Likelihood Models for Template Matching Using the PDF Projection Theorem

A. Thayananthan * R. Navaratnam * P. H. S. Torr † R. Cipolla *

* University of Cambridge
Department of Engineering
Cambridge, CB2 1PZ, UK
http://mi.eng.cam.ac.uk/ cipolla

† Oxford Brookes Univeristy
Department of Computing
Oxford, OX33 1HX, UK
http://wwwcms.brookes.ac.uk/ philiptorr

Abstract

Template matching techniques are widely used in many computer vision tasks. Generally, a likelihood value is calculated from similarity measures, however the relation between these measures and the data likelihood is often incorrectly stated. It is clear that accurate likelihood estimation will improve the efficiency of the matching algorithms. This paper introduces a novel method for estimating the likelihood PDFs accurately based on the PDF Projection Theorem, which provides the correct relation between the feature likelihood and the data likelihood, permitting the use of different types of features for different types of objects and still estimating consistent likelihoods. The proposed method removes the normalization and bias problems that are usually associated with the likelihood calculations. We demonstrate that it significantly improves template matching in pose estimation problems. Qualitative and quantitative results are compared against traditional likelihood estimation schemes.

1 Introduction

Template matching is commonly used in many computer vision applications such as feature-based tracking, object recognition and stereo-matching. Templates can either be learned from exemplar images [6, 18] or created from models [1, 14, 16]. Matching methods differ according to the type of features used: edge pixels [6, 13, 17], intensity image patches [3] and wavelets coefficients [19] to name a few. In general template matching requires similarity measures between the features of a template and the query image. Image intensity patches are often compared by normalized cross-correlation whereas Hausdorff and chamfer measures are popular with edge-based features.

For object detection, template matching is performed by matching the template at all locations, scales and orientations. If the likelihood value, obtained from one or several of the matching measures, is above a threshold, then a possible matching of the template is reported. However many template matching schemes calculate the likelihoods in adhoc ways (e.g simply exponentiating a similarity measure). For example, the raw chamfer score is thresholded in many chamfer matching schemes [6, 17]. Such methods ignore the necessity for normalizing the likelihood values and bias towards likelihoods

BMVC 2004 doi:10.5244/C.18.97

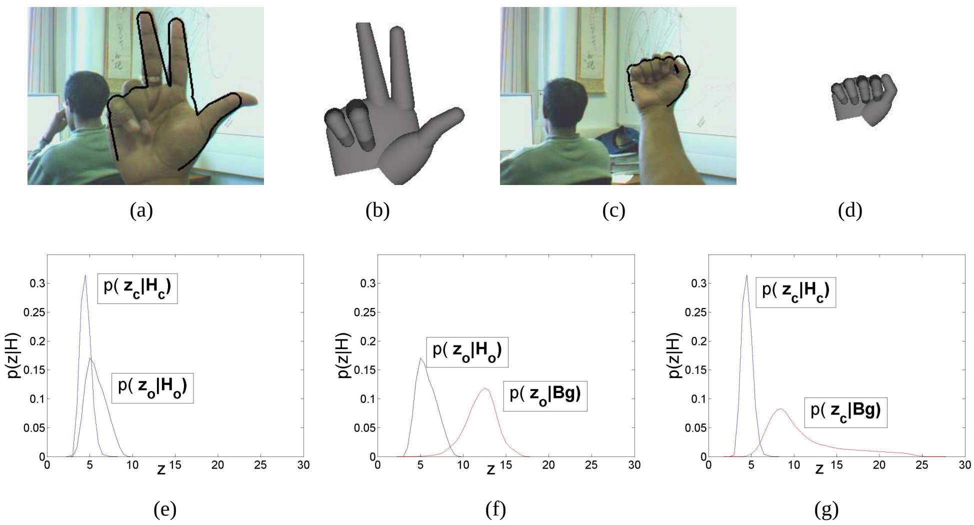


Figure 1: **Likelihood distributions for chamfer matching.** *(a, c) show images of open and closed hands with the corresponding templates manually superimposed. (b, d) the equivalent poses of the 3D model from which the templates are created. (e, f, g)* $p(z_o|H_o)$ *is the likelihood distribution of the chamfer score, z, obtained from matching the open hand template with open hand images. Similarly* $p(z_c|H_c)$ *is the distribution for matching a closed hand template with closed hand images.* $p(z_o|Bg)$ *and* $p(z_c|Bg)$ *are the distributions learned from matching the open hand template and closed hand template with background clutter. Each distribution was learned from 1000 images.*

