Multiple Kernel Gaussian Process Classification for Generic 3D Object Recognition

Erik Rodner, Doaa Hegazy, and Joachim Denzler

Chair for Computer Vision
Friedrich Schiller University of Jena, Germany
Email: Erik.Rodner@uni-jena.de
www.inf-cv.uni-jena.de

Abstract

We present an approach to generic object recognition with range information obtained using a Time-of-Flight camera and colour images from a visual sensor. Multiple sensor information is fused with Bayesian kernel combination using Gaussian processes (GP) and hyper-parameter optimisation. We study the suitability of approximate GP classification methods for such tasks and present and evaluate different image kernel functions for range and colour images. Experiments show that our approach significantly outperforms previous work on a challenging dataset which boosts the recognition rate from 78% *to* 88%*.*

Keywords: Gaussian Processes, Time-of-Flight Camera, Kernel Combination, Bag-of-Features

1 Introduction

In the last decade, research in machine learning and computer vision mainly concentrated on image categorisation using a single visual sensor or photos from the web [1]. Despite the great success of those methods, the importance of depth information for reliable generic object recognition is mostly ignored.

In the following work we present an approach to generic object recognition using the combined information of a visual sensor and range information obtained from a Time-of-Flight (ToF) camera [2]. A ToF camera offers real-time depth images obtained by modulating an outgoing beam with a carrier signal and measuring the phase shift of that carrier when received back at the ToF sensor. Incorporating the advantages of this new camera technology into computer vision systems is current research and is successfully applied to 3d reconstruction tasks [3] or marker-less human body motion capture [4].

We use a ToF camera together with a CCD camera to solve generic object recognition problems. To combine the information of our two sensors, we utilise Gaussian process classification [5], which allows efficient hyperparameter estimation and integration of multiple sensor information using kernel combination in a Bayesian framework. In contrast to previous work [6], which used GP regression to approximate the underlying discrete classification problem, we also study Laplace approximation (LA) which directly tackles the discrete nature of the categorisation problem. Hyperparameter estimation is done by extending multi-task techniques for GP regression [7, 6] to LA. Additionally, we show how to compute image based kernel functions using range images by applying the framework of Spatial Pyramid Matching Kernels [8] (SPMK) to different kinds of local range features. Figure 1 presents an overview of the proposed approach and the main steps involved.

The main contributions of this paper are as follows:

1. We present multiple kernel learning with Gaussian process classification for sensor data fusion (previous papers are limited to visual sensors only [6]).
2. Different noise models of GP classification are studied for image categorisation and sensor data fusion (previous papers are limited to Gaussian noise and GP regression [6]).
3. Our method yields a significant improvement in terms of recognition performance compared to the current state-of-the-art of object recognition with Time-of-Flight range images [9].

The remainder of the paper is organised as follows. Section 2 gives an overview about related work in the area of image categorisation with range images and applications of GP classification. We briefly review classification and regression with Gaussian processes in Sect. 3, followed by presenting our

978-1-4244-9631-0/10/$26.00

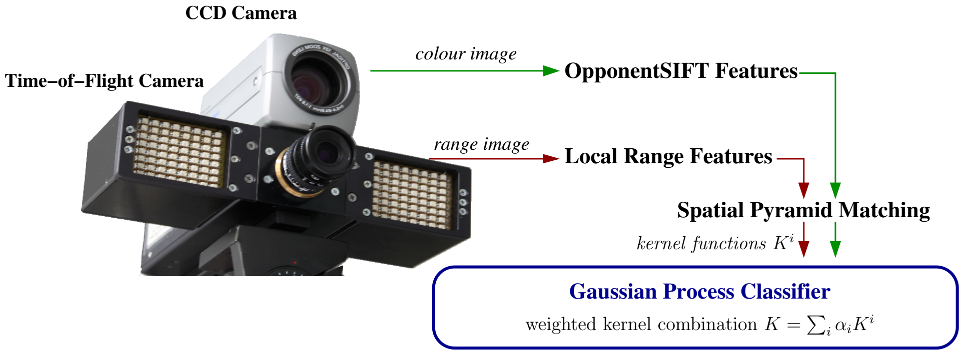

Figure 1: An overview of the proposed approach: colour and range images are captured using a CCD camera and a Time-of-Flight camera (PMD Vision Technologies 19k); Local features are extracted from the data of both sensors and "kernelized" using Spatial Pyramid Matching, yielding different kernel functions.

choice of local range features (Sect. 4) and their use for Pyramid Matching Kernels (Sect. 5) Experiments in Sect. 6 study the different parts of our approach and evaluates them with a public available dataset. A summary of our findings and a discussion of future research directions conclude the paper.

