Bayesian Inferences for Object Detection

Rozenn Dahyot

[1] School of Computer Science and Statistics, Trinity College Dublin, Ireland, Rozenn.Dahyot@cs.tcd.ie

Moghaddam et al. (Moghaddam and Pentland, 1997) first proposed to perform object detection by modelling the marginal probability density function (pdf) of high dimensional features of appearance. Based on gaussian hypotheses, their approach is however not robust to outliers that can occur in images due, for instance, to cluttered backgrounds or partial occlusions. In (Dahyot et al., 2004), robustness has been improved by using better priors for the distribution of the errors encountered in the observations. Because the marginal pdf was not analytically available, the Maximum A Posteriori (MAP) pdf was used instead for detection (Dahyot et al., 2004). If both pdfs, marginal and MAP, are proportional under gaussian assumptions (MacKay, 1995), and therefore perform equivalently, we aim in this paper to compare both densities (marginal and MAP) for object detection using the same robust priors.

Dimensionality reduction and intra class variability.

A collections of high dimensional features $\{\mathbf{x}_k\}_{k=1,\cdots,K}$ are used to capture the variability of appearance in an object class $\mathcal{O}$. The N-dimensional vector $\mathbf{x}_k$ usually corresponds to colour or grey-level values of image pixels. It is assumed that the information contains in this data set can be captured by a variable Θ of smaller dimension $J << N$. The latent variable Θ retains the main variabilities in the class $\mathcal{O}$ and is linked with the variable $\mathbf{x}$ by $\mathbf{x} = f(\Theta) + \mathbf{w}^{\mathbf{r}}$ where $\mathbf{w}_{\mathbf{r}}$ is a reconstruction error. Several relations f, linear or non-linear, have been proposed in the literature (Saul and Roweis, 2003). Here, a simple linear relation f has been learned from the training set using the Principal Component analysis (PCA) (Murase and Nayar, 1995; Moghaddam and Pentland, 1997; Dahyot et al., 2004).

Modelling Observations.

Features $\{\mathbf{x}_k\}_{k=1,\cdots,K}$ used for training are assumed to be good representative of the object class, therefore clean from outliers. However, when comparing a new observation to the learned manifold, occurrences of gross errors due to cluttered background or partial occlusions, is a common issue. The observation $\mathbf{y}$ is then explicitly modelled as $\mathbf{y} = \mathbf{x} + \mathbf{w}^o = f(\Theta) + \mathbf{w}$ where $\mathbf{w}^o$ represents the observation noise, and $\mathbf{w}$ is the sum of the reconstruction error and the observation noise.

Bayesian inferences.

A standard approach to obtain a measure of similarity between the observation $\mathbf{y}$ and the class of interest is to compute the value of the marginal pdf $p_Y(\mathbf{y}|\mathcal{O})$. Assuming the independence of variables Θ and $\mathbf{w}$, the pdf can be computed by:

$$\begin{array}{ll} p_Y(\mathbf{y}|\mathcal{O}) & = \int p_{Y,\Theta}(\mathbf{y},\Theta|\mathcal{O})\ d\Theta \\ & = \int p_{Y|\Theta}(\mathbf{y}|\Theta,\mathcal{O}) \cdot p_\Theta(\Theta|\mathcal{O})\ d\Theta \\ & = \int p_W(\mathbf{w}|\mathcal{O}) \cdot p_\Theta(\Theta|\mathcal{O})\ d\Theta \end{array} \tag{1}$$

Another possible similarity measure corresponds to the value in $\mathbf{y}$ of the MAP pdf (Jordan and Weiss, 2002; Dahyot et al., 2004):

$$p_Y^{MAP}(\mathbf{y}|\mathcal{O}) = \max_\Theta p_{Y,\Theta}(\mathbf{y},\Theta|\mathcal{O}) = p_{Y,\Theta}(\mathbf{y},\hat{\Theta}_{MAP}|\mathcal{O}) \tag{2}$$

The class $\mathcal{O}$ is now characterised by a family of pdfs indexed by the latent variable Θ. This approach consists in first selecting the best function that explained the observation $\mathbf{y}$, and then computing its corresponding pdf value. The first step is performed in estimating $\hat{\Theta}_{MAP} = \arg\max_\Theta p_{Y,\Theta}(\mathbf{y},\Theta|\mathcal{O})$, and then computing $p_{Y,\Theta}(\mathbf{y},\hat{\Theta}_{MAP}|\mathcal{O})$.
Finally, now considering Θ as a parameter indexing the family of pdfs characterising the class $\mathcal{O}$, the value in $\mathbf{y}$ of the maximum likelihood (ML) pdf can also be computed as a similarity measure:

$$p_Y^{ML}(\mathbf{y}|\mathcal{O}) = \max_\Theta p_{Y|\Theta}(\mathbf{y}|\Theta,\mathcal{O}) = p_{Y|\Theta}(\mathbf{y}|\hat{\Theta}_{ML},\mathcal{O}) \tag{3}$$

where $\hat{\Theta}_{ML} = \arg\max_\Theta p_{Y|\Theta}(\mathbf{y}|\Theta,\mathcal{O})$.

Priors.