with large partition function values. Figure 1 shows two different templates from a hand pose matching application [16]. The feature likelihood distributions for these templates are quite distinct as shown in figure 1 (e). If the matching decisions are made from one chamfer threshold for both templates, we will either end up with a large number of false positives (with a large threshold) or low detection rate (with a low threshold). Performance can be improved by using different thresholds for different templates. But optimal performance (in terms of detection and false positive rate) will be obtained by using the likelihood distributions.

Another more subtle but equally important, issue in likelihood estimation is the ambiguity arising from the discriminatory power of the matching features. In chamfer matching, for example we are depending on the chamfer score obtained from matching an object template to an image, to make a decision. How well does the range of chamfer scores for the object separate the background clutter and other type of objects? Figure 1 (f, g) compares the feature likelihood PDFs (Probability Distribution Function) and the clutter likelihood PDF of two hand templates. The feature likelihood PDFs is obtained by matching the template with similar looking hand poses, whereas the clutter likelihood PDF is estimated from matching the template to background clutter. At a chamfer score of 6.5 the object likelihood is 0.11 and the clutter likelihood is 0.005 for the open hand template. The clutter likelihood seems to provide an uncertainty measure associated with matching the particular template at this particular chamfer score. It is also connected to the false positive rate of the template matching scheme. Hence, it intuitively makes sense to somehow incorporate this uncertainty measure within the matching framework. However, many template matching schemes ignore this potential statistic for capturing the inherent uncertainty.

The above example suggests that discriminability of the matching features could be approximately quantified by obtaining a matching score in a reference hypothesis. This is the underlying principle of the Likelihood Ratio Test (LRT). It could be easily shown that using likelihood ratios minimizes the Bayesian error in two-class decision problems [20]. It is used in many areas of computer vision in simpler formulations (e.g. skin colour classification [8] and limb detector [15]). But it is not clear how to evaluate accurate likelihoods in more complex situations such as template-based matching schemes.

In this context, Baggenstoss [2] proposed a novel way of estimating raw data likelihood PDFs, from the feature likelihood PDFs, using the *PDF projection theorem* and its corollary *class-specific feature theorem*. The PDF projection theorem explains the idea of a likelihood ratio and its implications within a broader statistical interpretation and provides a general framework to estimate the likelihood PDFs more accurately and consistently than most of the adhoc schemes commonly used. The Class-specific feature theorem extends this mechanism to specific features extracted by each class or hypothesis, thus avoiding the curse of dimensionality. First, it permits us to estimate PDFs only on features that are specific to a class. Then it transforms the class-specific feature likelihoods into raw data likelihoods. We can now use different features for different classes and still obtain comparable likelihoods in raw data space. We shall see later that this is important when we compare likelihoods obtained from matching different templates to different sets of edge features. Caputo and Niemann [5] obtained significant improvement in object recognition tasks using class-specific features. Minka [12] proposed the use of the PDF projection theorem for obtaining exemplar-based likelihoods.

In this paper, we develop a likelihood PDF estimation scheme for template-based matching through the utilization of the PDF projection theorem. By incorporating the uncertainty of matching to features in a principled manner, we not only improve the efficiency of matching, but also obtain a mechanism for quantifying the uncertainty involved. The advantages are manifold; firstly the false positive rates are greatly reduced, secondly the likelihoods are now comparable across the different regions of the template space and thirdly the likelihood values can be seamlessly integrated with a temporal prior to produce a more precise posterior distribution.