2 Related Work

One of the earlier works on local features computed on range images is the work of Hetzel et al. [10]. They present different feature types and similarity measures for histograms which can be used for nearest neighbour classification. Toldo et al. [11] propose a preliminary segmentation of complete 3d models and a description of the resulting parts utilising the Bag-of-Features idea. Classification is done with multiple equally weighted histogram intersection kernels and support vector machines. The suitability of SIFT features for image matching is studied by Zhang et al. [12], who propose to compute SIFT features on normal texture and shape index images [10]. In contrast, Lo et al. [13] present an extension to SIFT features especially suitable for range images. Special feature types for Time-of-Flight cameras are studied by Haker et al. [14]. A key idea of their work is the non-equidistant Fourier transformation to represent range images.

Our method for feature computation is mainly based on Spatial Pyramid Matching as presented by Lazebnik et al. [8] for standard image categorisation and utilised by Li et al. [15] for range images. Similar to our previous work [9, 16] we concentrate on the combination of colour or texture information obtained from a standard CCD camera and range information from a Time-of-Flight camera. A beneficial combination of these two information sources with kernel-based methods requires an efficient method for hyperparameter optimisation, which is available in Bayesian frameworks, such as Gaussian process classification [5]. As shown in Kapoor et al. [6] regression with Gaussian process priors can be successfully integrated in an image categorisation setting and can handle multiple kernels.

3 Multi-Kernel Classification with Gaussian Processes

In the following section, we briefly review Gaussian process regression and binary classification. We also explain multi-class classification with GP priors and efficient kernel combination by hyperparameter estimation.

3.1 Bayesian Estimation with Gaussian Process Priors

We first give some initial motivation before describing the use of GP priors in more mathematical terms: Machine learning can often be described as estimating the relation between inputs $\boldsymbol{x}$ and outputs or labels y. This relation is often described using a regression function f and a noise term ϵ, e.g. $y = f(\boldsymbol{x}) + \epsilon$. The idea of Bayesian estimation is now that instead of searching for a specific function f, one can regard f as a latent random variable which can be marginalised out. Gaussian processes allow to nicely model the prior distribution of regression functions f by defining a suitable covariance function controlling the smoothness of the function samples.

Let $\mathcal{X}$ be the space of all possible input data (feature vectors or images). Given n training examples $\boldsymbol{x}_i \in \mathcal{X}_T \subset \mathcal{X}$ and corresponding binary labels $y_i \in \{-1, 1\}$ (multi-class classification is explained in subsequent sections), we would like to predict the label y_* of an unseen example $\boldsymbol{x}_* \in \mathcal{X}$. The two main assumptions of Gaussian processes for regression or classification are:

1. There is an underlying latent function $f : \mathcal{X} \to \mathbb{R}$, so that labels y_i are conditionally independent from the input $\boldsymbol{x}_i$ given $f(\boldsymbol{x}_i)$. Thus, the labels are distributed according to the so called noise model $p(y_i \mid f(\boldsymbol{x}_i))$.

2. The function f is a sample of a Gaussian process (GP) prior and represents itself a random variable: $f|\mathcal{X} \sim \mathcal{GP}(\mathbf{0}, \mathcal{K}(\mathcal{X}, \mathcal{X}))$ with zero mean and covariance or kernel function $\mathcal{K} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$.

The Gaussian process prior enables to model the covariance of outputs $f(\boldsymbol{x}_*)$ as a function of inputs $\boldsymbol{x}_*$. With $\mathcal{K}$ being a kernel function describing the similarity of two inputs, the common smoothness assumption that similar inputs should lead to similar function values (and similar labels) can be modelled. Incorporating our assumptions into a Bayesian framework and marginalising over the latent function values, we have the following equations for the inference of the label y_* of an unseen example $\boldsymbol{x}_*$:

$$p(y_*|\boldsymbol{x}_*, \boldsymbol{y}, \mathcal{X}_T) = \int_{\mathbb{R}} p(y_*|f_*)\; p(f_*|\boldsymbol{x}_*, \boldsymbol{y}, \mathcal{X}_T) df_* \tag{1}$$

where we marginalise the latent function value $f_* = f(\boldsymbol{x}_*)$ corresponding to $\boldsymbol{x}$. The conditional distribution of f_* is also available with marginalisation of all latent function values $\boldsymbol{f} = (f(\boldsymbol{x}_i))_{i=1}^n$ of the training set $\mathcal{X}_T$:

$$p(f_*|\boldsymbol{x}_*, \boldsymbol{y}, \mathcal{X}_T) = \int_{\mathbb{R}^n} p(f_*|\boldsymbol{x}_*, \boldsymbol{f}) p(\boldsymbol{f}|\boldsymbol{y}, \mathcal{X}_T) d\boldsymbol{f} \tag{2}$$

