local spatial frequencies from the images, and
- empirical motivation: the Gabor filters have been found out to produce a distortion tolerant feature space for pattern recognition tasks.<sup>8</sup>

The basic form of the 2D Gabor filter is

$$h(x, y) = \frac{f^2}{2\pi\sigma_x\sigma_y} e^{-f^2(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2})} e^{i(f \cos(\theta)x + f \sin(\theta)y)}, \quad (3)$$

where  $f$  is the central frequency of the filter,  $\theta$  the directional angle of the filter and  $\sigma_x^2$  and  $\sigma_y^2$  the spatial-domain variances of the filter. The coefficient  $\frac{f^2}{2\pi\sigma_x\sigma_y}$  compensates for the frequency-related decrease of the power spectrum in natural images.<sup>6</sup> The Fourier transform or the transfer function of the filter is the Gaussian

$$F(\omega_x, \omega_y) = e^{-\frac{1}{2f^2}(\sigma_x^2(\omega_x - f \cos(\theta))^2 + \sigma_y^2(\omega_y - f \sin(\theta))^2)}. \quad (4)$$

The above filters contain a DC component, i.e. their means are nonzero, which makes the filters sensitive to the absolute gray level of the image. This can be avoided by subtracting a suitable Gaussian from (3), resulting in the impulse response<sup>6</sup>

$$h(x, y) = \frac{f^2}{2\pi\sigma_x\sigma_y} e^{-f^2(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2})} (e^{i(f \cos(\theta)x + f \sin(\theta)y)} - e^{-\frac{1}{2}(\sigma_x^2 \cos^2(\theta) + \sigma_y^2 \sin^2(\theta))}). \quad (5)$$

In the work reported here we used six orientations and three frequencies, in one octave spacing, to yield self-similar set of filters, where all filters were scaled and rotated versions of any one basic filter.

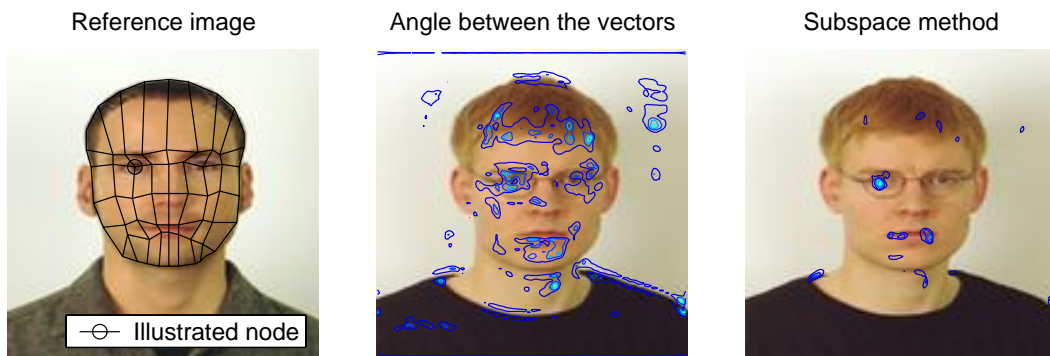
Comparison of the outputs of Gabor filter banks in two image locations is usually based on the amplitudes of the filters only.<sup>9</sup> Typically, the amplitudes of the filter outputs are stacked as a vector (sometimes called a Gabor-jet),

and common vector similarity measures are used, such as Euclidean distance (often after normalization of the length to remove the effect of image contrast) or inner product or angle between the vectors.<sup>9</sup>

Neglecting the information about phase, however, causes some limitations in the feature space. Consider some fiducial details in the image, which give highly distinctive Gabor filter amplitudes (such as an edge, corner, human eye, etc.) The Gabor jet (amplitude) vector contains only indirect information of the distance from such a detail. For example, the only clue to the distance from an edge is the faster decay of outputs of the smaller receptive field filters. Then robust matching of point chosen from close to some distinct detail is difficult. Figure 3 shows an example of this where the best match to the feature close to eye is found from the middle of the eye, using the standard angle similarity measure.

