An Empirical-Statistical Agenda for Recognition

David Forsyth

Computer Science Division, U.C. Berkeley, Berkeley, CA 94720, USA
daf@cs.berkeley.edu,
WWW home page: http://www.cs.berkeley.edu/~daf

This piece first describes what I see as the significant weaknesses in current understanding of object recognition. We lack good schemes for: using unreliable information — like radiometric measurements — effectively; integrating potentially contradictory cues; revising hypotheses in the presence of new information; determining potential representations from data; and suppressing individual differences to obtain abstract classes. The problems are difficult, but none are unapproachable, given a change of emphasis in our research.

All the important problems have a statistical flavour to them. Most involve a change of emphasis from the detailed study of specific cues to an investigation of techniques for turning cues into integrated representations. In particular, all have a statistical flavour, and can be thought of as inference problems. I show an example that suggests that methods of Bayesian inference can be used to attack these difficulties.

We have largely mapped out the the geometrical methods we need. Similarly, all the radiometric information that conceivably could be useful already exists. I believe that the next flowering of useful vision theories will occur when we engage in an aggressive study of statistics and probabilistic modelling, particularly methods of Bayesian inference.

1 What we do badly

The weaknesses in current understanding of object recogntion all appear to come from our conception of an object model as a passive, unstructured repository of detailed geometric information.

1.1 Managing information

The literature is rich with individual cues to object identity, from surface colour to geometric primitives. It is unusual to have these cues agree on anything; the resulting embarrassment is avoided by not comparing the cues, or by ignoring such effete measures as colour or texture. This is a serious mistake. It is quite clear that it is generally better to have more cues, even if some are more reliable than others.

Our use of colour and shading information is still uncomfortably weak, probably because we don't have models that handle (or ignore) the nasty physical effects of interreflections or colour bleeding. I'm not aware of the use of texture in recognition, except in template matching approaches. Texture doesn't really

lend itself to direct template matching — not every tiger has every stripe in the same place — but it can be described and these descriptions contain information. We ignore this information, probably because we don't have mechanisms for integrating it with other cues, and because we haven't properly formulated models that cope with individual variations (abstraction again!).

1.2 Abstraction

Performing recognition at a properly abstract level (where the "fish", "person" or "bicycle" decision is made before the species, height or manufacturer are known) is something we know nothing about. We know from computer science that hierarchies are the proper way to access large numbers of objects efficiently, and there is some evidence [20, 21] that people make different kinds of decision about object identity ("category-level" and "instance-level" discrimination), suggesting the presence of a hierarchy. In turn, this suggests that abstraction may help with efficiency.

All current recognition algorithms perform recognition at the level of individual instances in a flat modelbase; any category-level recognition is performed later. This leads to the characteristic ludicrously inefficient search through models. Implicitizing this search by using geometric invariants simply creates clogged hash tables; it does not solve the problem because it does not scale.

1.3 Segmentation

There is a suspicious profusion of white objects on black backgrounds in the current recognition literature (significant variants include multihued objects on green backgrounds, and black objects on white backgrounds). This is a product of the view that segmentation and recognition are distinct problems, and that when the early vision community gets round to solving segmentation, everything will be ok. The view is pernicious, and the practices it encourages are dubious. A recognition system that operated at a sensible level of abstraction would draw broad, generic distinctions between objects at first; a knowledge of these kinds of distinctions is precisely what the starting stages of segmentation needs. In other words, objects are segmented in wholes because they are objects; we should rig the segmentation process to find the kind of evidence that leads to the objects we are looking for.

This view of segmentation is usual in practice — this is what interest operators, lines and conics are all about — but exploring the full power of the view hasn't been easy because we have only the vaguest notion of what the evidence should be. The old-fashioned view is that objects should be modelled as composites of primitives — the term "part" is often used — and that finding image evidence of the presence of these primitives is what segmentation is about [2]. These primitives might be shapes (hence the GC/SHGC/Geon/etc. debate), but they might also be characteristic shading or colour events. The strength of this view is that one can see, dimly, how to build sensible recognition systems like this: one builds repeated segmentation processes that find very general, widely

used primitives first and then assemble these primitives into larger and larger regions of image evidence. The segmentation is practicable, because *at each stage, we know what we're looking for*. Again, this is old news, but basically sensible news — for example, now the fuss about appearance based recognition has died down, that community is moving very quickly in this direction [9], precisely because it is the only way to deal with segmentation and complex model bases. The weakness is that we don't know what the primitives are, what the models look like, or how to build the models.