Finally, the incorporation of the noise model and our assumption of independent training examples leads to:

$$p(\boldsymbol{f} \mid \boldsymbol{y}, \mathcal{X}_T) = \frac{p(\boldsymbol{f} \mid \mathcal{X}_T)}{p(\boldsymbol{y} \mid \mathcal{X}_T)} \left(\prod_i p(y_i \mid f_i) \right) \tag{3}$$

The distribution $p(\boldsymbol{f} \mid \mathcal{X}_T)$ is a n-dimensional normal distribution with zero mean and covariance matrix $\boldsymbol{K} = (\mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_j))_{i,j}$, which is often called kernel matrix. The noise model can take various forms depending on the nature of the labels y_i, which leads to the distinction between Gaussian process regression and classification.

3.2 Regression with Gaussian Process Priors

Gaussian process regression uses a Gaussian noise model

$$p(y_i \mid f_i) = \mathcal{N}(y_i \mid f_i, \sigma^2) \tag{4}$$

with variance σ^2. The advantage of this model is that we do not have to use approximated inference methods to handle the involved marginalisation in equation (1) and (2). A given binary classification problem is solved as a regression problem which regards y_i as real-valued function values instead of discrete labels. The GP regression model assumptions lead to analytical solutions of the integrals and allow to directly predict the label y_*. Let $\boldsymbol{k}_*$ be the kernel values $(\boldsymbol{k}_*)_i = \mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_*)$ corresponding to a test example $\boldsymbol{x}_*$. The GP model for regression leads to the following equation for the prediction $\bar{y}_*$ [5]:

$$\bar{y}_*(\boldsymbol{x}_*) = \boldsymbol{k}_*^T (\boldsymbol{K} + \sigma^2 \boldsymbol{I})^{-1} \boldsymbol{y} \tag{5}$$

with $\boldsymbol{I}$ denoting the identity matrix.

3.3 Classification with GP Priors and Laplace Approximation

To incorporate knowledge about the discrete nature of the labels, we can use noise models such as the cumulative Gaussian

$$p(y_i \mid f_i) = \frac{1}{2} \left(\mathrm{erf} \left(\ell \; \frac{y_i f_i}{2} \right) + 1 \right) \tag{6}$$

with length-scale parameter ℓ. Exact inference with this model is intractable and approximation methods are required. One common efficient approximation method is Laplace approximation [5]. The key idea to handle the integral in equation (2) is to approximate $p(\boldsymbol{f}|\mathcal{X}_T, \boldsymbol{y})$ with a Gaussian distribution. This can be done by finding the mode $\hat{\boldsymbol{f}}$ with nonlinear optimisation techniques and approximating the covariance matrix by utilising the Hessian at $\hat{\boldsymbol{f}}$. An important advantage of the cumulative Gaussian noise model is the availability of analytical solutions for the marginalisation involved in equation (1). For implementation details we refer the reader to the excellent text book of Rasmussen and Williams [5].

3.4 Kernel Combination for Classification with Multiple Sensors

Let us assume m kernel functions $\mathcal{K}^i(\cdot, \cdot)$ are given. In the multi-sensor setting, these kernel functions can be computed from different data modalities. To use the information of all kernels, they can be combined by a single kernel function $\mathcal{K}$, e.g. in a linear fashion with exponential weights $\boldsymbol{\alpha} \in \mathbb{R}^m$:

$$\mathcal{K}^{\boldsymbol{\alpha}}(\boldsymbol{x}, \boldsymbol{x}') = \sum_{i=1}^{m} \exp(\alpha_i)\; \mathcal{K}^i(\boldsymbol{x}, \boldsymbol{x}') \tag{7}$$