In automatic matching of two objects there is need to find point correspondence for arbitrary points in the images, so the reference locations are not necessarily in the fiducial points of the Gabor filter outputs. In this study we apply a similarity measure sensitive to the filter phases, with amplitude matching based on subspaces instead of single Gabor jets. Another similarity measure sensitive to the phase of the Gabor filters has been presented by Wiscott.<sup>10</sup>

To model the variation of the Gabor filter responses on the corresponding points in various images, the features associated to the object locations are represented as subspaces, with basis  $M_i$ , and the similarity measure  $S$  is the angle between a test Gabor jet  $J$  and the subspace  $M_i$ , (Eq. 6), with extra term from the phase error (Eq. 7). The phase information associated to each feature consists of the medians of the phases for all frequencies  $f$ , denoted by  $\Phi_{f,i}$ , of the orientation giving on average the strongest response in the images used to compile the feature subspace.



**Figure 1.** Example of the performance of the subspace based similarity measure for the Gabor jets. The contour lines show the similarity as  $\exp -S_{amp}$ , in order to enhance the visibility of the well matching areas. All the closed contour lines represent local maxima of the similarity. Left image shows the reference image and the location where the feature is picked from. Middle image shows the standard similarity measure, the angle between the jets. Right image shows the similarity with respect to a 5-dimensional subspace, built automatically from 12 images.

The likelihood, or probability of feature  $M_i$  producing the Gabor jet  $J$ , is based on *ad hoc* exponential distribution:

$$S_{amp}(J, M_i) = \text{acos} \left( \frac{J^T M_i^T M_i J}{\|J^T\| \|M_i^T M_i J\|} \right) \quad (6)$$

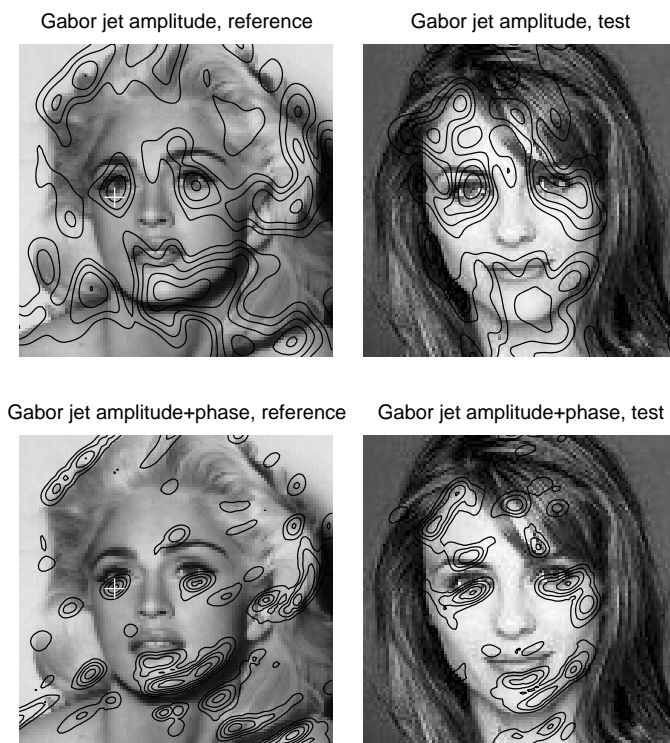
$$S_{phase}(\phi, \Phi_i) = \sum_f (\phi_f - \Phi_{i,f})^2 \quad (7)$$

$$p(J|M_i) = \exp(-\beta(S'_{amp} + S'_{phase})) \quad (8)$$

where  $\phi$  denotes the phases of the filters respective to those stored in the feature  $\Phi_i$ , and  $S'_{amp}$  and  $S'_{phase}$  denote the similarity measures normalized over the image to have maximum value of one. Note that the angle  $S_{amp}$  is always positive, in  $[0, \pi/2]$ , as all the amplitude vectors are in the positive quadrant.