This paper describes the adaptation of PDF projection theorem for chamfer matching, however the methods derived could be easily adapted to other types of template matching. Section 2 of this paper reviews some related work in chamfer matching. Section 3 introduces the PDF projection theorem and its corollary, the class specific feature theorem. We explain how we adapt the PDF projection theorem to the chamfer matching framework in section 4. Section 5 details the experiments conducted to highlight the advantages of accurate likelihood estimation in hand pose detection through chamfer matching. We conclude in section 6.

2 Related Work

We use PDF projection theorem to calculate robust likelihoods for chamfer matching with edge features, However, the theory is applicable to all type of features. Some previous work on chamfer matching and likelihood estimation is reviewed in this section.

Chamfer matching is frequently used as a cost function in many template-based matching schemes and improved versions have been used for object recognition and contour alignment. Borgefors [4] introduced hierarchical chamfer matching, in which a coarse-to-fine search is performed using a resolution pyramid of the image. Olson and Huttenlocher [13] use a template hierarchy to recognize three-dimensional objects from different views.

Gavrila uses approximately 4,500 shape templates to detect pedestrians in images [6]. To avoid exhaustive search, a template hierarchy is formed by bottom-up clustering based on the chamfer score. A number of similar shape templates are represented by a cluster prototype. This prototype is first compared to the input image, and only if the error is below a threshold value, the templates within the cluster are compared to the image. The use of a template hierarchy is reported to result in a speed-up of three orders of magnitude compared to exhaustive matching. However, [6, 13] do not estimate proper likelihood PDFs and operate on raw cost function values losing some of the advantages aforementioned in the previous section.

Jojic and Frey [7] proposed a probabilistic framework for tracking with templates. Toyama and Blake [18] use a similar approach to track humans. The templates are grouped into only one level of clusters, but they did not have a template hierarchy. A prototype template is chosen to represent the whole cluster and the chamfer function is used to measure the shape variation from the prototype. They assume that the shape variation measured by the chamfer function follows a χ^2 distribution. A χ^2 distribution is parameterized by only two parameters, the variance and the dimensionality of the distribution. These were learned from the members of each template cluster. The dimensionality of the χ^2 distribution directly corresponds to the dimensionality of the shape-space spanned by the templates in each cluster. Hence, they are able to obtain parameterized likelihood distribution with a proper partition function for each cluster group separately. However, they fail to quantify the likelihood of their templates matching to the background.

Sidenbladh [15] and Black derive a robust likelihood function by trying to explain features both belonging to the foreground object (in their case, limbs) and the features belonging to the background simultaneously . PDFs of the filter responses were learned for foreground objects and background images during the training. The likelihood ratio penalises the data that is well explained by both foreground model and background model.

$$p(\text{all feartures} \mid \text{fgrnd model, bgrnd model}) \propto \frac{p(\text{fgrnd feartures} \mid \text{fgrnd model})}{p(\text{fgrnd feartures} \mid \text{bgrnd model})} \qquad (1)$$

Equation (1) is similar in spirit to what we derive as a formulae for robust likelihood function through PDF projection theorem. However, the PDF projection theorem allows us to interpret the likelihoods in raw image data domain, irrespective of the features we use.

Stenger *et al.* [16] uses two similarity measures, chamfer score and skin colour overlap to match hand templates. They assume independence between the two similarity measures and the likelihood is formulated as a weighted sum of chamfer score and skin coloured area within the projected model.

This paper is limited to matching a set of templates using likelihood PDFs obtained through the PDF projection theorem. Extending the PDF projection theorem to a template hierarchy is left for future work. Furthermore, we do not use skin colour information in any of our experiments as the main aim of this paper is to show improvements to chamfer matching through the proper modelling of the likelihood distributions.