For binary classification this relates to an equivalent representation of the latent function $\boldsymbol{f}$ as a linear sum $\boldsymbol{f} = \sum_{i=1}^m \exp(\alpha_i)\, \boldsymbol{f}^i$ of independent latent functions $\boldsymbol{f}^i$ sampled from GP priors $\mathcal{GP}(\mathbf{0}, \mathcal{K}^i)$. Another combination technique we experimented with, is to plug the linear combination into an exponential kernel [17] which leads to a similar classification performance. The parameters α_i are hyperparameters of the kernel function and can be optimised using GP model selection techniques, such as likelihood optimisation. The negative log-likelihood of all training examples for the binary GP regression model is given by [5]:

$$-\log p(\boldsymbol{y} \mid \mathcal{X}_T, \boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{y}^T \left(\boldsymbol{K}^{\boldsymbol{\alpha}} + \sigma^2 \boldsymbol{I}\right)^{-1} \boldsymbol{y} + \frac{1}{2}\log\left|\boldsymbol{K}^{\boldsymbol{\alpha}} + \sigma^2 \boldsymbol{I}\right| + \frac{n}{2}\log 2\pi \quad (8)$$

with $\boldsymbol{K}^{\boldsymbol{\alpha}}$ being the parameterised kernel matrix according to equation (7). Further details regarding the likelihood of the Laplace Approximation and its gradients can be found in [5]. Maximisation of the likelihood can be done with iterative nonlinear optimisation methods. To reduce over-fitting problems, [18] propose to choose the parameter according to the minimum leave-one-out error among the parameters calculated in each iteration of the likelihood optimisation. However, we did not observed a performance gain using this technique.

3.5 Multi-Class Classification and Model Selection

So far, we presented GP classification with binary labels. However, our image categorisation application requires to solve a multi-class classification problem. Rasmussen and Williams [5] present an extension of the Laplace Approximation for multi-class classification tasks with $y \in \{1, \ldots, M\} = \Omega$. This method requires MCMC techniques and is thus computationally demanding.

We follow Kapoor et al. [6] by utilising the one-vs-all technique. For each class $i \in \Omega$ a binary classifier is trained which uses all images of i as positive examples and remaining images as a negative training set. Classification is done by returning the class with the highest probability estimated by the corresponding binary classifier. The one-vs-all approach also offers to perform efficient model selection by joint optimisation of hyperparameters for all involved binary problems [7]. The objective function is simply the sum of all binary negative log-likelihoods as given in equation (8) for GP regression. An equivalent idea can be applied to Laplace approximation.

4 Local Features for ToF Range Images and Colour Images

In the following we present local features which can be calculated using range images from a Time-of-Flight camera. The extraction process is performed in two steps: (1) preprocessing and sampling of regions and (2) the calculation of local feature descriptors. Figure 2 gives an overview of the proposed processing pipeline and guides through the following two sections.

4.1 Preprocessing and Sampling

The range data obtained with the ToF camera (in our case a Photonic Mixer Device) suffer from severe statistical noise. In order to filter this noise and smooth the range data, a median filter of size 5×5 is applied. One idea to select appropriate regions for local features is to use an interest detector [16]. However, due to the low resolution of the intensity image provided by the ToF camera this often leads to uninformative interest points and a small number of local features. Therefore, we compute local features on a predefined grid as suggested by [9].

Dense sampling has shown to be beneficial also for image categorisation tasks using colour images [17]. Thus, the same sampling strategy is applied to colour images of the visual sensor and we extract local descriptors at a predefined grid with a horizontal and vertical pixel spacing of 10 pixels.

4.2 Local Range Features

Range images have the advantage of providing direct information about the shape of objects. Therefore, it is beneficial to give preference to features that capture different shape aspects. Local shape descriptors are preferable as they provide some robustness to clutter and occlusion. Hetzel et al. [10] introduce and use three shape-specific local feature histograms (pixel depth, surface normals and curvature) for the task of free-form specific 3d object recognition. The main advantages of these features are the robustness to viewpoint changes and the discriminative information they contain for specific 3d object recognition [10] and generic object recognition [16].

Pixel Depth A local histogram of depth values is the simplest available feature. All values are naturally invariant with respect to image plane translations and rotations and provide a rough representation of the object shape. As pointed out by [10], depth histograms can be misleading in scenarios with a large amount of background clutter and the presence of multiple objects. In this paper, a histogram of 64 bins of pixel distances is calculated and used.