#### 4. OBJECT MODEL

The model we are fitting to the face is composed of a set of locations  $r_i = \{x_i, y_i\}$  with a Gabor-filter based features associated to each location. Basically the set contains all points on the object, but for computational efficiency we use



**Figure 2.** Example of the effect of utilizing the phase information with the Gabor filters. The black contour lines show the likelihood of the feature picked from the location marked by white plus sign. All the closed curves represent local maxima of the likelihood. Top row: likelihood of the feature in reference image (left) and test image (right), based on the angle between the vector of filter amplitudes. Bottom row: feature likelihood using also phase information. Note how the amplitude based similarity finds the best match in the middle of the eye, while the phase term restores the correct displacement.

sparse set of points that should contain the most important details on the object, i.e., a labeled graph representation. The goal of the matching is to find the probability distribution for a set of locations  $q_i$  on the analyzed image so that the corresponding points in the image are mapped together.

We attempt to let the system learn the spatial variability of the faces (the probability of mapping given points  $\mathbf{r}$  to locations  $\mathbf{q}$ ) and utilize that knowledge in the sampling process as prior of the object model. The most simple model for the spatial variability is based on linear assumption for the variation with respect to some reference object. On this assumption, any object can be expressed as a linear combination of the eigenvectors of the correlation matrix of the grid point positions. This is the linear object class model suggested by Poggio and Vetter<sup>11,12</sup> for view-based 3D object modeling. In the linear object classes, the linear transformation to produce a novel (orthographic) view of an object can be learned from other objects, for which the initial and target views are available.

The spatial model is represented by the reference grid  $m_i$ , and the covariance matrix  $C$ , where  $C_{ij} = E\{(r_i - m_i)'(r_j - m_j)\}$ . In practice,  $m$  and  $C$  are approximated by sample means from a set of sample objects. The covariance matrix  $C$  can be computed from manually labeled images, but we aim to estimate it in bootstrap fashion to make the object learning as automatic as possible.

In the initial stage we set up the reference grid  $r_i$  on the face (either automatically, or by hand for more informative feature locations) and assume diagonal covariance (uncorrelated nodes). On images to be fitted, the grid is first scaled by  $s_x$  and  $s_y$ , in x- and y-directions, and translated to center coordinates  $x_0, y_0$  to yield positions  $\mathbf{r}' = T(\mathbf{r}, x_0, y_0, s_x, s_y)$ . The grid node locations have then Gaussian probability distribution around  $\mathbf{r}'$ ,

$$p(q_i) = N(r'_i, \alpha), \quad (9)$$

where  $N(m, \alpha)$  denotes a normal distribution with mean  $m$  and variance  $\alpha$ . Next we estimate the covariance model for the objects from a set of fitted images. Figure 5 shows an example of the covariance model. The subfigures show the mean grid and the principal components in decreasing order added to the mean. Due to coarse preprocessing of the images, the first principal components appear to code slight changes in camera angle. Example of priors arising from the covariance model are shown in Fig. 6.

In the next section we explain in detail the MCMC sampling of the model parameters, for the initial stage with no covariance model, and for the latter stage using the linear covariance model.

## 5. MCMC SAMPLING OF THE MODEL PARAMETERS

The model variables we are interested in are the translation  $x_0, y_0$ , scaling  $s_x, s_y$ , and positions of the nodes  $\mathbf{q} = \{q_i^{(x)}, q_i^{(y)}\}, i = 1, \dots, N$ . The variance of the node positions around the reference grid ( $\alpha$  in Eq. 9) is difficult to know in advance, and too small value will bias solutions towards positions in the reference grid, while too large value causes the nodes to be associated to the best matching (maximum likelihood) positions, without preserving the topographic relations of the face. To avoid guessing the variance we take it as unknown model attribute, set a prior for it, and integrate it out from the posterior distributions of the actual goal variables. In MCMC approach, this only requires sampling from the joint posterior of all unknown and it can be efficiently implemented by Gibbs sampling the variance from the full conditional of it given all the other model attributes, so the value depends on the prior and the evidence from the current realized deviations from the reference grid, as detailed below (see Eq. 11).