3 PDF Projection Theorem

The PDF projection theorem [2] provides a mechanism to work in the raw data or image domain, $\mathbf{I}$ (rgb values of the image pixels), rather than performing likelihood comparisons in extracted feature space, $\mathbf{z}$. This is done by projecting the PDF estimates from the feature space back to the raw data space. The Neyman-Fisher [11] factorization theorem states that if $\mathbf{z} = \tau(\mathbf{I})$ is a *sufficient statistic* for H, then the function $p(\mathbf{I}|H)$ can be factored into two functions,

$$p(\mathbf{I}|H) = g(\tau(\mathbf{I})|H)\, h(\mathbf{I}) \tag{2}$$

This is a product in which one function, h, does not depend on H and the other, g, depends on H only through $\tau(\mathbf{I})$. A statistic $\mathbf{z}$ can be any function of raw data $\mathbf{I}$. However, in the context of this paper it refers to features extracted from the raw image data.

We can now derive the PDF Projection theorem by applying equation (2) to two hypotheses H and H_0. Let $\mathbf{z} = \tau(\mathbf{I})$ be any feature set computed from the raw data $\mathbf{I}$. Let H_0 be some fixed hypothesis such that the PDF $p(\mathbf{I}|H_0)$ and $p(\mathbf{z}|H_0)$ are known and greater than zero everywhere in the range of $\mathbf{I}$. If $\mathbf{z}$ is sufficient statistic for both H and H_0 and $p(\mathbf{z}|H)$ is a PDF, then the following function is a PDF.

$$p(\mathbf{I}|H) = \frac{p(\mathbf{I}|H_0)}{p(\mathbf{z}|H_0)} p(\mathbf{z}|H) \tag{3}$$

It is interesting to observe that $p(\mathbf{I}|H)$ invariant to the chosen feature $\mathbf{z}$, if $\mathbf{z}$ is a sufficient statistic. Baggenstoss [2] and Kay [9] introduced class-specific methods by extending equation (3) to include class-specific features as

$$p(\mathbf{I}|H_j) = p(\mathbf{I}|H_0) \frac{p(\mathbf{z}_j|H_j)}{p(\mathbf{z}_j|H_0)}, \tag{4}$$

where a different feature set $\mathbf{z}_j$ is extracted for each hypothesis H_j. In theory the reference hypothesis H_0 can be any hypothesis as long as $\mathbf{z}_j$ is sufficient statistic for H_j and H_0. However in practice we can improve the accuracy of the likelihood estimation by carefully choosing the reference hypothesis. The following points need to be taken into consideration.

Sufficient statistics: It is difficult to prove the sufficiency requirement in most cases. However, near optimal performance can be obtained even if the sufficiency requirement is not completely satisfied. Moreover, it is not always known if sufficiency is ever achieved with most Bayes classifiers. The point is that we achieve optimal performance if sufficiency is satisfied, otherwise the best possible near-optimal performance is obtained given the 'insufficiency' of the selected features [2].

Accurate evaluation of reference PDF: This is the pertinent problem in the application of the PDF projection theorem. The denominator density $p(\mathbf{z}_j|H_0)$ needs to be accurately estimated at the tails, where even a small error will have large effect on the likelihood estimation. Bagenstoss suggests choosing an analytical PDF as a reference hypothesis and has provided analytical solutions for a number of statistics [10]. However, it is difficult to come up with analytical PDFs in many computer vision problems. An easier solution is to formulate it in a way such that it can be estimated empirically. But much will depend on its accuracy in the tail and we need to make sure that $p(\mathbf{z}_j|H_0) > 0$ whenever $p(\mathbf{z}_j|H_j) > 0$.

In simple terms, the PDF projection theorem allows us to estimate the data likelihood as a ratio of the feature likelihood and a reference likelihood. The numerator describes how well the features support the object hypothesis and the denominator describes to what extend they resemble the reference hypothesis. The decision can be now based on how good the class-specific features are at separating the object from the clutter.

4 Template Matching

In this paper we illustrate the application of the PDF projection theorem using chamfer matching. A common approach is to have a number of prototype shape templates and search for them in the image. Chamfer score and a closely related measure Hausdorff score have been used in many shape matching schemes. We use truncated chamfer score which makes it more robust to outliers.

