

A BAYESIAN APPROACH TO 3D OBJECT RECOGNITION USING LINEAR COMBINATION OF 2D VIEWS

Vasileios Zografos, Bernard F. Buxton

Department of Computer Science, University College London, Malet Place, London, WC1E 6BT, UK

v.zografos@cs.ucl.ac.uk, b.buxton@cs.ucl.ac.uk

Keywords: Object recognition, linear combination of views, simplex, differential evolution, Bayes, Markov-Chain Monte-Carlo

Abstract: In this work, we introduce a Bayesian approach for pose-invariant recognition of the images of 3d objects modelled by a small number of stored 2d intensity images taken from nearby but otherwise arbitrary view-points. A linear combination of views approach is used to combine images from two viewpoints of a 3d object and synthesise novel views of that object. Recognition is performed by matching a target, scene image to such a synthesised, novel view using an optimisation algorithm, constrained by construction of Bayes prior distributions on the linear combination. We have experimented with both a direct search and an evolutionary optimisation method on a real-image, public database. The Bayes priors effectively regularised the posterior distribution so that all algorithms were able to find good solutions close to the optimum. Further exploration of the parameter space has been carried out using Markov-Chain Monte-Carlo sampling.

1 INTRODUCTION

The detection and subsequent categorisation of three-dimensional objects has been one of the most important and recurring problems in computer vision, mainly due to the inherent difficulties in dealing with objects that may be imaged from a variety of view-points.

In this work, we shall examine a particular kind of *extrinsic* variation; that is, variation of the image due to changes in the viewpoint from which the object can be seen. Traditional approaches for solving this specific problem have relied on an explicit 3d model of the object and the generation of 2d projections at various viewpoints, such as the work by (Lee and Ranganath, 2003). Although such methods can accurately model pose variations, generating a 3d model can be a complex process and often requires the use of specialised hardware. Other methods (Lamdan et al., 1988; Beymer, 1994) have tried to capture the viewpoint variability by using multiple views of the object from different directions covering a large portion of the view-sphere. Since dense coverage is usually necessary for accurate recognition results such methods

require capture, storage and manipulation of a vast number of images of each object of interest, limiting their practicality. In order to overcome these limitations, alternative methods have been introduced that try to work directly with a smaller number of images. They are called *view-based* methods and represent an object as a collection of a small number of 2d views. Their advantage is that they do not require construction of a 3d model while keeping the number of required stored views to a minimum. Examples include the recent work by (Bebis et al., 2002) and (Li et al., 2004).

We propose a similar method which works directly with pixel values and thus avoids the need for low-level feature extraction and the solution of the correspondence problem (Bebis et al., 2002). In addition, by employing a Bayesian approach, we can use all the available prior knowledge about the linear combination parameters and restrict our search to regions where valid and meaningful solutions are likely to exist. We adopt a “generate-and-test” strategy using an optimisation algorithm to recover a set of linear combination of view (LCV) coefficients that will synthesise a novel image which is as similar as possible to

the target image. If the similarity between the synthesised and the target images is above some threshold then an object is determined to be present in the scene and its location and pose are defined (at least in part) by the LCV coefficients.

In the next section, we introduce the theory behind the LCV and explain how it can be used to synthesise realistic images over a range of viewpoints. In section 3, we present our extension of the LCV method to incorporate prior probabilistic information and offer one possible interpretation for the Bayesian prior probabilities. Section 4 outlines the main components of our proposed recognition system while section 5 presents some experimental results of our approach on real imagery. Finally, we conclude in section 6 with a critical evaluation of our method and suggest possible future work.

2 Linear combination of views

LCV is a technique which belongs in the general theory of the tri- and multi-focal tensors, or Algebraic Function of Views (AFoV) (Shashua, 1995) and provides a way of dealing with variations in an object's appearance due to viewpoint changes. This theory is based on the observation that the geometry of the set of possible images of an object undergoing 3d rigid transformations and scaling may, under most imaging conditions, to a good approximation, be embedded in a linear space spanned by the landmark points of a small number of images taken from different viewpoints. Ullman and Basri (Ullman and Basri, 1991) were the first to apply this theory to line drawings and edge maps.