The selection of used MCMC method is important for computationally efficient sampling. As an example, consider sampling of the position of one grid node. The likelihood field, computed as explained in section 3 over the image, defines the probability of the feature appearing in each location, and contains several maxima. The prior on the position is Gaussian (circular centered on reference grid position, or the conditional from the covariance model, Eq. 14), and thus the posterior has irregular multimodal shape limited by the support of the Gaussian prior. Metropolis sampling is equal to random perturbation of the node position, and acceptance of the new position according to the posterior probability. Gibbs sampling is equal to choosing the new position proportional to the posterior probabilities, which is algorithmically more difficult, but it can cope with multimodal feature likelihood much more efficiently. Gibbs sampling chooses from all the likelihood maxima within the prior support, so that it 'sees', on each draw, all the possible matches for the feature, which are allowed by the prior, no matter how sharply peaked the maxima are. Metropolis, on the other hand, changes the mode from a sharp peak to another much less frequently, as it requires that the random perturbation hits the other peak. For this reason we prefer to use Gibbs sampling for the node positions, especially as the feature matching produces rather sharp peaks in likelihoods. Figure 3 shows example of the probabilities during the Gibbs sampling.

The global translation and scaling parameters  $x_0, y, s_x, s_y$  can be sampled directly with the Metropolis algorithm. Then Gibbs sampling for the individual node positions is carried out, and the whole loop is repeated. The priors and hyperpriors need not be updated at each iteration of the algorithm - sampling every 5-10 steps is sufficient.

The full probability model in the task should contain a decision process for each grid node, expressing the posterior probability that the corresponding point is visible in the image, to account for occlusion etc. In the current implementation this is done by simple *ad hoc* rule: if the posterior probability of the grid point is very low, compared to other grid nodes, we assume the feature is not visible and the position is drawn from the prior.

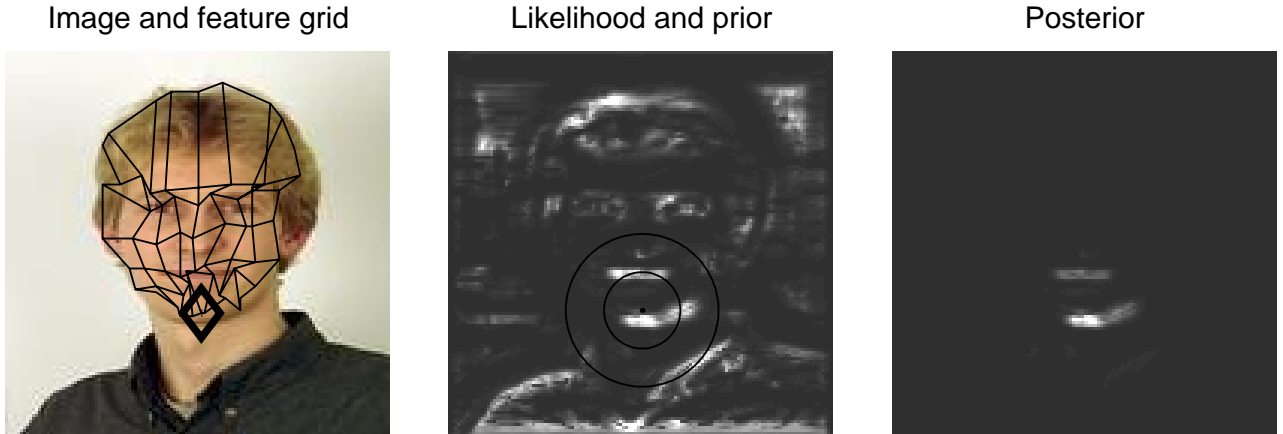
In the next sections we explain in detail the Gibbs sampling of the node positions.