Truncated Chamfer score: The similarity between two shapes can be measured using their chamfer score. Given the two point sets $\mathscr{U} = \{\mathbf{u}_i\}_{i=1}^{n}$ and $\mathscr{V} = \{\mathbf{v}_j\}_{j=1}^{m}$, the truncated chamfer score is the mean of the truncated distances between each point, $u_i \in \mathscr{U}$ and its closest point in $\mathscr{V}$, where τ is the threshold distance:

$$d_{cham,\tau}(\mathscr{U},\mathscr{V}) = \frac{1}{n}\sum_{u_i\in\mathscr{U}} \min\left(\min_{v_j\in\mathscr{V}} ||u_i - v_j||, \tau\right) \tag{5}$$

The chamfer score between two shapes can be efficiently computed using a distance transform (DT) of the edge image. This transformation takes a binary edge image as input, and assigns to each pixel in the image the distance to its nearest edge pixel. The chamfer score between a template and an edge map can then be computed as the mean of the DT values at the template point coordinates. Edge orientation is included by computing the chamfer score only for edges with similar orientation, in order to make the distance function more robust [6]. We use 8 orientation groups for edges and hence eight separate distance maps are calculated per image. Similarly the template points are also divided into 8 groups.

Class-specific features: When matching a template T_j with the image, the chamfer score $\mathbf{z}_j$ is obtained by adding the closest distances from the projected template points to the nearest edge pixels of the image. The feature $\mathbf{z}_j$ is class-specific to template T_j in the sense that subset of edge pixels involved in the calculation e_j are chosen by the shape of the projected template T_j. Hence different edge features are used by different templates.

Common Reference Hypothesis: By choosing a common reference hypothesis H_0 for all the templates $j = 1,..,N$, $p(\mathbf{I}|H_0)$ becomes a constant for all templates. Thus, we obtain likelihood values over raw image data which are consistent and comparable using (4). H_0 is chosen as the union of all possible hand images and background images. Mathematically, we define it as $H_0 = H_1 \cup H_2 \cup ... \cup H_N \cup H_b$, where H_b is the hypothesis that image contains only background. As a result, the reference likelihood for j^{th} template is defined as

$$p(\mathbf{z}_j|H_0) = \sum_{i}^{N} p(\mathbf{z}_j|H_i)p(H_i) + p(\mathbf{z}_j|H_b)p(H_b). \tag{6}$$

Intuitively, the reference likelihood measures the possibility of the template being matched to background and other types of objects (i.e. objects represented by other templates) at a given chamfer value.

Figure 2: **Training images**. *First and second rows show some of the synthetic hand poses used for estimating the feature and reference likelihood distributions of the template in Figure 1(b).*

5 Experiments

We have created 35 hand templates from a 3D hand model (5 gestures at 7 different scales, see figure 4). In our case, hypothesis H_j states that the image contains a hand pose similar to template T_j (in scale and gesture). The hypothesis H_j is tested by matching the template T_j to the image. The aim of the experiment is to decide which one of the following measure is superior in terms of matching performance.

1. $\mathbf{z}_j$, the chamfer score obtained by matching the template T_j to the image.
2. $p(\mathbf{z}_j|H_j)$, the feature likelihood of template T_j.
3. $p(\mathbf{I}|H_j) = K\frac{p(\mathbf{z}_j|H_j)}{p(\mathbf{z}_j|H_0)}$, the data likelihood value from the PDF projection theorem.

Most schemes such as [1, 6] explicitly threshold the chamfer score to arrive at matching decisions. Some [13, 18] learn the feature likelihood distributions. However, to our knowledge, no one has so far attempted to model the data likelihood through the PDF projection theorem. Our experiments below illustrate the advantage of using the data likelihoods to make the matching decisions of the templates.