1.4 Learning

It would be a good thing if our recognition systems could recognise many objects, and if the recognition process did not require substantial reengineering each time a new object was encountered. Ideally, models of new objects could be obtained by showing the system various instances in various views. This is not necessarily an argument for *statistical* learning theory; these properties are have been built into geometric reasoning systems [22], although doing so generally requires some form of implicit statistical reasoning.

The selection of measurements, broadly interpreted, is poorly understood. Typical current recognition systems will use some set of geometrical measurements to determine whether an object is present; of course, there will be alternative sets of measurements that could be used. Which one should we use, and why? This is a narrow version of the problem of primitives — which primitives should we use in recognition, and why? Both problems appear, in abstract, to be model selection problems.

1.5 Nagging issues

The state of verification is a disgrace; either we shouldn't do it (which could be preferred on the grounds that we're only doing it because we haven't assembled much of the available evidence for our hypothesis before testing it), or we should do it properly. The current practice of counting edge points and crossing fingers simply isn't good enough. The problem is one of hypothesis testing (or of discrete model selection) — to what degree does the image evidence in this region support the following object hypothesis? This is a fairly straightforward problem with relatively straightforward statistical solutions — model selection, again.

2 A brief sketch of Bayesian inference

Probability provides a mechanism for comparing instances with a collective of previous examples; this mechanism makes it possible to combine evidence from various sources. The process of comparison is through a probabilistic model of the measurements produced given the state of the world. For example, we might take a large number of pictures of sunsets and estimate

$$P(\text{big red blob}|\text{picture is a sunset})$$

by a frequency estimate. This number could be interpreted either as a statement of expected frequencies, or as a "degree of belief" that a big red blob will appear in a picture of a sunset. It gives the probability of measurements, given the state of the world; this term is often referred to as the *conditional*, or *likelihood*.

The main use for a probabilistic model in recognition is inference. Generally, we expect to have a probabilistic model of the state of the world before we measured it. This is the prior — in the example, the probability that a picture taken from our collection is a picture of a sunset, or

$$P(\text{picture is a sunset})$$

The Bayesian philosophy says that all our knowledge of the state of the world is encapsulated in the *posterior* — the probability of a world state given observations. The posterior accounts for the effect of observations on the probability that (in our example) a picture taken from our collection is a picture of a sunset. By Bayes' rule, the posterior is proportional to the product of the prior and the likelihood, so we have

$$\begin{aligned} P(\text{picture is a sunset}|\text{big red blob} \propto & P(\text{big red blob}|\text{picture is a sunset}) \\ & \times P(\text{picture is a sunset}) \end{aligned}$$

Most computer vision researchers have seen this expression many times without any great delight. Its great importance is that *generative* models — which give the way that data is produced, given the state of the world — can be turned into recognition models just by multiplying by the prior. One common objection — that priors can be arbitrary — is, I think, empty; all but the silliest choices of prior are overwhelmed by data in the kinds of problem we wish to solve. The real problem is that we need to be able to compute with the resulting posterior, and that is where difficulties arise.

2.1 How probabilistic models can address our problems

The simple relationship between generative models and inference is the attraction of Bayesian models. If one accepts the basic hypothesis of the Bayesian philosophy — that the posterior encapsulates our knowledge of the world — information integration is simple. One just forms the posterior corresponding to the measurements available. No additional complications appear, because the generative model (or likelihood — which describes the process that leads to the measurements) is usually easy to obtain.

In the following piece, Joe Mundy makes the case that there is no canonical structure of class abstraction and so no basis for dealing with the world at any level other than that of instances. This may well be sound philosophy, but it is impractical system architecture. The advantage of class abstraction is, basically, that one kind of measurement can do multiple duty. Thus, it is worth finding extended regions with near parallel sides, because there are an awful lot of things that look rather like cylinders. Class hierarchies appear to offer efficient representations by making relatively small numbers of decisions to identify

relatively large numbers of objects. If this promise can be realised, the fact that the hierarchy is artificial is irrelevant.