### 5.1. Sampling the node positions without covariance model

In the initial stage, without the covariance model, we use hierarchical prior for the variances of the node positions (i.e., heteroscedastic residual model), where each node has different variance  $\alpha_i^0$ , expressing the prior knowledge that we do not know the range of variation in each node. The variances have common prior, or hyperprior, with unknown mean  $\alpha^1$ , which has a fixed higher level prior. For detailed discussion of similar prior structure in regression model, see the review paper by Lampinen and Vehtari.<sup>5</sup>

For the variance of the Gaussian, a conjugate distribution is the inverse Gamma. Thus to use Gibbs sampling for variance, we set a inverse Gamma prior

$$\sigma^2 \sim \text{Inv-gamma}(\sigma_0^2, \nu_\sigma), \quad (10)$$



**Figure 3.** Example of feature probabilities during the Gibbs sampling. Left image: the test image and the grid of feature locations, during early stage of sampling, when the variances  $\alpha_i^1$  are still large. The probabilities associated to the node marked by the diamond are shown in the next images. Middle image: likelihood of the feature shows as gray scale, and the Gaussian prior for the location of the node, determined by all the other nodes and the shape of the reference grid. Right image: the posterior probability of the node location. During the Gibbs sampling the new location for the node is drawn proportional to the shown probabilities.

with parametrization  $\text{Inv-gamma}(\sigma^2 | \sigma_0^2, \nu) \propto (\sigma^2)^{-(\nu/2+1)} \exp(-\frac{1}{2}\nu\sigma_0^2\sigma^{-2})$ . The parameter  $\nu$  is the number of degrees of freedom and  $\sigma_0^2$  is a scale parameter. The conditional posterior for the variances are then obtained from inverse Gamma distribution:

$$p(\alpha_i^0 | \alpha^1, \nu_0, \mathbf{q}, I) = \text{Inv-gamma}\left(\frac{\nu_0\alpha^1 + \|q_i - f(\mathbf{q}_{\setminus i})\|^2}{\nu_0 + 1}, \nu_0\right) \quad (11)$$

$$p(\alpha^1 | \alpha_2, \nu_2, \nu_0, \alpha^0, N) = \text{Inv-gamma-inv-gamma}(\alpha_2, \nu_2, \nu_0, \alpha^0, N) \quad (12)$$

where  $\mathbf{q}_{\setminus i}$  denotes  $\mathbf{q}$  with  $i^{\text{th}}$  element removed,  $f$  gives the “expected” position of the feature according to all the grid parameters,  $\nu_0, \alpha_2$  and  $\nu_2$  are distribution parameters and is  $N$  the number of features. Inverse Gamma prior for  $\alpha_1$  gives Inverse-Gamma-Inverse-Gamma posterior, from which we use ARS (adaptive rejection sampling), modified from the FBM software\*, where Gamma-Gamma-distribution was used for sampling the inverse of second level variance hyperparameters.

The full conditional posterior for positions  $q_i$  is simply Gaussian

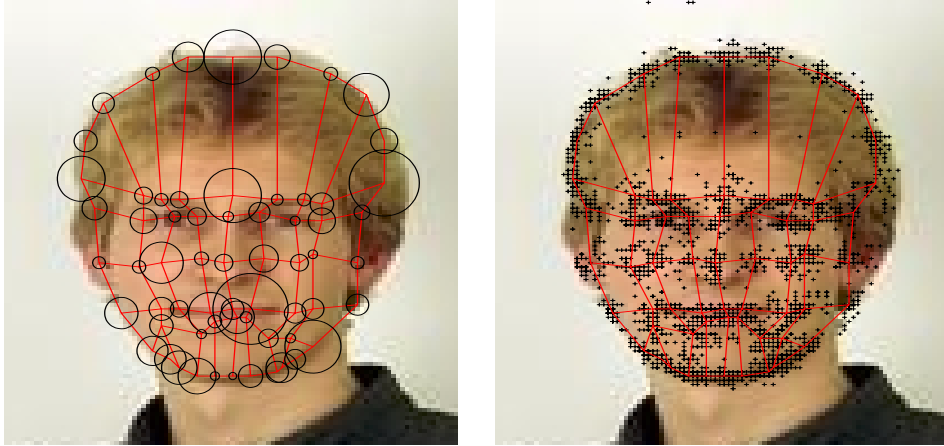
$$p(q_i | \mathbf{q}_{\setminus i}, x_0, y_0, s_x, s_y, \alpha^0, \alpha^1) = N(f(\mathbf{q}_{\setminus i}), \alpha_i^0) \quad (13)$$

The distribution parameters of the hyperpriors can be set quite arbitrarily so that results are satisfactory. As they are located high in the hierarchical model, their exact values are not of much importance. The same applies to the Metropolis algorithm deviation parameters. Figure 4 shows an example of the posterior distribution of matching a face with no covariance model. Figure 5 shows an example of the covariance model, computed from a set of 30 images.