Thus, given two images I' and I'' of an object from different (basis) views with corresponding image coordinates (x', y') and (x'', y'') , we can represent any corresponding point (x, y) in a novel target image I_T according to, for example:

$$\begin{aligned} x &= a_0 + a_1x' + a_2y' + a_3x'' + a_4y'' \\ y &= b_0 + b_1x' + b_2y' + b_3x'' + b_4y'' \end{aligned} \quad (1)$$

The target view is reconstructed from the above two equations given a set of valid coefficient (a_i, b_j) . Provided we have at least 5 corresponding landmark points in all three images (I_T, I', I'') we can estimate the coefficients (a_i, b_j) by using a standard least squares approach. Several others have taken this concept further from its initial application to line images and edge maps to the representation of real images I_T (Bebis et al., 2002; Koufakis and Buxton, 1998; Hansard and Buxton, 2000; Peters and von der Malsburg, 2001).

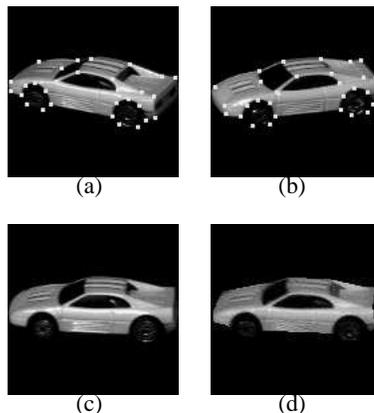


Figure 1: Example of real data from the COIL-20 database. The two basis view images I' (a) and I'' (b) with landmark points selected at prominent features. I_T (c) is the target image. The synthesised image I_S (d) is at the correct pose identified by our algorithm.

Such results suggest that it is possible to use LCV for recognition by matching images I_T of target views of an object to a combination of stored, basis views of the object. The main difficulty in applying this idea within a pixel-based approach is the selection of the LCV coefficients (a_i, b_j) .

2.1 Image synthesis

To synthesise a single, target image using LCV and two views we first need to determine its geometry from the landmark points. For pixel-based object recognition in which we wish to avoid feature detection a direct solution is not possible, but we instead use a powerful optimisation algorithm to search and recover the LCV coefficients for the synthesis.

Given a set of hypothesised landmark points (see Fig. 1(a),(b)) we can, to a good approximation, synthesise an image I_S to match to the target I_T :

$$\begin{aligned} I_T(x, y) &= w'I'(x', y') + w''I''(x'', y'') + \varepsilon(x, y) \\ &= I_S(x, y) + \varepsilon(x, y). \end{aligned} \quad (2)$$

The weights w' and w'' may be calculated from the LCV coefficients and the image I_S synthesised as described in (Koufakis and Buxton, 1998). Thus, $w' = \frac{d''^2}{d'^2 + d''^2}$, $w'' = \frac{d'^2}{d'^2 + d''^2}$ with $d'^2 = a_3^2 + a_4^2 + b_3^2 + b_4^2 + a_0^2 + b_0^2$ and $d''^2 = a_1^2 + a_2^2 + b_1^2 + b_2^2 + a_0^2 + b_0^2$. The synthesis essentially warps and blends images I' and I'' to produce I_S . It is important to note therefore that (2) applies to all points (pixels) (x, y) , (x', y') and (x'', y'') in images I_S, I' and I'' with the dense correspondence defined by means of the LCV equations

(1) and a series of piecewise linear mappings (Gosh-tasby, 1986; Koufakis and Buxton, 1998). If (x', y') and (x'', y'') do not correspond precisely to pixel values, bilinear interpolation is used. The same idea may be extended to colour images by treating each spectral band as a luminance component (e.g. I_R, I_G, I_B).

2.2 Coefficient variation

Since the pose information is implicitly encoded in the 10 coefficients (a_i, b_j) , it is useful to investigate their variation as the object is viewed from different directions. We are particularly interested in viewing directions on the portion of the view-sphere between the two basis views. We thus devised a simple experiment using a synthetic 3d model on which we selected a number of landmarks on prominent features and along edges. The model was placed over a black background and then allowed to rotate between $\pm 20^\circ$ from the frontal position and 2d snapshots of the scene were taken, under orthographic projection, at 1° intervals. The two images at $\pm 20^\circ$ were chosen as the basis views. We then solved (1) at each view and obtained a set of coefficients as a function of the angle ϑ as illustrated in Fig. 3(b).