Probabilistic models simplify constructing class hierarchies because they can encode explicitly the variation between instances of a class. This yields an immediate solution to an old problem with geometric primitives: it is hard to prove anything useful about objects that are not *exactly* instances of the primitive class, but vary only slightly from instances. The solution is to build a likelihood model around the measurement. Thus, for example, if we are looking for human limbs we can benefit from the fact that the outlines are not only *not* straight and parallel, but the way that they differ from being straight and parallel is structured.

Probabilistic models may also help to determine appropriate primitives. I do not agree with Mundy's argument that a primitive decomposition is — or even should be — canonical. It seems more constructive to view a decomposition into parts, or primitives, as a convenience for representational efficiency. In this view, the distinctive properties of primitives are that they occur on many objects in similar forms; that their presence is a useful guide to the presence of an object; and that their presence leads to distinctive image properties to guide inference. These appear to me to be statistical criteria.

All these comments offer some hope for a rational theory of segmentation, *if we can extract information from posteriors easily.*

2.2 Using posteriors

The attraction of the Bayesian view is that (given a decent generative model) all problems with integrating information disappear. Of course, the tricky bit is extracting information from the posterior, and this has tended to be the problem in the past. It is very easy to set up problems where the priors and the likelihoods all appear simple, and yet the posterior is very hard to handle (the colour constancy example given below is of this kind). Vision problems are too big and too disorderly to use conjugate distributions (an old-fashioned dodge of using models where the prior, likelihood and posterior all turn out to have an "easy" form). We might decide to choose a world model that maximises the posterior, but how do we get this maximum? Bayesian segmentation using Markov random fields foundered on this point.

The currently fashionable view in the statistical community says that information can be extracted from the posterior by drawing a large number of samples from that distribution. Thus, for example, if we wanted to decide whether to blow something up or not based on image evidence, we would form the posterior, draw samples from this, compute an expected utility for blowing it up and for not blowing it up by averaging the utilities for these samples, and choose the option with the larger expected utility[1].

Drawing samples from a posterior is not at all easy. Markov chain Monte Carlo methods appear to be the answer. A typical algorithm is the Metropolis-

[1] Right now, it would have to be a fairly slow target; but computers get faster.

Hastings algorithm, which would produce in this case a sequence of hypotheses, by taking an hypothesis T_i and proposing a revised version, T_i'. The new hypothesis T_{i+1} is either T_i or T_i', depending (randomly) on how much better the posterior associated with T_i' is. Once sufficient iterations have completed, all subsequent T_i are samples drawn from the posterior; the number of iterations required to achieve this is often called the *burn in* time. These samples may or may not be correlated; if this correlation is low, the method is said to *mix* well. It is known how to apply this algorithm to chains whose domain of support is complicated (for example, the number of hypotheses may not be known *a priori*) [7].

Metropolis-Hastings algorithms should be viewed as a kind of souped up hypothesize and test process. We propose a representation of the world and accept or reject it based on the posterior; our representation of the *posterior* then consists of a large set of accepted proposals. This view justifies using current vision algorithms as a source of proposals. The crucial improvement is that we can use different, incompatible algorithms as distinct sources of proposals, and the samples we obtain represent the posterior incorporating all available measurements. The example in section 3 illustrates this approach in greater detail.

There are some serious algorithmic problems: it is not possible to tell reliably whether a chain has burnt in by looking at the samples the chain produces; chains can mix extremely slowly, and often do if not very carefully designed [10]; and the difference between a successful algorithm and a catastrophic failure rests on the proposal process. Exact sampling, or coupling from the past, is a (not terribly practical) method for dealing with the first problem [19, 17]; the other two are not going away anytime soon. The advantages of representing ambiguity and error explicitly appear to outweigh these difficulties.

3 An example: colour constancy by sampling

Colour constancy is a good simple example that has some of the flavour of recognition, in the sense that we are using a model to make inferences from image observations. In the simplest version, we are in a world of flat frontal surfaces whose diffuse reflectances belong to a low dimensional linear family, illuminated by a coloured source. We do not know the colour of the source, and wish to determine surface colour, whatever the source colour. The problem can seldom, if ever be solved exactly.