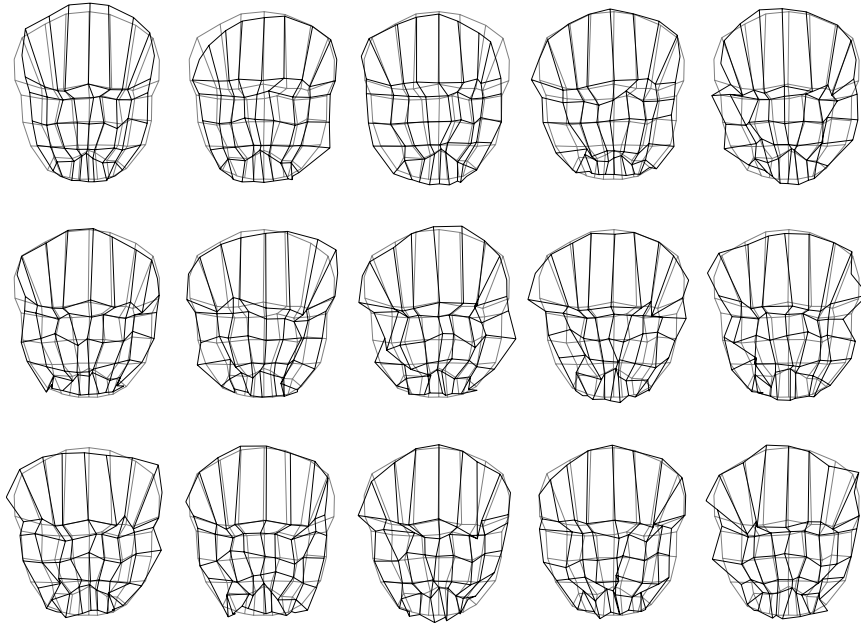
## 5.2. Sampling the node positions with covariance model prior

In order to use Gibbs sampling for the node positions, we need the full conditional distribution of the location of one node  $i$  given the locations for all the other nodes, marked by  $v$ . As the spatial variability model is Gaussian (we have linear object class, with no higher order relations between displacement of the positions) all conditional densities are also Gaussian.

\*<http://www.cs.toronto.edu/~radford/fbm.software.html>



**Figure 4.** The final posterior distribution of the feature locations. The left image shows the grid as the median of the distribution, and the posterior values for the variances  $\alpha_i^0$  for each node are shown as circles, drawn at two-sigma distances. The right image shows the samples of the grid locations on top of the image.



**Figure 5.** Covariance model computed from 30 real images. Gray lines: the reference grid. Black lines: principal components of the covariance model added to the reference grid. Note that considerable variation (principal components 2 and 3) seem to be related to slight changes in the view angle.

Let  $C_v$  denote the  $2N - 2 \times 2N - 2$  block of the covariance matrix  $C$  that contains the rows and columns related to the fixed nodes, and  $C_i$  the  $2 \times 2$  block related to the sampled node, and  $C_{iv}$  the  $2 \times 2N - 2$  block containing the cross-variance terms. The conditional distribution of  $p(r_i|r_v)$  is then Gaussian with mean and variance