As we can see, coefficient a_0 follows a quadratic curve, coefficients a_1 and a_3 are linear and the remaining coefficients are constant. Note also that a_1 and a_3 have a range of $[0, 1]$ with $a_1, a_3 \approx 0.5$ for $\vartheta = 0^\circ$ (frontal view). Additionally, at the same angle, a_0 is at a minimum. Finally, we observe that, $a_2, a_4, b_0, b_1, b_3 \approx 0$ and $b_2, b_4 \approx 0.5$.

This information could be used to bound the parameter space in the optimisation or to approximate regions within which we initialise our algorithm. Alternatively, we can predict the solution set at each ϑ . This information, combined with knowledge of the range of poses we are likely to encounter in a specific experiment, may thus be used to construct the Bayesian priors, as we shall see in section 3

We note here that the form of the coefficients is largely independent of the actual object used and therefore the results generalise to any generic object under similar imaging conditions. One could also carry out similar experiments with other 3d rigid transformations and recover how the LCV coefficients vary in such cases.

3 Bayesian model

In this section we propose an extension of the basic LCV equations (1) and (2) by incorporating prior information on the coefficients (a_i, b_j) and building a

Bayesian model. We start with (2) in which $\epsilon(x, y)$ is assumed to be a vector of i.i.d. random noise. Given the data (target image I_T) and a model represented by the basis views I' and I'' , the posterior probability of the LCV coefficients is:

$$P(a_i, b_j | I_T, I', I'') \propto P(I_T | a_i, b_j, I', I'') P(a_i, b_j). \quad (3)$$

$P(I_T | a_i, b_j, I', I'')$ is the probability of observing the target image I_T given the coefficients (a_i, b_j) and the basis views I' and I'' . $P(a_i, b_j)$ is the prior probability of the LCV coefficients.

If we then use the general assumption that the deviations ϵ of the synthesised image I_S from the target image I_T are drawn from a Gaussian distribution the log-likelihood of observing the target image may, up to constant terms that we shall ignore, be written as:

$$-\log[P(I_T | a_i, b_j, I', I'')] = \frac{1}{2\sigma_\epsilon^2} \sum_{x,y} [I_T(x,y) - I_S(x,y)]^2, \quad (4)$$

which is quadratic in the residuals, and the summation is over all image pixels. The other term on the right in (3) comes from the prior p.d.f..

The prior represents general information about the LCV parameters based on the analysis we have carried out previously. We have used a Gaussian prior for the coefficients a_i and b_j centered at the values identified in section 2.2. Under the assumption of statistical independence of the coefficients with each having its own mean and variance and, if we again ignore uninteresting constants we get:

$$-\log(P(a_i, b_j)) \propto \sum_{i,j=0}^4 \left[\frac{(a_i - \mu_a)^2}{\sigma_i^2} + \frac{(b_j - \mu_b)^2}{\sigma_j^2} \right], \quad (5)$$

where μ_a, μ_b are the means and σ_i, σ_j the standard deviations of the prior probabilities for coefficients a_i and b_j respectively.

From (3),(4) and (5) we see that, if we continue to ignore uninteresting constants, the log posterior probability becomes:

$$-\log[P(a_i, b_j | I_T, I', I'')] \propto \frac{\sum_{x,y} [I_T(x,y) - I_S(x,y)]^2}{\sigma_\epsilon^2} + \sum_{i,j=0}^4 \left[\frac{(a_i - \mu_a)^2}{\sigma_i^2} + \frac{(b_j - \mu_b)^2}{\sigma_j^2} \right]. \quad (6)$$

We usually require a single synthesised image to be presented as the most probable result. A typical choice for that single image is the one which maximises the posterior probability (MAP), or minimises the negative log-posterior (6) with respect to the parameters a_i and b_j . The latter can be minimised using standard optimisation techniques.

As we can see from (6), the prior is used to bias the MAP solution toward the means μ_a and μ_b , away

from the maximum likelihood (ML) solution which is where $\sum_{x,y}[I_T(x,y) - I_S(x,y)]^2$ is at a minimum (i.e. there is little difference between I_T and I_S). How much the prior affects the solution in relation to the likelihood depends on the ratio:

$$k = \frac{\sigma_\epsilon^2}{\sum_{i,j}(\sigma_i^2 + \sigma_j^2)}. \quad (7)$$

As the influence of the prior vanishes (i.e. σ_i, σ_j become very big and the Gaussian prior resembles a uniform distribution) the MAP solution approaches the ML solution. Careful selection of the ratio k is therefore very important.