There are many different mechanisms, each of which gives a different estimate. It is usual to assume that the illuminant changes slowly over space; cues include an assumption of constant average surface colour [3, 8, 11], the subspace of receptor space to which the surfaces map [15], the presence of sharp changes in image brightness [13], the fact that specularities typically take the source colour [12, 14], and physical constraints on reflectance and/or illuminant [5, 4]. All of these cues are basically valid and all should be used. Let us go back to the case of colour constancy. We would like to use the following constraints:

- Illuminants vary only slowly over space.
- Specularities yield cues to surface colour.
- Illuminants and reflectances are drawn from finite dimensional linear families.
- Reflectances are everywhere above 0.012 and below 0.96 in value (this gives an 80:1 dynamic range at each wavelength, consistent with other methods and physical evidence
- Illuminants are everywhere positive.

We ignore average reflectances, as they will be covered by the prior.

3.1 The generative model

We model surface reflectances as a sum of basis functions $\phi_j(\lambda)$, and assume that reflectances are piecewise constant:

$$s(x, y, \lambda) = \sum_{j=0}^{n_s} \sigma_j(x, y)\phi_j(\lambda)$$

Here $\sigma_j(x, y)$ are a set of coefficients that vary over space according to some model; in the example we describe, they are constant in a grid of boxes, where the grid edges are not known in advance.

Similarly, we model illuminants as a sum of (possibly different) basis functions ψ_i and assume that the spatial variation is given by the presence of a single point source positioned at $\mathbf{p}$. The diffuse component due to the source is:

$$e_d(x, y, \lambda, \mathbf{p}) = d(x, y, \mathbf{p}) \sum_{i=0}^{n_e} \epsilon_i \psi_i(\lambda)$$

where ϵ_i are the coefficients of each basis function and $d(x, y, \mathbf{p})$ is a gain term that represents the change in brightness of the source over the area viewed. The specular component due to the source is:

$$e_m(x, y, \lambda, \mathbf{p}) = m(x, y, \mathbf{p}) \sum_{i=0}^{n_e} \epsilon_i \psi_i(\lambda)$$

where $m(x, y, \mathbf{p})$ is a gain term that represents the change in specular component over the area viewed.

Standard considerations yield a model of the k'th receptor response as:

$$\begin{aligned} p_k(x, y) &= \int s(x, y, \lambda)(e_d(x, y, \lambda, \mathbf{p}) + e_m(x, y, \lambda, \mathbf{p}))\rho_k(\lambda)d\lambda \\ &= d(x, y, \mathbf{p}) \sum_{i,j} g_{ijk}\epsilon_i\sigma_j(x, y) + m(x, y, \mathbf{p}) \sum_i h_{ik}\epsilon_i \end{aligned}$$

where $g_{ijk} = \int \rho_k(\lambda)\psi_i(\lambda)\phi_j(\lambda)d\lambda$ and $h_{ik} = \int \rho_k(\lambda)\psi_i(\lambda)d\lambda$. In this case, $\sigma_j(x, y)$ is piecewise constant with some spatial model — in what follows, we

assume that it is piecewise constant on a grid, *but we do not know what the grid edges are*. The spatial model for the illuminant follows from the point source model, where $\mathbf{p}$ is the position of the source and $m(x, y, \mathbf{p})$ is obtained using Phong's model of specularities.

We choose a uniform prior for reflectance coefficients. We expect illuminants to have no chromatic bias, and so use a Gaussian prior, whose mean is white; we allow a fairly substantial standard deviation, to allow for illuminants that are coloured.

Thus, the generative model is:

- sample the number of reflectance steps in x and in y (k_x and k_y respectively);
- now sample the position of the steps ($\mathbf{e}_x$ and $\mathbf{e}_y$ respectively);
- for each tile, sample the reflectance for that interval from the prior (σ_j^m for the m'th tile;
- sample the illuminant coefficients ϵ_i from the prior;
- sample the illuminant position $\mathbf{p}$ from the prior;
- and render the image, adding Gaussian noise.