Surface Normals A histogram of surface normals represented as a pair of two angles (ϕ, θ) in sphere coordinates provides a first-order statistic of the local object shape. Surface normals can be easily derived from approximating the gradient

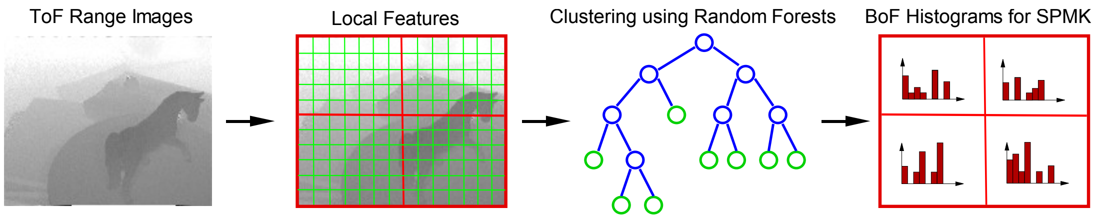

Figure 2: Local features for range images are computed on a grid without using interest point detectors. All local features are clustered with a Random Forest and used to compute Bag-of-Features (BoF) histograms for each cell. Only one level of the pyramid is shown in this figure.

with finite differences. We use a two dimensional histogram with 8×8 bins.

Curvature Representing the curvature can be done by using the shape index introduced in [19] and utilized in [10]:

$$S(\boldsymbol{p}) = \frac{1}{2} - \frac{1}{\pi} \arctan\left(\frac{k_{\max}(\boldsymbol{p}) + k_{\min}(\boldsymbol{p})}{k_{\max}(\boldsymbol{p}) - k_{\min}(\boldsymbol{p})}\right) \quad (9)$$

where $k_{\max}(\boldsymbol{p})$ and $k_{\min}(\boldsymbol{p})$ denote the principle curvatures computed at point $\boldsymbol{p}$. The principal curvatures are proportional to the eigenvalues of the structure matrix [20]. Therefore, the second derivatives have to be computed in order to calculate the shape index. The histogram calculated of all S values consists of 64 bins.

4.3 Local Colour Features

Many papers have been published about comparing different local descriptors of colour images for image categorisation tasks. The most recent work is the comparison of local colour feature descriptors by Van de Sande et al. [17]. We use Opponent-SIFT[1] features which was among the best discriminative descriptors in the evaluation of [17].

5 Image-Based Kernel Functions

In the following section we show how to compute image-based kernel functions using the local features types presented in the last section.

One of the state-of-the-art feature extraction approaches for image categorisation is the Bag-of-Features (BoF) idea. A quantisation of local features which is often called a codebook, is computed at the time of training. We use the supervised method of Moosmann et al. [21] for clustering which provides discriminative codebook elements. For each image a histogram is calculated which counts the number of matching local features for each codebook entry. On the one hand, the intuition behind this idea is that the clustering corresponds to an automatic decomposition of objects into corresponding parts. Thus, the histogram counts the occurrence of several latent object parts in an image, which results in a highly discriminative descriptor. On the other hand, current local features are often unable to describe highly complex object parts. In these cases the global descriptor contains sophisticated statistics about the image content.

A standard way to apply the BoF idea to kernel-based classifiers is to use the calculated histogram as a feature vector and apply a traditional kernel function such as the radial basis function kernel. In contrast, we define the kernel function directly on images. The Spatial Pyramid Matching Kernel as proposed by Lazebnik et al. [8] extends the BoF idea and divides the image recursively into cells (e.g. 2×2 cells). In each cell the BoF histogram is calculated and the kernel value is computed using a weighted combination of histogram intersection kernels corresponding to each cell. Each of the four local feature types presented in Sect. 4 yields a different kernel, which can be combined with GP model selection as explained in Sect. 3.4.

6 Experiments

In the following we empirically support the following hypotheses:

1. Our method significantly outperforms our previous approach [9] (Sect. 6.3).
2. GP regression outperforms Laplace approximation (Sect. 6.3).
3. Generic object recognition benefits from range information (Sect. 6.2).
4. Combining only range image kernels by GP likelihood optimisation leads to better results than fixed equal kernel weights. However, weight optimisation is not beneficial when all kernels and few training examples are used (Sect. 6.2).
5. Local range features computed using surface normals lead to the best recognition performance (Sect. 6.1).

[1] We used the software provided by Koen van de Sande at http://www.colordescriptors.com.