The effect of using such Gaussian priors to bias the posterior can be seen in Fig. 3(a). Here we show the negative log probability of the likelihood, prior and posterior for the coefficient a_2 . This plot was generated by isolating and varying this coefficient while having conditioned the remaining coefficients to the optimal prior values identified previously. The image I_S was synthesised using (2) and compared to the target image I_T with the log probabilities recorded at each value of a_2 . We used images from the COIL-20 database (Nene et al., 1996).

What we should note in particular from this example is the effect of the prior on the likelihood, especially near the tails of the p.d.f. where we have large error residuals. The prior widens the basin of attraction of the likelihood resulting in a convex posterior that is much easier to minimise even if we initialise our optimisation algorithm far away from the optimal solution.

On the other hand, near the global optimum we wish the prior to have as little impact as possible in order for the detailed information as to the value of a_2 to come from the likelihood. This allows for small deviations from the values for the coefficients encoded in the prior means, since every synthesis and recognition problem differs slightly, due to object type, location, orientation and perspective camera effects.

4 The recognition system

In principle using the LCV for object recognition is straightforward. All we have to do is select the object and find the coefficients in an equation such as (1) which will minimise the error ϵ from (2) and check that it is small enough so that our synthesised and target images I_S and I_T are sufficiently similar to enable us to say that they match.

4.1 Template matching

The first component of our system is the two stored basis views I' and I'' which define the library of known modelled objects. These are rectangular bitmap images that contain grey-scale (or colour) pixel information of the object without any additional background data. It is important not to choose a very wide angle between the basis views to avoid I' and I'' belonging to different aspects of the object with landmark points being occluded.

Having selected the two basis views, we pick a number of corresponding landmark points, in particular lying on edges and other prominent features. We then use constrained Delaunay triangulation (Shewchuk, 2002) and the correspondence to produce similar triangulations on both the images. The above processes may be carried out during an off-line, model-building stage and are not examined here.

The set of LCV coefficients is then determined by minimising the negative log posterior (6) and the object of interest in the target image I_T recognised by selecting the best of the models, as represented by the basis views, that explain I_T sufficiently well. Essentially, we are proposing a flexible template matching system, in which the template is allowed to deform in the LCV space, restricted by the Bayesian priors to regions where there is a high probability of meaningful solutions, until it matches the target image.

4.2 Optimisation

To recover the LCV coefficients, we need to search a high-dimensional parameter space using an efficient optimisation algorithm. For this purpose, we have chosen to examine two different methods, a global, stochastic algorithm, and a local, direct-search approach. The stochastic method used is called Differential Evolution (DE) and was introduced by (Storn and Price, 1997). Briefly, DE works by adding the weighted difference between two randomly chosen population vectors to a third vector, and the fitness of the solution represented by the resultant is compared with that of another individual from the current population.

For the local method, we have selected the simplex algorithm by (Nelder and Mead, 1965) since it is very easy to implement and does not require calculation (or approximation) of first or second order derivatives. A simplex is a polytope of $N+1$ vertices in N dimensions with each vertex corresponding to a single matching function evaluation. In its basic form the simplex is allowed to take a series of steps, the most common of which is the *reflection* of the vertex



Figure 2: Image samples from the COIL-20 database

having the poorest value of the objective function. It may also change shape (*expansion* and *contraction*) to take larger steps when inside a valley or flat areas, or to squeeze through narrow cols. It can also change direction (*rotate*) when no more improvement can be made in a current path. Since the simplex is a local, direct search method it can become stuck in local optima and therefore some modifications of its basic behaviour are necessary. We thus introduced the *reducing-step restarting* simplex (Zografos and Buxton, 2007) which can make variable length “jumps” in order to avoid local optima and burrow deeper into the basin of attraction.

5 Experiments

We have performed a number of experiments on real images using the publicly available COIL-20 database. This database contains examples of 20 objects imaged under varying pose (horizontal rotation around the view-sphere at 5° intervals) against a constant background with the camera position and lighting conditions kept constant (see Fig. 2).