$$\mathbf{E}(r_i|r_v) = m_i + C_{iv}C_v^\dagger(r_v - m_v) \quad (14)$$

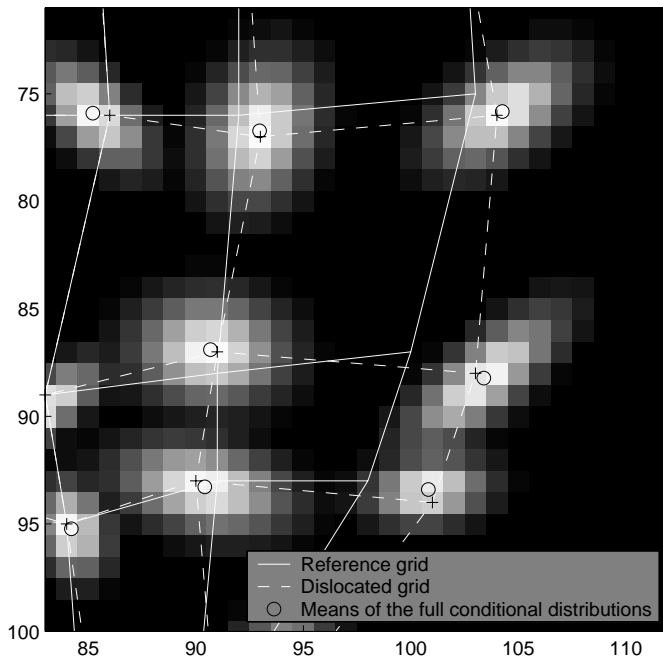
$$\text{Var}(r_i|r_v) = C_i - C_{iv}C_v^\dagger \quad (15)$$

where  $C_v^\dagger$  is the generalized inverse, or pseudo-inverse of  $C_v$  (Note that the rank of  $C$  is much less than  $N$ , depending on the number of samples used to estimate the covariances and the number of eigenvectors retained in the covariance

estimate).

In the Gibbs sampling, the prior term for each node is computed from equations above, and multiplied by the likelihood terms to yield the posterior, from which the new position is drawn. This is repeated for each node. The matrix inverse terms  $C_i v C_i^\dagger$  for all nodes  $i$  can be computed in advance to speed up the sampling.

Figure 6 shows an example of the conditional distributions for the node locations. The distributions are computed for each node and overlaid on the image. The displacement of the nodes is in the direction of the second eigenvector. Fixing the  $N - 1$  node positions gives strong evidence that the displacement is in the direction of the second eigenvector, and thus the conditional density of the last node is concentrated near that location. Note that means of the conditional distributions are not equal to the grid positions, as the displacement vector of the  $N - 1$  nodes is not exactly orthogonal to the other eigenvectors, even though the full displacement vector is, and thus the conditional densities contain contributions from other eigenvectors, too.

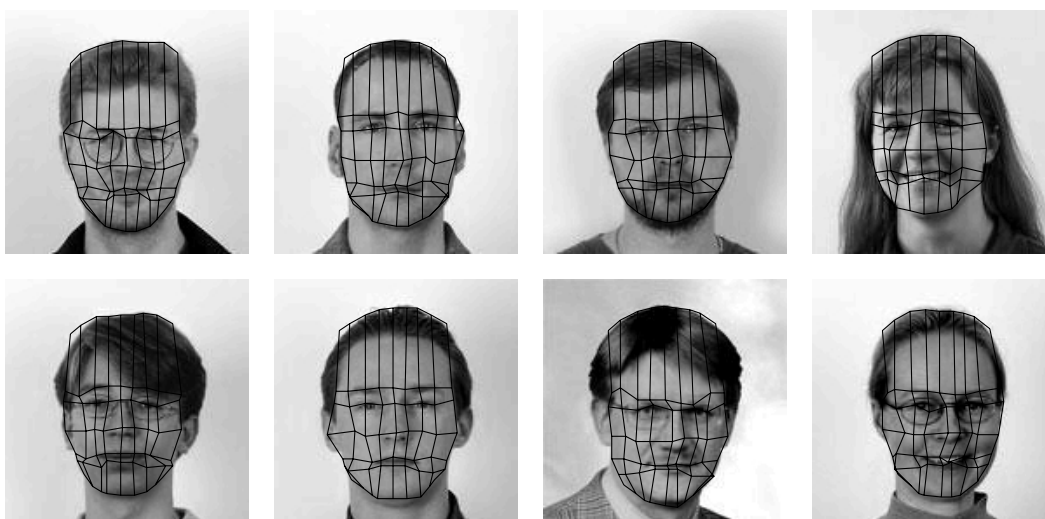


**Figure 6.** Example of conditional distributions of the node locations. The grid drawn by solid line is the reference grid of the covariance model. The grid shown as dashed line is changed towards the eigenvector with the second largest eigenvalue. The gray scale shows the conditional distributions for each node, given that all the other nodes are in the locations shown by the grid.

