Robust Pose Estimation using the SwissRanger SR-3000 Camera

Sigurjón Árni Guðmundsson, Rasmus Larsen and Bjarne K. Ersbøll

Technical University of Denmark, Informatics and Mathematical Modelling.
Building 321, Richard Petersens Plads, DTU DK-2800 Kgs. Lyngby,
sag@imm.dtu.dk,
WWW home page: http://www2.imm.dtu.dk/~sag

Abstract. In this paper a robust method is presented to classify and estimate an objects pose from a real time range image and a low dimensional model. The model is made from a range image training set which is reduced dimensionally by a nonlinear manifold learning method named Local Linear Embedding (LLE). New range images are then projected to this model giving the low dimensional coordinates of the object pose in an efficient manner. The range images are acquired by a state of the art SwissRanger SR-3000 camera making the projection process work in real-time.

1 Introduction

The pose says how the object is oriented in space. Detecting an objects pose is an active research field in computer vision and closely related to other very active topics such as object recognition, classification and face recognition.

In this work a statistical machine learning method of dimensionality reduction of data is proposed to detect the pose of an object; on which side it lies and the planar angle from the cameras view. A low dimensional model is trained using images of an object in various positions. The model then gives the low dimensional pose coordinates of the data points. Finally a new image can be projected onto the model resulting in the objects pose coordinates. Analyzing intensity image data with manifold methods have shown good results before [1–3]. In the present work a model is constructed from range images from a SwissRanger camera that produces range images in real time

1.1 Related Research

The pose estimation problem has been approached in numerous ways. Many industrial methods are based on matching objects to a CAD model [9]. Other methods such as extended gaussian images [8] recognize the pose by analyzing the surface normal behavior of the object. Machine learning has been a "hot" subject in academic research in this field, especially dimensionality reduction techniques. Eigen shapes and optimal component projections have shown good

results on a wide range of data [10]. In recent years manifold learning methods such as isomap, Laplacian eigenmaps and locally linear embedding (LLE) have shown how various data lie on nonlinear manifolds that can be uncovered [1, 3, 4]. These methods strive at the same goal but have different different properties with regard to classification and projecting new data etc. In this work the LLE will be investigated and shown how its properties suit the pose problem.

1.2 Contributions

The potential of this nonlinear pose estimator based on range data is considerable. LLE interprets the change between the training images in the simplest fashion while at the same time giving a possibility to project a new data point onto the model without needing to go through the heavy eigen - calculations.

The effect of using range data from the SwissRanger camera also gives the model added robustness; eliminating the effect of variable lighting and giving a possibility of dimensionally scaling new images before projecting onto the model.

1.3 Outline of Paper

This paper proposes a pose estimation technique based on the locally linear embedding (LLE). It is composed of four main sections. In section 2 the the theoretical background of LLE is given and the SwissRanger is introduced. In section 3 data, the experiments and their results are discussed. Finally section 4 shows the direction for further research.

2 Technical Approach

The approach presented is based on the following:

- Images of an object in different poses lie on a non-linear manifold in the high dimensional pixel space.
- LLE finds the intrinsic dimension of the manifold.
- New data points are projected onto the embedded dimensions using Locally Linear Projecting (LLP)
- Range data adds robustness to the model and simplifies preprocessing of both training data and the new data.

2.1 Reduction of Dimensionality

One of the hottest topics in statistical machine learning is dealing with high dimensional data. A sequence of 100×100 pixel images can be seen as points in a 10000 dimensional pixel space. If these images are in some way correlated it is likely that a much lower dimensional feature space can be found; where the features in some way describe the sequence.

The classical approach to dimensionality reduction is Principal Component Analysis (PCA). PCA linearly maps the dataset to the maximum variance subspace by eigen-analyses of the covariance matrix, where the eigenvectors are the principal axes of the subspace, the eigenvalues give the projected variance of the data and the number of the significant eigenvalues gives the "true" dimensionality. PCA is an optimal linear approach and through its usage nonlinear structures in the data can be lost. It has been shown ([4, 1, 3]) that many very interesting data types lie on nonlinear low dimensional manifolds in the high dimensional space and several so called nonlinear manifold learning methods have been developed to describe the data in terms of these manifolds. One example of such methods is Locally Linear Embedding (LLE) which is the subject of the next section.

2.2 Local Linear Embedding

LLE assumes that locally each data-points' close neighborhood is linear and optimally preserves the geometry in these neighborhoods. It is an elegant unsupervised, non-iterative method that avoids the local minima problems that plague many other methods. For a dataset $\mathbf{X}$ with N points in D dimensions ($D \times N$), an output $\mathbf{Y}$ of N, d dimensional points is found ($d \times N$) where $d \ll D$.

The algorithm has three basic steps. In the first step each points K-nearest neighbors are found, second the points are approximated by a linear weighted combination of its neighbors. Finally a linear mapping is found through an eigenvalue problem to reduce the dimensionality of the approximated points to the embedded dimensions d. A full description and proofs of the algorithm is given in [2].