We constructed LCV models from 5 objects, using as basis views the images at $\pm 20^\circ$ from a frontal view, while ensuring that the manually chosen landmarks were visible in both I' and I'' (Fig. 1). Comparisons were carried out against target images from the same set of modelled objects taken in the frontal pose at 0° .

In total, we carried out 500 experiments (250 with each optimisation method \times 10 tests for each model-target image combination) and constructed two 5×5 arrays of model \times image results. Each array contains information about the matching scores represented by the cross-correlation coefficient (Zografos and Buxton, 2007). The highest scores were along the main diagonal, where each model of an object is correctly matched to a target image of the same object. The

recognition response should be less (or have a higher negative log probability) when comparing a specific model with images of other objects.

For the simplex method, we set the maximum number of function evaluations (NFEs) to 1000 and a fixed initialisation of: $a_0, a_1, a_3, b_0, b_1 = 1$, $a_2, a_4, b_3 = 0.5$, $b_2 = 0.9$, $b_4 = 1.4$, deliberately chosen far away from the expected solution, in order not to influence the optimisation algorithm with a good initialisation. In the case of DE, we chose a much higher NFEs=20000 (100 populations \times 200 generations) and specified the boundaries of the LCV space as: $-5 \leq a_0, b_0 \leq 5$, $-1 \leq a_1, a_2, a_3, a_4 \leq 1$, $-1 \leq b_1, b_2, b_3, b_4 \leq 1$. This includes the majority of the coefficient ranges identified in section 2.2.

The results of the above experiments, averaged over 10 test runs, are summarised in the heatmap plot Fig. 3(c). As expected, we can see a well defined diagonal of high cross-correlation where the correct model is matched to the target image. This observation, combined with the absence of any significant outlying good matches when model \neq image, leads us to the conclusion that, on average, both methods perform equally well in terms of recognition results. The only difference is how close these methods can get to the global optimum, and in how many NFEs.

To explore this further, we have included the plot in Fig. 4 which compares the average cross-correlation responses for both methods. We observe that both methods have a consistently good performance, with the DE converging to solutions of higher cross-correlation in the majority of cases. The simplex failed to converge to the correct solution in a few cases, particularly in some of the tests for models 1 and 9. This of course may be explained in part by the smaller NFEs that were allowed for this algorithm, although preliminary experiments had indicated the NFE value chosen should generally have sufficed.

We also include boxplot diagrams, Fig. 3(d),(e), which illustrate the diversity of the 10 coefficients in the recovered solutions. For this, results were taken from the test runs along the main diagonal, using both of the optimisation algorithms. We can see that there is little diversity in the coefficients with few or no outliers, indicating a stability in the values already identified by the 3d experiment (section 2.2) across different objects. The only significant variation is, as expected, in the translation coefficients a_0 and b_0 .

Finally, we have included in Fig. 3(g),(h) for test runs on the main diagonal when the correct model is matched to a target image, values of the negative log probability as the number of function evaluations increases. From these, we observe the different optimisation behaviours of the two algorithms, DE and

simplex, and how much earlier the latter can reach the global minimum. DE is much slower, but it has the advantage that it can avoid locally optimum solutions, which the simplex sometimes clearly cannot.

5.1 Markov-Chain Monte-Carlo

In previous sections we have introduced the posterior distribution and have provided some graphs of the variation of the individual coefficients. Nevertheless, graphical analysis of the 10-dimensional posterior distribution is difficult, so to obtain a more specific and complete idea of its characteristics we have used a Markov-Chain Monte Carlo (MCMC) (Gelman et al., 1995) approach in order to generate a sample of the distribution and further analyse it.

Briefly, MCMC is method for sampling from an unknown distribution, requiring only that its density can be calculated at each point. Initially MCMC draws values from a known, starting distribution and then gradually adjusts these draws to converge to an approximation of the posterior distribution. These samples are drawn sequentially with the draws forming a Markov Chain. In this particular implementation, we have used the Metropolis-Hastings rule (Metropolis et al., 1953; Hastings, 1970) and a product of ten 1-dimensional normal distributions, such as those used in section 3 for the priors, as a starting distribution. In addition, the first half of the samples were discarded in order to reduce any residual correlation in the draws.