Learning the Likelihood PDFs: It is a formidable task to learn both $p(\mathbf{z}_j|H_j)$ and $p(\mathbf{z}_j|H_0)$ from real image data if the number of templates involved is large. Hence, we choose to learn the PDFs using synthetic images (see figure 2). Learned statistics from synthetic images have been successfully used to detect human poses in real images recently [14]. The feature likelihood distribution, $p(\mathbf{z}_j|H_j)$, of the template T_j is learned from the synthetic images of hands belonging to template region of H_j. Each template T_j is matched to 1000 synthetic images and chamfer scores from the immediate neighbourhood of the matched location are obtained. The feature likelihood distribution is approximated by histogramming these chamfer scores. Similarly the reference likelihoods $p(\mathbf{z}_j|H_0)$ are learned from 10000 synthetic images created from all possible hand hypotheses and plain background images. We learned 70 PDFs, a feature likelihood and a reference likelihood distributions for each of the 35 templates in the above manner.

Results from synthetic data: We tested template matching on 1000 randomly created synthetic images. Each synthetic image contains a hand pose similar in scale and pose (with small perturbations) to a randomly chosen template. Three Receiver Operating

Method	Within first 10	Within first 20	Within first 50
Chamfer score	63%	70%	79%
Feature Likelihood	66%	76%	85%
Data Likelihood	74%	85%	95%

Table 1: **Detection Rates**

The second, third and fourth columns show the percentages for obtaining the correct template within the first 10, 20 and 50 matches, respectively. This experiment was conducted on 150 real images.

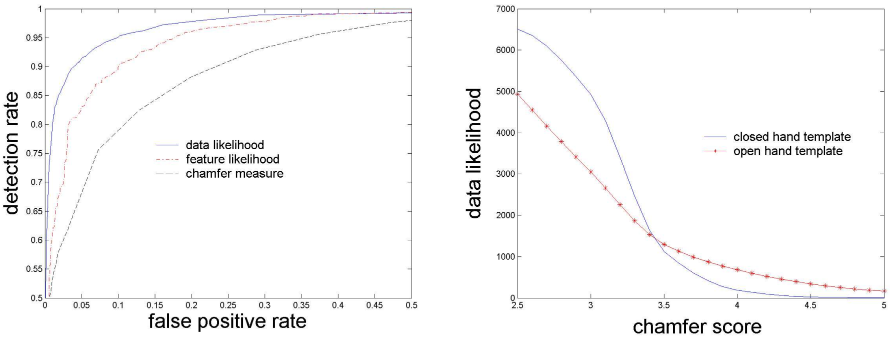
Characteristic (ROC) curves were obtained for each measure described in section 5 (see figure 3 (a)). It is clear that matching with the data likelihood measure provides the best performance followed by the performance of using the feature likelihoods. The ROC curves show that a very significant improvement is gained by using the data likelihood value over the raw chamfer score.

Results from matching to real images: Some interesting qualitative results from matching to real images are shown in figure 4. Columns 2 and and 5 show the templates corresponding to maximum data likelihood value and the minimum chamfer score respectively (for raw chamfer score the lower the better). Column 3 and 6 display the values of the different measures for those particular template matches.

Consider the example in row 3 in figure 4. The correct template, chosen by the data likelihood value has a larger chamfer score, $\mathbf{z}_j = 3.72$, and a smaller feature likelihood, $p(\mathbf{z}_j|H_j) = 13.15 \times 10^{-2}$, than the wrong template with chamfer score, $\mathbf{z}_j = 3.5$, and feature likelihood, $p(\mathbf{z}_j|H_j) = 20.15 \times 10^{-2}$. However, the reference likelihood of the correct template is small, $p(\mathbf{z}_j|H_0) = 8.5 \times 10^{-5}$, at the chamfer score of 3.72, whereas the the reference likelihood of the wrong template is quite large, $p(\mathbf{z}_j|H_0) = 108.0 \times 10^{-5}$. This is reflected in the data likelihoods for the correct and wrong templates, $p(\mathbf{I}|H_j) = 1.547 \times 10^3$ and 0.191×10^3. Hence, the reference likelihood value plays a crucial role in assigning larger likelihood values to the correct templates. Table 1 lists detection rates for finding the correct template within the first 10, 20 and 50 matches for each matching measure. We used 150 real hand images for these experiments. These matches are ranked from large number of candidates, $35 \times 240 \times 320 = 268800$, for each image. We consider a template match incorrect if it has an alignment error of more than 2 pixels. Because of these stringent criteria and the large amount of clutter present in the images, we obtained somewhat low detection rates given in the table. Detection rates can be greatly improved if other cues such as skin colour to guide the search. However, the purpose of this paper is to show the improvement by the PDF projection theorem on chamfer matching. Hence we did not use other cues to improve the detection rates.