The integral in equation (1) has been solved analytically using normal assumptions (Tipping and Bishop, 1997; Moghaddam and Pentland, 1997). It is however difficult to solve it when expressions of $\mathcal{P}_W(\mathbf{w}|\mathcal{O})$ and $\mathcal{P}_\Theta(\Theta|\mathcal{O})$ are complex. Following (Dahyot et al., 2004), the distribution of the error is modelled using hard redescender robust M-estimator function. This expression allows us to deal with gross error occurring in the observations. The pdf of Θ is modelled in this experiment with the empirical distribution inferred using each training sample $\mathbf{x}_k$ for which the latent variable Θ_k is estimated with f in the training stage:

$$p_\Theta(\Theta|\mathcal{O}) = \frac{1}{K}\sum_{k=1}^{K} \delta(\|\Theta - \Theta_k\|) \tag{4}$$

Under those assumptions, the marginal is computed by summing all probabilities computed for all possible K latent variables Θ_k. The MAP probability retains only the maximum which allows the possibility of speeding up the exhaustive search. For the ML pdf, deterministic algorithms are used to compute for each observation $\mathbf{y}$ the maximum likelihood estimate $\hat{\Theta}_{ML}$ (Dahyot et al., 2004). Other priors can also be used (Dahyot et al., 2004; Vik et al., 2003) with different robust estimation strategies (Dahyot et al., 2004; Vik et al., 2003; Dahyot and Wilson, 2005).

Detection: classification between $\mathcal{O}$ and $\overline{\mathcal{O}}$

The detection of objects is performed by first spanning the set of observations and computing their probabilities. Receiver Operating Characteristic (ROC) curves are used to assess our detectors, by computing the detection rate (number of detected occurrences of $\mathcal{O}$ divided by the number of occurrences of $\mathcal{O}$ present in the test set) w.r.t. the false alarm rate (number of detected occurrences of $\overline{\mathcal{O}}$ divided by the number of occurrences of $\overline{\mathcal{O}}$ present in the test set) over a test set of images of objects and non-objects (~ 500 occurrences of $\mathcal{O}$ and ~ 60000 occurrences of $\overline{\mathcal{O}}$). Most samples of the objects in the test set present cluttered background and partial occlusions.
Figure 1 compares the ROC curves computed using all measures. In the test, the MAP pdf gives the best results, followed by the ML pdf and then the marginal pdf. The ML pdf performance is surpassed by the marginal at the top of the ROC curves. It is understood that in some difficult cases, when the observation shows too many outliers, the estimation of $\hat{\Theta}_{ML}$ is not accurate and, as a consequence, the ML pdf becomes less efficient than the marginal one. Concerning the computation time, the ML pdf, using a dedicated deterministic algorithm is the fastest followed by the MAP pdf and then the marginal.

Discussion.

If both pdfs, marginal and MAP, are proportional under gaussian assumptions (MacKay, 1995), this first experiment shows that better performances for detection can be expected when using the MAP under more realistic non-gaussian hypotheses. Both the MAP and ML pdfs perform better than the marginal. This experience gives better insights on the difference of the detection strategies presented in (Moghaddam and Pentland, 1997) and (Dahyot et al., 2004), and shows the potential superiority of using the MAP pdf as a detector.

References

Dahyot, R., Charbonnier, P., and Heitz, F. (2004). A bayesian approach to object detection using probabilistic appearance-based models. *Pattern Analysis and Applications*, 7(3):317–332.

Dahyot, R. and Wilson, S. (2005). Robust scale estimate for the generalized

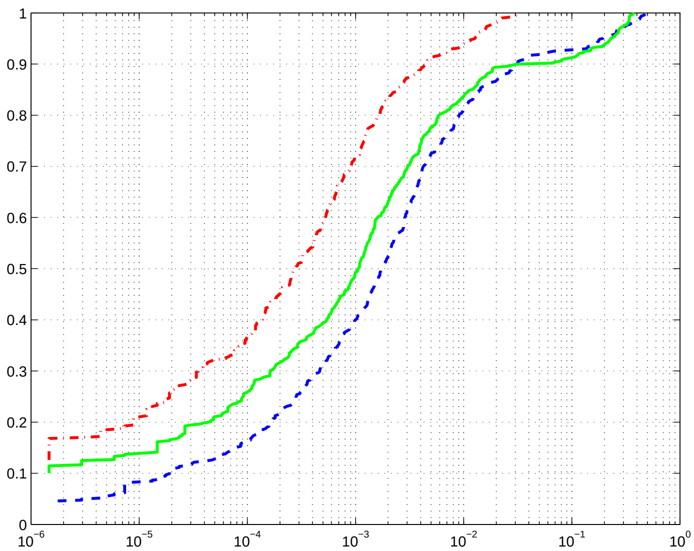

FIGURE 1. ROC curves: the MAP pdf (dash dot red, top), the ML pdf (continuous green, middle) and the marginal (long dash blue, bottom).

gaussian probability density function. Technical report, 05/08 Department of Statistics, Ireland.

Jordan, M. I. and Weiss, Y. (2002). Graphical models: Probabilistic inference. *in The Handbook of Brain Theory and Neural Networks*.

MacKay, D. (1995). Probable network and plausible predictions - a review of practical bayesian methods for supervised neural networks. *Network*, pages 469–505.

Moghaddam, B. and Pentland, A. (1997). Probabilistic visual learning for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):696–710.

Murase, H. and Nayar, S. K. (1995). Visual learning and recognition of 3-d objects from appearance. *International Journal of Computer Vision*, 14(1):5–24.

Saul, L. K. and Roweis, S. T. (2003). Think globally, fit locally: unsupervised learning of low dimensional manifolds. *The Journal of Machine Learning Research*, 4:119 – 155.

Tipping, M. and Bishop, C. (1997). Probabilistic principal component analysis. Technical Report NCRG/97/010, Neural Computing Research Group, Aston University.

Vik, T., Heitz, F., and Charbonnier, P. (2003). Mean shift-based bayesian image reconstruction into visual subspace. In *IEEE International Conference on Image Processing*, Barcelona, Spain.