We chose a single experiment (matching to a frontal view of object 1 at 0^0) and generated a set of 10000 samples of the posterior (6) from areas of high probability using a single Markov Chain. The MCMC converged to a point very close to the global optimum, with a cross-correlation of 0.98508 (the best solution recovered by the simplex was 0.987301 and that for the DE was 0.973954). The way the MCMC gradually improved the posterior probability can be seen in Fig. 3(i) and compared with the simplex and DE methods illustrated in Fig. 3(g) and (h).

The first step of the analysis is to determine the major modes of the p.d.f. and find where other locally optimal solutions may be situated. For that purpose, we used repeated runs of a k-means clustering algorithm (Bishop, 1995), the best of which recovered 3 main clusters in close proximity and all near the global optimum. This leads to the conclusion that the distribution has a single mode (i.e. peak) perhaps with some subsidiary, nearby peaks caused by noise effects. The main point is that there is no significant local optimum elsewhere nearby in the distribution.

We have also included boxplots of the samples of

the LCV coefficients taken in the MCMC in Fig. 3(f). The samples are tightly bundled with a few outliers, another indication that the algorithm has converged to the optimum. This reinforces the notion that the posterior distribution we have constructed is unimodal.

For a quantitative, numerical analysis we have calculated the statistical characteristics of the coefficients $a_0 - a_4$, $b_0 - b_4$ produced by the MCMC sampling as shown in Table 1. The characteristics of the coefficients $a_0 - a_4$, $b_0 - b_4$ described in the first two columns are arranged in two rows in the last five columns of the table. Our first observation is that the means, modes and medians of each of the coefficients almost coincide, consistent with the posterior being an approximately symmetric, unimodal distribution. If we look at the range values, we see once more that the values are tightly concentrated signifying a basin of attraction that bottoms-out into a few close-by points. This is further affirmed by the low standard deviations of the sampled values of all 10 coefficients.

If we now examine the kurtosis and skewness we can see positive skewness in the samples of certain coefficients and negative in others. The distributions of the samples of all coefficients, except b_1 , are quite strongly skewed, reflecting strong influence of the likelihood near the optimum posterior, a property that is highly desirable. This is due to the shape of the likelihood function since the priors are symmetric. The values for the kurtosis are small for some coefficients whose posteriors are therefore almost Gaussian near the optimum, whilst other coefficients strongly affected by the priors are leptokurtic, in particular a_4 , b_2 and b_4 .

6 Conclusion

We have shown how the linear combination of views (LCV) method may be used in view-based object recognition. Our approach involves synthesising intensity images using LCV and comparing them to the target, scene image. In addition, we incorporate prior probabilistic information on the LCV parameters by means of a Bayesian model. For the priors, we chose Gaussian distributions centered at the expected values of those parameters given a specific transformation (in this case rotation of the object about a vertical axis in 3d).

Matching and recognition experiments were then carried out on data from the COIL-20 public database using two different optimisation algorithms in order to recover the optimal LCV parameters. These experiments have shown that our method works well for