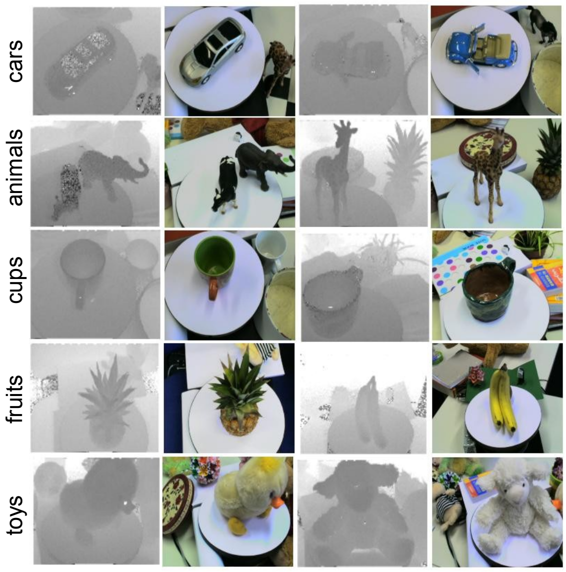

Figure 3: Some examples of image pairs included in the dataset of [16] with images obtained from a ToF camera and a visual sensor. From top to bottom: *cars*, *animals*, *cups*, *fruits* and *toys*. The dataset includes high intraclass variabilities (*animals*, *fruits*, *toys*) and small interclass distance (*animals* and *toys*).

We used the object category dataset presented in [16] in all evaluation experiments. Note that this database is the only available dataset for generic object recognition that provides both Time-of-Flight camera images and images from a CCD camera.

The database of [16] consists of a large set of 2d/3d images obtained from a CCD camera and a Time-of-Flight camera [2], which is a PMD Vision 19k with a resolution of 160×120 pixels. Examples can be found in Figure 3. The images belong to five different generic object categories (*cars*, *toys*, *cups*, *fruits* and *animals*). Each category consists of seven object instances with 32 image pairs. Images can contain multiple instances of the same class, large viewpoint and orientation variations, partial occlusion (e.g. by other objects), truncation (e.g. by the image boundaries) as well as background clutter.

In contrast to our previous work [16], which use a predefined split into a training set with 100 images and a test set of 60 images for each category, we evaluate our approach using a varying number g of object instances for training ($32g$ image pairs) and the remaining images of the dataset for testing. As a performance measure we use the mean of the average recognition rate obtained from 50 evaluations with a random selection of training instances.

6.1 Evaluation of Range Feature Types

First of all, we compare the performance for all local range features presented in Sect. 4.2. Classification is done with GP regression and without additional hyperparameters. The results are shown in Figure 4 (a) for different numbers of object instances used for training. The performance ranking of the respective methods is clearly visible and local features calculated using surface normals result in the best average recognition rate.

6.2 Evaluation of GP Classification and Kernel Combination

Let us have a look on the performance of GP regression compared to approximate GP classification with Laplace approximation (LA). Figure 4 (c) shows the performance of both methods using the image kernel function of the CCD camera. GP regression significantly outperforms LA, which is a surprising result because of the theoretical suitability of LA for classification problems. We also tested LA with hyperparameter optimisation of the length-scale ℓ included in the noise model and defined by equation (6). This additional hyperparameter optimisation leads to a small performance gain, but is still inferior to GP regression.

Figure 4 (b) shows that by combining multiple range kernels, the categorisation performance increases compared to single range kernel functions. The best method is GP regression with weights optimised by likelihood optimisation as presented in Sect. 3.4. As also observed in the previous experiment, LA does not lead to a performance gain, even with hyperparameter optimisation.

The results for combined image kernels of both sensors are shown in Figure 4 (d). Combining range data from the ToF sensor and images of the visual sensor leads to a superior categorisation performance (76.3% using 1 instance) compared to the best results using a single sensor (71% with 1 instance). The performance gain due to sensor combination is most prevalent for few training examples or object instances. An interesting fact is that in this case we do not benefit from weight optimisation for kernel combination. This is likely due to over-fitting in the presence of the highly discriminative colour kernel. Therefore, one should prefer to use equal weights in those scenarios.