6 Conclusion

Ideally, one would learn the likelihood distributions from natural images, which is feasible with a small number of templates. However, as the number of templates increases, the number of images needed to estimate the PDFs increases dramatically. This is a significant problem in many object recognition tasks. Currently we are estimating them from synthetic images created from projecting a 3D hand model into real background images. But we have obtained good matching results on real images using distributions learned from synthetic images. Commercially available modelling softwares such as ‘Poser’ can



(a) (b)

Figure 3: **ROC and Data Likelihood Curves** *(a) shows the* ROC *curves obtained from the performance of data likelihood, feature likelihood and raw chamfer score on 1000 synthetic images. (b) compares the data likelihood distributions of the open hand (large scale) and the closed hand (small scale) templates.*

render realistically looking images of hands and humans. In addition to providing an accurate anatomical model for hand and human body, they also model other natural image effects such as shadows accurately. They have also been successfully used in other learning methods in computer vision [14].

The training, even though time consuming, is an off-line process and does not affect the speed of the online matching at all. The advantages are numerous. The false positive rates are dramatically reduced and we obtain consistent likelihoods in raw data space. This will turn out to be very important if we need to incorporate the likelihood information with other types of information such as a temporal prior.

Acknowledgments

The authors would like to thank Tom Minka at Microsoft Research, Cambridge, UK and Bjorn Stenger at Toshiba R&D Center, Japan for many productive discussions. This work is supported by the Gates Cambridge Trust, the Overseas Research Scholarship Programme and Toshiba Corporation.

References

[1] V. Athitsos and S. Sclaroff. An appearance-based framework for 3D hand shape classification and camera viewpoint estimation. In *IEEE Conf. on FGR*, pages 45–50, 2002.

[2] P. M. Baggenstoss. Class-specific features in classification. *IEEE Transactions on Signal Processing*, pages 3428–3432, December 1999.

[3] E. Borenstein and S. Ullman. Class-specific, top-down segmentation. In *Proc. 7th European Conf. on Computer Vision*, volume I, pages 109–122, Copenhagen, Denmark, May 2002.

[4] G. Borgefors. Hierarchical chamfer matching: A parametric edge matching algorithm. *IEEE Trans. Pattern Analysis and Machine Intell.*, 10(6):849–865, November 1988.

[5] B. Caputo and H. Niemann. To each according to its need: kernel class-specific classifer. In *Proc. 16th Int. Conf. on Pattern Recognition*, volume IV, pages 94–97, Quebec City, Canada, August 2002.

[6] D. M. Gavrila. Pedestrian detection from a moving vehicle. In *Proc. 6th European Conf. on Computer Vision*, volume II, pages 37–49, Dublin, Ireland, June/July 2000.

[7] N. Jojic and B. Frey. Learning mixture models of images and inferring spatial transformations. In *Proc. Conf. Computer Vision and Pattern Recognition*, volume I, pages 416–422, June 1999.