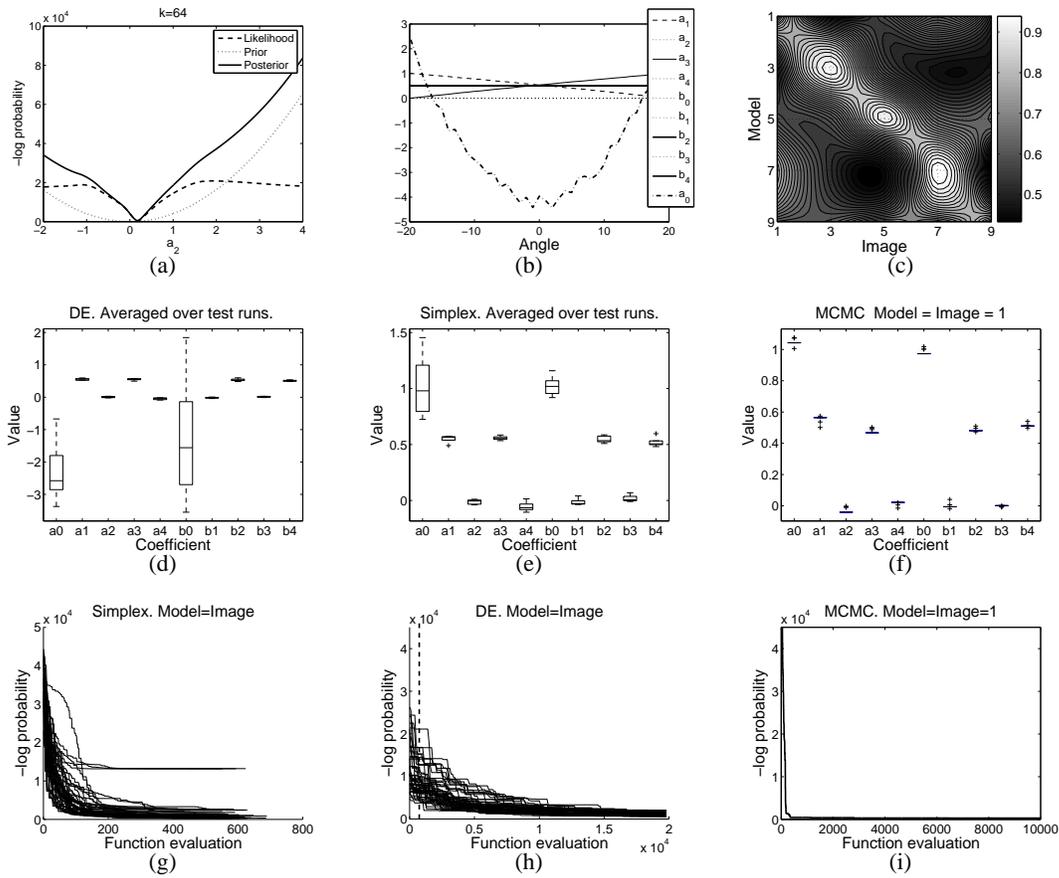


Figure 3: (a) The negative log posterior resulting from the combination of the prior and likelihood. (b) Variation plot of the 10 LCV coefficients. (c) Model \times Image heatmap array with high cross-correlation on the main diagonal. (d)-(f) Coefficient diversity plots for the DE, Simplex and MCMC methods respectively. (g)-(i) Iterative optimisation results for the three methods. Note the different scales on the ordinate and the vertical line in (h) indicating the NFEs for the simplex method.

Table 1: Statistical characterisation of the MCMC algorithm.

Dispersion measures:						
	range	a_0 : 0.069	a_1 : 0.072	a_2 : 0.038	a_3 : 0.034	a_4 : 0.036
		b_0 : 0.045	b_1 : 0.058	b_2 : 0.037	b_3 : 0.009	b_4 : 0.043
	std. dev.	a_0 : 0.013	a_1 : 0.024	a_2 : 0.015	a_3 : 0.012	a_4 : 0.010
		b_0 : 0.015	b_1 : 0.006	b_2 : 0.008	b_3 : 0.002	b_4 : 0.004
Location measures:						
	mean	a_0 : 1.050	a_1 : 0.551	a_2 : -0.031	a_3 : 0.474	a_4 : 0.016
		b_0 : 0.981	b_1 : -0.004	b_2 : 0.481	b_3 : -0.000	b_4 : 0.510
	median	a_0 : 1.042	a_1 : 0.564	a_2 : -0.040	a_3 : 0.467	a_4 : 0.022
		b_0 : 0.972	b_1 : -0.005	b_2 : 0.480	b_3 : 0.001	b_4 : 0.510
	mode	a_0 : 1.042	a_1 : 0.564	a_2 : -0.040	a_3 : 0.467	a_4 : 0.022
		b_0 : 0.972	b_1 : -0.005	b_2 : 0.480	b_3 : 0.001	b_4 : 0.510
Distributional measures:						
	skewness	a_0 : 1.171	a_1 : -1.458	a_2 : 1.214	a_3 : 1.275	a_4 : -2.053
		b_0 : 1.3685	b_1 : 0.305	b_2 : 2.520	b_3 : -1.689	b_4 : -2.071
	kurtosis (-3)	a_0 : -0.568	a_1 : 0.309	a_2 : -0.466	a_3 : -0.256	a_4 : 3.301
		b_0 : 0.193	b_1 : 1.523	b_2 : 6.325	b_3 : 0.874	b_4 : 5.668