So we have a likelihood,

$$P(\text{image}|k_x, k_y, \mathbf{e}_x, \mathbf{e}_y, \sigma_j^m, \epsilon_i, \mathbf{p})$$

The posterior is proportional to:

$$\begin{aligned} P(\text{image}|k_x, k_y, \mathbf{e}_x, \mathbf{e}_y, \sigma_j^m, \epsilon_i, \mathbf{p}) \times & \; Prior(kv_x)Prior(kv_y) \\ \times & \; Prior(\mathbf{s}_x)Prior(\mathbf{s}_y) \\ \times & \prod_{m \in \text{tiles}} Prior(\sigma_j^m) \\ \times & \; Prior(\epsilon_i)Prior(\mathbf{p}) \end{aligned}$$

and all we have to do is draw samples from this.

3.2 The sampling process

The sampling process is straightforward MCMC. Proposal moves are of four types:

- **Birth of a step:** the details are as in example 1 of [7], with the exception that we derive a proposal distribution for the position of the step from the image gradient, so that we are more likely to propose a step where the gradient is high.
- **Death of a step:** as in example 1 of [7].
- **Change position of a step:** as in example 1 of [7].
- **Change reflectance and illumination:** this has some subtle pitfalls. It is tempting to fix illumination and change reflectance, then fix reflectance and change illumination, because both steps will involve sampling a Gaussian, which is easy. This, it turns out, is a bad idea, because the chain moves

extremely slowly if we do this. The explanation is quite simple; given a reflectance/illumination pairing that is quite good, small changes lead to huge increases in the error. Thus, the process will lead to a reflectance that works well with the current illuminant or an illuminant that works well with the current reflectance and will move extremely slowly. Instead, we use a method due to Neal [18] that suppresses random walk by adjoining momentum variables and then modelling the state as the position of a particle in an energy field; typically, the state moves to positions of large posterior value quickly, at which point we throw away the momentum variables.

3.3 Experimental examples

I show some results obtained using a real dataset from [5]. This data was photographed with a CCD camera, then displayed on a CRT, photographed from that screen with a film camera, subjected to unknown printing processes and then scanned from the published paper; I used such battered data deliberately, to illustrate the potential power of the process. There are no specular components in this dataset, making the specular reasoning simple. I obtained a basis using the mechanisms of Marimont and Wandell [16]. Constraints on coefficients were estimated using a graphical method.

In each case, the spatial model placed edges at the right points; this is hardly surprising, as each column and row edge has some high contrast points. Figure 1 shows a scatter plot of reflectance samples estimated for various corresponding tiles for images obtained under different coloured illuminants, under the assumption that each image was completely independent. Since the samples lie in reasonably close groups compared with the receptor response groups, the algorithm is displaying constancy. These samples contain not only estimates of reflectance, but also *information about a reasonable range of solutions*. This is crucial, and powerful.

Current colour constancy algorithms cannot deal with prior knowledge about the world. This algorithm can. For example, consider the effect of knowing that tile i in an image under white light is "the same" as tile j in an image obtained under purple light. It is quite plausible to want to know this — knowing that an object is a bus can and should affect our reports on its colour. In our representation, we can perform this calculation by *resampling* the set of samples. We sample pairs of representations — one obtained under the white light, the other under the purple light — so that pairs where the two reflectances are similar are represented more often (this can all be made formal). Figure 2 shows the effect on results; this added information has reduced uncertainty. Further details appear in [6].

4 How inference methods can help address our problems

Probabilistic models are difficult to use and to set up and there are no reliable algorithms for handling probabilistic models on the scale that vision will require.

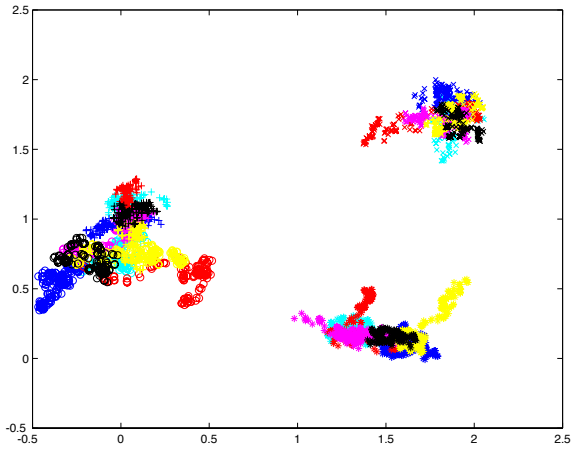

Fig. 1. *The first two components of surface reflectance samples, plotted on the same axes for four different surfaces. Each sample is colour keyed to the image from which it was obtained; red samples for the red image, etc, with black corresponding to the white image. The circles show samples of the reflectance coefficients for the blue surface at the top left corner of the Mondriaan; the stars for the yellow surface in the second row; the plusses show samples for the orange surface in the top row of the Mondriaan and the crosses for the red surface in the bottom row. Notice that the smear of samples corresponding to a particular surface in one image intersects, but is not the same as, the smear corresponding to that surface in another. This means that the representation envisages the possibility of their being the same, but does not commit to it.*