Figure 7 shows some examples of fitting the model on the faces. The shown grids are medians of 50 MCMC samples. Visual inspection of the quality of the fitting is not straightforward, especially as the grid is slightly different from the one in section 5.1, but the method seems rather robust. Based on these results, we are currently expanding the object priors to be three-dimensional, so that the shape of the head could be recovered from one image.

## 6. CONCLUSION

We have reported some preliminary results of using MCMC sampling for object matching. The most notable advantage of the Bayesian approach is that it provides a theoretically sound basis for including problem specific background information (handcrafted or learned from samples) into the inference process. In the face matching task studied in the paper, we model the spatial variability of different faces by linear correlations (leading to linear object class representation<sup>11</sup>), and show how the posterior distribution for any point can be computed from the positions of the other points and the likelihood function of features that measure the point correspondence. We present Metropolis



**Figure 7.** Examples of fitting the feature grids on faces, shown as the median of 50 MCMC samples. The actual distribution of the grids is difficult to illustrate, and the considerable ambiguity in the matching point locations decreases the smoothness of the median grids.

and Gibbs sampling methods for drawing samples from the posterior distribution of all the model parameters, and show how some auxiliary (nuisance) parameters can be integrated out from the posterior distributions.

In the future work we attempt to expand the object priors to three dimensions, so that the learned 3D covariance model aids in guessing the 3D shape of the head from one 2D image.

## REFERENCES

1. S. Geman and D. Geman, “Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**(2), pp. 721–741, 1984.
2. A. Gelman, J. B. Carlin, H. S. Stern, and D. R. Rubin, *Bayesian Data Analysis*, Texts in Statistical Science, Chapman & Hall, 1995.
3. J. P. Kaipio, V. Kolehmainen, E. Somersalo, and M. Vauhkonen, “Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography,” *Inverse Problems* **16**, pp. 1487–1522, 2000.
4. W. Gilks, S. Richardson, and D. Spiegelhalter, eds., *Markov Chain Monte Carlo in Practice*, Chapman & Hall, 1996.
5. J. Lampinen and A. Vehtari, “Bayesian approach for neural networks – review and case studies,” *Neural Networks* **14**, pp. 7–24, April 2001. (Invited article).
6. J. Daugman, “Complete discrete 2-d Gabor transforms by neural networks for image analysis and compression,” *IEEE Trans. Acoustics, Speech, and Signal Processing* **36**, pp. 1169–1179, 1988.
7. J. Jones and L. Palmer, “An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex,” *J. Neurophysiol.* **58**(6), pp. 1233–1258, 1987.
8. J. Lampinen and E. Oja, “Distortion tolerant pattern recognition based on self-organizing feature extraction,” *IEEE Transactions on Neural Networks* **6**, pp. 539–547, May 1995.
9. M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. v.d. Malsburg, R. Wurtz, and W. Konen, “Distortion invariant object recognition in the dynamic link architecture,” *IEEE Transactions on Computers* **42**(3), pp. 300–310, 1993.
10. L. Wiskott, “Phantom faces for face analysis,” *Pattern Recognition* **30**(6), pp. 837–846, 1997.
11. T. Vetter and T. Poggio, “Linear object classes and image synthesis from a single example image,” Tech. Rep. 1531, Massachusetts Institute of Technology, March 1995.
12. D. Beymer and T. Poggio, “Face recognition from one example view,” Tech. Rep. 1536, Massachusetts Institute of Technology, 1995.