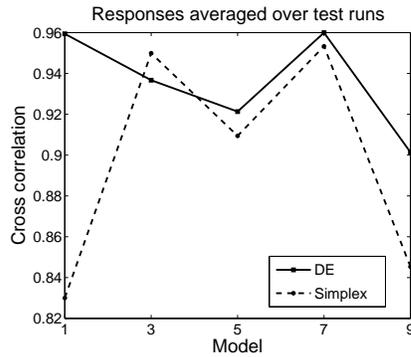


Figure 4: Comparison of the average response between the DE and simplex algorithms.

pose variations where the target view lies between the basis views. The experiments further show the beneficial effects of the prior distributions in “regularising” the optimisation. In particular, priors could be chosen that produced a good basin of attraction surrounding the desired optimum without unduly biasing the solution. Finally, we used MCMC to draw a sample from the posterior distribution and carried out additional tests to characterise the shape of the distribution. The use of the MCMC approach as an optimisation solution was also briefly explored with, because of the form of the posterior, satisfactory results.

Nevertheless, additional work is required. The kurtosis of the posterior distribution calculated from the samples taken in the MCMC experiment suggests that there may be effects arising from the fact that the LCV coefficients are not all independent. We would thus like to reformulate the LCV equations (1) by using the affine tri-focal tensor and introducing the appropriate constraints in the LCV mapping process. In addition, in this paper we have only addressed extrinsic viewpoint variations, but it should also be possible to include intrinsic, shape variations using the approach described by (Dias and Buxton, 2005).

REFERENCES

- Bebis, G., Louis, S., Varol, T., and Yfantis, A. (2002). Genetic object recognition using combinations of views. *IEEE Transactions on Evolutionary Computation*, 6(2):132–146.
- Beymer, D. J. (1994). Face recognition under varying pose. In *Proc. IEEE Conf. CVPR*, pages 756–761.
- Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press.
- Dias, M. B. and Buxton, B. F. (2005). Implicit, view invariant, linear flexible shape modelling. *Pattern Recognition Letters*, 26(4):433–447.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman and Hall, London, 2nd edition.
- Goshtasby, A. (1986). Piecewise linear mapping functions for image registration. 19(6):459–466.
- Hansard, M. E. and Buxton, B. F. (2000). Parametric view-synthesis. In *Proc. 6th ECCV*, 1:191–202.
- Hastings, W. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.
- Koufakis, I. and Buxton, B. F. (1998). Very low bit-rate face video compression using linear combination of 2dfaceviews and principal components analysis. *Image and Vision Computing*, 17:1031–1051.
- Lamdan, Y., Schwartz, J., and Wolfson, H. (1988). On recognition of 3d objects from 2d images. *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1407–1413.
- Lee, M. W. and Ranganath, S. (2003). Pose-invariant face recognition using a 3d deformable model. *Pattern Recognition*, 36:1835–1846.
- Li, W., Bebis, G., and Bourbakis, N. (2004). Improving 3d object recognition based on algebraic functions of views.
- Metropolis, A., Rosenbluth, W., Rosenbluth, M., Teller, H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1092.
- Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, 7:308–313.
- Nene, S. A., Nayar, S. K., and Murase, H. (1996). Columbia Object Image Library (COIL-20). Technical Report CUCS-006-96, Department of computer science, Columbia University, New York, N.Y. 10027.
- Peters, G. and von der Malsburg, C. (2001). View reconstruction by linear combination of sample views. In *Proc. British Machine Vision Conference BMVC 2001*, 1:223–232.
- Shashua, A. (1995). Algebraic functions for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):779–789.
- Shewchuk, J. R. (2002). Delaunay refinement algorithms for triangular mesh generation. *Computational Geometry: Theory and Applications*, 22:21–74.
- Storn, R. and Price, K. V. (1997). Differential evolution - a simple and efficient heuristic for global optimization overcontinuous spaces. *Journal of Global Optimization*, 11(4):341–359.
- Ullman, S. and Basri, R. (1991). Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(10):992–1006.
- Zografos, V. and Buxton, B. F. (2007). Pose-invariant 3d object recognition using linear combination of 2d views and evolutionary optimisation. *ICCTA*, pages 645–649.