The paradigm is the right one; we learned to use geometric models well, and we should now be directing our efforts toward using probabilistic models well.

4.1 Information integration

While sampling algorithms are currently slow and difficult to build well, the advantages that compensate for this are:

- We get a good representation of all the conclusions that can reasonably be drawn from the data (i.e. the posterior).
- It is easy to see how to include other forms of information in the reasoning process; rewrite the likelihood, add more proposal mechanisms and proceed. If a form of conditional independence applies (which it does in useful cases), we can resample an existing set of samples.
- We do not need to abandon current algorithms to achieve this; instead, we can cloak them in probability and use them as a proposal process (as with the use of the gradient to propose step positions).

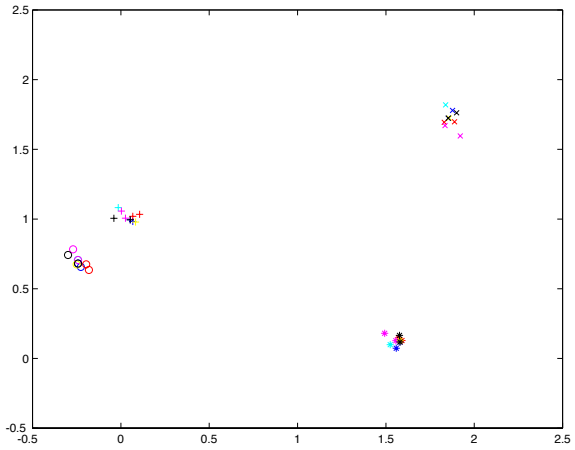

Fig. 2. *The first two components of surface reflectance samples, plotted on the same axes for four different surfaces. These come from the samples of figure 1,* resampled *under the assumption that the blue surface in the top left hand corner of the Mondriaan is the same for each image. We use the same representation and axes as in that figure. Notice that this single piece of information hugely reduces the ambiguity in the representation.*

4.2 Segmentation

The colour constancy algorithm described above had a segmenter built into it with very little fuss — this is the spatial model of reflectances being piecewise constant in tiles. This illustrates a strength and a weakness of sampling methods.

The strength is that the relationship between segmentation and recognition is built in; the image is segmented into regions that are given by the model, and we are not required to perform generic segmentation. The weakness is that the model "hides" segmentation in the proposal processes; if we have good proposal processes, the problem of segmentation largely disappears, but if they are bad, we may never see a sensible interpretation. The good news is that there is some general information about what makes a good proposal process (e.g. [18]), and that it is easy to obtain proposal processes from current knowledge of vision.

4.3 Learning, representation and primitives

The mechanisms for learning likelihoods ($P(measurement|model)$) from data exist. We expect that a modelbase would be represented by a complicated likelihood function, that would have a hierarchical structure and intermediate representations of one form or another. For example, a model for recognising clothed people is likely to manipulate a representation of what clothing looks like at one stage, and what arms look like at another. The ideal is to lay out efficient and accurate models automatically from pictures. We need to determine what primitives to use. Geometric reasoning has been very little help here in the past. The

difficulty comes from obvious sources: a useful geometric primitive will have a characteristic appearance in an image that is informative about its 3D structure, and must also be useful for many objects and unlikely to arise from non-objects. These criteria are statistical, rather than geometric in nature. Amit and Geman have constructed some promising algorithms for automatically determining simple image primitives [1], but there is a long way to go.

4.4 Feature selection

Feature selection has received embarassingly little attention in the recognition community. The question is simple; which measurements should be used to recognise an object? It appears in all current algorithms, in both the indexing and the verification phase. The answer can only be statistical.