6.3 Comparison with Previous Work

We also compared our method with our previous approach [16] and its extension using dense sampling [9]. Note that these works are the only ones providing methods for ToF-based categorisation. In this experiment the same number of training examples is used to allow direct comparison of the recognition rates. The results are shown in Table 1. Our approach utilising GP regression and hyperparameter optimisation significantly outperforms

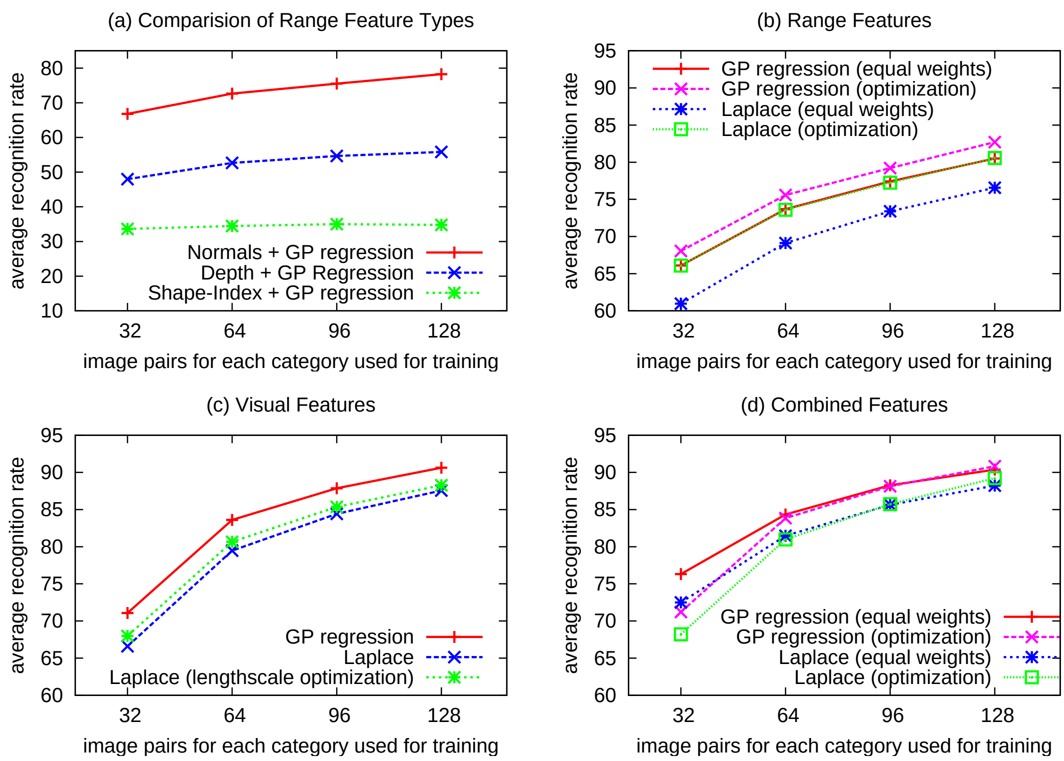

Figure 4: Evaluation of (a) different types of range features; (b) GP methods with multiple range features; (c) GP methods with colour features; (d) and combined features from the CCD camera and the ToF camera. 32 image pairs corresponds to one object instance or type.

Table 1: Comparison of our GP based approach to previous work with an equal number of training examples. (s) denotes range features computed on interest points only.

Ref.	Method	Features	Recog. Rate
[16]	Boosting	range feat. (s)	39.8 %
[9]	Boosting	range features	62.8 %
Ours	GP Reg.	range features	**79.2** %
[16]	Boosting	comb. feat. (s)	64.2 %
[9]	Boosting	combined feat.	78.4 %
Ours	GP Reg.	combined feat.	**88.1** %

their approach for single sensor data from the ToF camera and combined information of both sensors. Even by using range features only we achieve 79.2% average recognition rate which is superior to the overall classification system of [9] with a recognition rate of 78.4%. The confusion matrix in Figure 5 highlights the still existing classification difficulties.

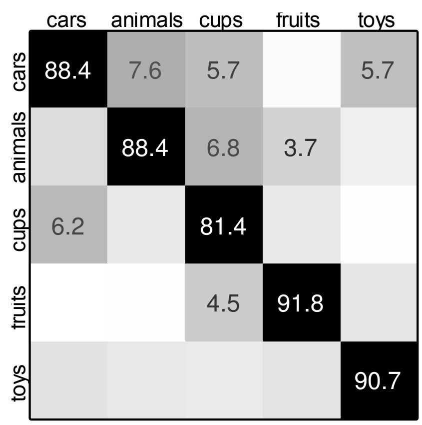

Figure 5: Results of our GP approach represented as a confusion matrix. Only values above 3% are displayed and highlight difficult cases.