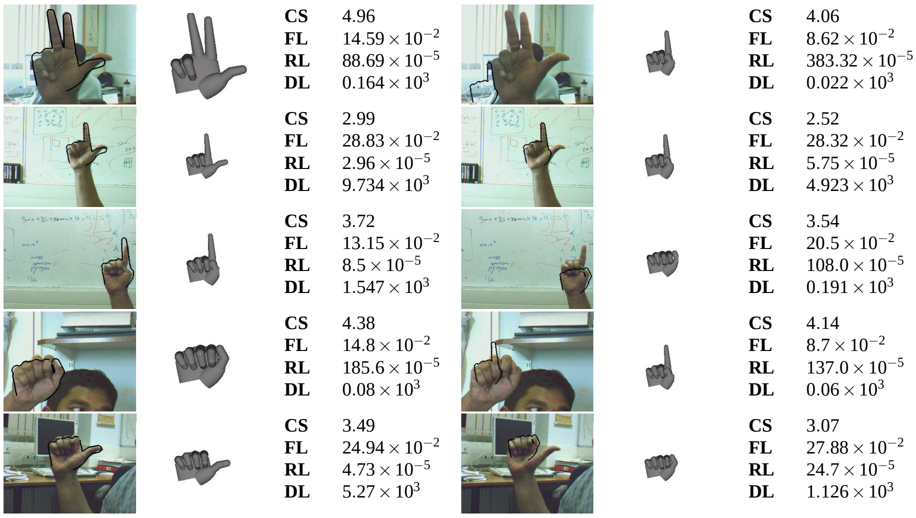


Figure 4: **Hand template matching** *Rows show some of the images where using the data likelihood in matching outperformed matching using just the chamfer score. Columns 2 and 5 show the best templates corresponding to maximum data likelihood value and the minimum chamfer score (for raw chamfer score the lower the better). Column 3 and 6 display the values of the different measures for those particular template matches.* **CS***: chamfer score,* **FL***: feature likelihood,* **RL***: reference likelihood,* **DL***: data likelihood. It can be seen from the examples that the reference likelihood plays a crucial role in maximizing the data likelihood for the correct template.*

[8] M. J. Jones and J. M. Rehg. Statistical color models with application to skin detection. *Int. Journal of Computer Vision*, 46(1):81–96, January 2002.

[9] S. Kay. Sufficiency, classification, and the class-specific feature theorem. *IEEE Trans. Information Theory*, 46(4):1654–1658, July 2000.

[10] S. Kay, A. Nuttall, and P. Baggenstoss. Multidimensional probability density approximation for detection, classification and model order selection. *IEEE Trans. on Sig. Proc.*, pages 2240–2252.

[11] M. Kendall and A. Stuart. *The Advanced Theory of Statistics*, volume 2. Charles Griffin, London, 1979.

[12] T. Minka. Exemplar-based likelihoods using the pdf projection theorem. Technical report, Microsoft Research Ltd., Cambridge, UK, March 2004.

[13] C. F. Olson and D. P. Huttenlocher. Automatic target recognition by matching oriented edge pixels. *Transactions on Image Processing*, 6(1):103–113, January 1997.

[14] G. Shakhnarovich, P. Viola, and T. Darrell. Fast pose estimation with parameter-sensitive hashing. In *Proc. 9th Int. Conf. on Computer Vision*, volume II, pages 750–757, 2003.

[15] H. Sidenbladh and Black. Learning image statistics for bayesian tracking. In *Proc. 8th Int. Conf. on Computer Vision*, volume 1, pages 709–716, Vancouver, Canada, 2001.

[16] B. Stenger, A. Thayananthan, P. H. S. Torr, and R. Cipolla. Filtering using a tree-based estimator. In *Proc. 9th Int. Conf. on Computer Vision*, volume II, pages 1063–1070, 2003.

[17] A. Thayananthan, B. Stenger, P. H. S. Torr, and R. Cipolla. Shape context and chamfer matching in cluttered scenes. In *CVPR*, volume I, pages 127–133, 2003.

[18] K. Toyama and A. Blake. Probabilistic tracking in a metric space. In *Proc. International Conference on Computer Vision*, volume II, pages 50–57, Vancouver, Canada, July 2001.

[19] J. Triesch and C. Von der Malsburg. A system for person-independent hand posture recognition against complex backgrounds. *IEEE Trans. Pattern Analysis and Machine Intell.*, 23(12):1449–1453, 2001.

[20] A. R. Webb. *Statistical Pattern Recognition*. John Wiley & Sons, 2002.