The main difficulty with statistical work on feature selection is the absence of a notion of computational or measurement cost. The Bayesian answer to feature selection is to use all measurements, appropriately weighted to take account of their reliability. This is all very well, but it misses the fact that measurements may be difficult or expensive to make and that there is a limited amount of computation available. It is quite surprising that this question, which is obviously central in object recognition, has received so little attention in the literature.

5 Summary

Probabilistic models can be used to address the greatest weaknesses in our understanding of object recognition: integration, abstraction, segmentation and feature selection. The great power of probabilistic models is that they can encapsulate sources of variation whose origins are too complex to bear investigation — for example, the variations in the outline of each person's limb due to different muscle bulk. Sampling methods are good enough to produce attractive solutions to simple vision problems. There are substantial challenges in building efficient computational implementations, but these difficulties are worth addressing. Geometry and radiometry are simply not our most important areas of ignorance. In fact, we can currently predict the appearance of objects rather well — i.e. we can build quite good likelihood models. We should now be studying modelling and inference instead, because these are the theories on which a systematic view of the overall process of recognition will be based.

References

1. Y. Amit, D. Geman, and K. Wilder. Joint induction of shape features and tree classifiers. *IEEE T. Pattern Analysis and Machine Intelligence*, 19(11):1300–1305, 1997.
2. T.O. Binford. Visual perception by computer. In *Proc. IEEE Conference on Systems and Control*, 1971.

3. G. Buchsbaum. A spatial processor model for object colour perception. *J. Franklin Inst.*, 310:1–26, 1980.
4. G. Finlayson. Colour in perspective. *IEEE T. Pattern Analysis and Machine Intelligence*, 18:1034–1038, 1996.
5. D.A. Forsyth. A novel algorithm for colour constancy. *Int. J. Computer Vision*, 5:5–36, 1990.
6. D.A. Forsyth Sampling, Resampling and Colour Constancy. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1991.
7. P.J. Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732, 1995.
8. H. Helson. Some factors and implications of colour constancy. *J. Opt. Soc. America*, 48:555–567, 1934.
9. C-Y. Huang, O.T. Camps, and T. Kanungo. Object recognition using appearance-based parts and relations. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 877–83, 1997.
10. Mark Jerrum and Alistair Sinclair. The markov chain monte carlo method: an approach to approximate counting and integration. In D.S.Hochbaum, editor, *Approximation Algorithms for NP-hard Problems*. PWS Publishing, Boston, 1996.
11. D.B. Judd. Hue, saturation and lightness of surface colors with chromatic illumination. *J. Opt. Soc. America*, 30:2–32, 1940.
12. G.J. Klinker, S.A. Shafer, and T. Kanade. A physical approach to color image understanding. *Int. J. Computer Vision*, 4(1):7–38, 199.
13. E.H. Land and J.J. McCann. Lightness and retinex theory. *J. Opt. Soc. Am.*, 61(1):1–11, 1971.
14. H.C. Lee. Method for computing the scene-illuminant chromaticity from specular highlights. *J. Opt. Soc. Am.-A*, 3:1694–1699, 1986.
15. L.T. Maloney and B.A. Wandell. A computational model of color constancy. *J. Opt. Soc. Am*, 1:29–33, 1986.
16. D.H. Marimont and B.A. Wandell. Linear models of surface and illuminant spectra. *J. Opt. Soc. Am.-A*, 9:1905–1913, 1992.
17. D. J. Murdoch and P. J. Green. Exact sampling from a continuous state space. *To appear in the Scandinavian Journal of Statistics*, 1998.
18. R.M. Neal. Probabilistic inference using markov chain monte carlo methods. Computer science tech report crg-tr-93-1, University of Toronto, 1993.
19. James G. Propp and David B. Wilson. Exact sampling with coupled markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9:223–252, 1996.
20. E. Rosch and C.B. Mervis. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7:573–605, 1975.
21. E. Rosch, C.B. Mervis, W.E. Gray, D.M. Johnson, and P. Boyes-Braim. Basic objects in natural categories. *Cognitive Psychology*, 8:382–439, 1976.
22. A. Zisserman, D.A. Forsyth, J.L. Mundy, C.A. Rothwell, and J.S. Liu. 3d object recognition using invariance. *Artificial Intelligence*, 78:239–288, 1995.