7 Conclusions and Further Work

We present an approach to generic object recognition using combined information obtained from a CCD camera and a Time-of-Flight camera. Our method is based on Gaussian process (GP) classification and kernel functions computed using different types of local features. This framework allows us to study various aspects of image categorisation with GP and lead to interesting results such as the superiority of GP regression compared to Laplace approximation. We also observe that Bayesian kernel combination does not always lead to better results compared to equally weighted kernels especially for few training examples and with a single highly discriminative kernel.

An interesting direction for future research would be to incorporate the variable sensor uncertainty of time-of-flight sensors directly into the problem of kernel combination. This uncertainty is directly available for each pixel of those cameras and might give hints about whether to trust the range-based kernel in situations where range estimation is not reliable (such as black surfaces).

8 Acknowledgments

The authors would like to thank Michael Kemmler for valuable discussions about Gaussian Processes and his comments on the paper.

References

[1] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," IJCV, vol. 73, no. 2, pp. 213–238, 2007.

[2] R. Lange, "3d time-of-flight distance measurement with custom solid-state image sensors in cmos/ccd-technology," Ph.D. dissertation, University of Siegen, 2000.

[3] Y. Cui, S. Schuon, C. Derek, S. Thrun, and C. Theobalt, "3d shape scanning with a time-of-flight camera," in CVPR, 2010.

[4] V. Ganapathi, C. Plagemann, D. Koller, and S. Thrun, "Real time motion capture using a single time-of-flight camera," in CVPR, 2010.

[5] C. E. Rasmussen and C. K. I. Williams, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning). The MIT Press, 2005.

[6] A. Kapoor, K. Grauman, R. Urtasun, and T. Darrell, "Gaussian processes for object categorization," IJCV, vol. 88, no. 2, pp. 169–188, 2010.

[7] N. D. Lawrence, J. C. Platt, and M. I. Jordan, "Extensions of the informative vector machine," in Deterministic and Statistical Methods in Machine Learning, 2004, pp. 56–87.

[8] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in CVPR, 2006, pp. 2169–2178.

[9] D. Hegazy, "Boosting for generic 2d/3d object recognition," Ph.D. dissertation, Friedrich Schiller University of Jena, 2009.

[10] G. Hetzel, B. Leibe, P. Levi, and B. Schiele, "3d object recognition from range images using local feature histograms," in CVPR, vol. 2, 2001, pp. 394–399.

[11] R. Toldo, U. Castellani, and A. Fusiello, "A bag of words approach for 3d object categorization," in MIRAGE: 4th Intl Conference on Computer Vision/Computer Graphics Collaboration Techniques, 2009, pp. 116–127.

[12] G. Zhang and Y. Wang, "Faceprint: Fusion of local features for 3d face recognition," in ICB: Intl Conf on Advances in Biometrics, 2009, pp. 394–403.

[13] T.-W. R. Lo and J. P. Siebert, "Local feature extraction and matching on range images: 2.5d sift," Computer Vision Image Understanding, vol. 113, no. 12, pp. 1235–1250, 2009.

[14] M. Haker, M. Böhme, T. Martinetz, and E. Barth, "Scale-invariant range features for time-of-flight camera applications," in CVPR Workshop on Time-of-Flight-based Computer Vision (TOF-CV), 2008, pp. 1–6.

[15] X. Li and I. Guskov, "3d object recognition from range images using pyramid matching," in ICCV, 2007, pp. 1–6.

[16] D. Hegazy and J. Denzler, "Combining appearance and range based information for multi-class generic object recognition," in CIARP, 2009, pp. 741–748.

[17] K. van de Sande, T. Gevers, and C. Snoek, "Evaluating color descriptors for object and scene recognition," PAMI, vol. 32, no. 9, pp. 1582–1596, 2010.

[18] Y. A. Qi, T. P. Minka, R. W. Picard, and Z. Ghahramani, "Predictive automatic relevance determination by expectation propagation," in ICML, 2004, pp. 85–92.

[19] J. J. Koenderink and A. J. van Doorn, "Surface shape and curvature scales," Image and Vision Computing, vol. 10, no. 8, pp. 557–564, October 1992.

[20] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," IJCV, vol. 60, no. 2, pp. 91–110, 2004.

[21] F. Moosmann, B. Triggs, and F. Jurie, "Fast discriminative visual codebooks using randomized clustering forests," in NIPS, 2006, pp. 985–992.