LLE properties: LLE is a "natural" classifier. The separate classes' samples are mapped to separate dimensions as they are seen as lying on separate manifolds in the data. With C classes all samples of a certain class are mapped onto a single point in $C - 1$ dimensions. The choice of dimensions for the third step in LLE is then twofold; choice of the local intrinsic dimensionality d_L which is the dimension of the manifold each class lies on, and the choice of global intrinsic dimensionality which is then used in step 3:

$$d_G = C \cdot d_L + (C - 1) \tag{1}$$

Finding d_L can be a complicated matter as the residual variances of each component cannot be measured such as is done in PCA and Isomap. On the other hand the intrinsic dimension can be found by using Isomap or by experiment in LLE to find what d_L describes the manifold. It has been shown in research ([1, 3]) that the dimensionality of a dataset made of images of a 3D object rotated 360°, then 2D are needed to describe the change between the points, i.e. the rotation is described by a circle.

LLE is not easily extended to out-of-sample data and new images x_n cannot simply be mapped to the low dimensional feature space. Calculating new

LLE coordinates for each new image with a large training set is way too heavy computationally and therefore out of the question for most fast or real time applications.

A simple method to map new data is sometimes called Local Linear Projecting (LLP, [6, 7, 2]). It utilizes the first two steps in LLE and omits the expensive third step. First it finds K neighbors' of x_n in the training set $\mathbf{X}$, then the weights are calculated as in *step 2* and these are used to make a weighted combination of the neighbors embeddings in $\mathbf{Y}$. The method thus exploits the local geometrical preservation quality of LLE. On the other hand the method is sensitive to translated input but this effect can be eliminated by preprocessing. This has been proved as a quick and effective method with good results.

2.3 The SwissRanger SR-3000 Camera and Data

The SwissRanger [11] camera is designed on the criteria to be a cost-efficient and eye-safe range finder solution.

Fig. 1. The CSEM SwissRanger SR-3000 Camera

Basically it has an amplitude modulated light source and a two dimensional sensor built in a miniaturized package (see fig. 1). The light source is an array of 55 near-infrared diodes (wavelength 850nm) that are modulated by a sinusoidal at 20MHz. This light is invisible to the naked eye.

The sensor is a 176×144 custom designed $0.8\mu m$ CMOS/CCD chip where each pixel in the sensor demodulates the reflected light by a lock-in pixel method, taking four measurement samples 90° apart for every period. From these samples the returning signal can be reconstructed. From the reconstructed signal two images are generated: An intensity (gray scale) image and a range image derived from the amplitude and the phase offset of the signal in each pixel.

The quality of the range-images is not as high as in many other range imaging devices. The spatial resolution is low and due to low emitted power the accuracy

is dependent on the reflection properties of the subject and in some situations the results can be distorted by high noise and effects such as multiple-paths. This makes these images difficult to use in some conventional 3D pose recognition methods where higher accuracy is needed. On the other hand the low latency and high frame rate giving near real-time images with near field accuracy within 2 mm makes the camera excellently suited in many scenarios. For the purposes of this experiment this camera was an excellent choice.

3 Testing Methodology

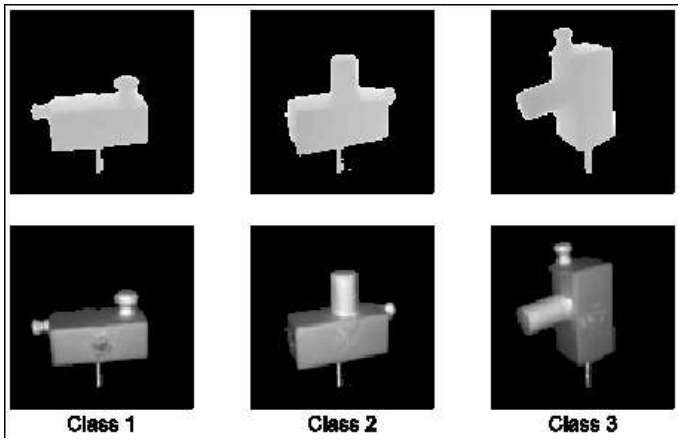

Fig. 2. The box dataset. The upper row is from the range dataset and the lower is from the intensity dataset. One image from each class is shown.

The experiment data was made of images of a cardboard box with wooden knobs attached to it (figure 2). This object can lie on 3 sides and is without symmetrical features. The images were acquired of the box while it was rotated 360° in each lying pose with 2° intervals (540 images in each dataset). The dataset was purposely made to make a model that could detect on which side a new image of the object was lying and detecting its orientation at the same time. Two models were constructed; one using the range data the other using intensity data for comparison purposes. The depth data was aligned so that the a smoothed version of the data's closest point was centralized.

The LLP projection of new data points was tested by choosing a subset of the images and using them as a test set on the model made from the remaining images. As the test points positions are known relative to the training points it can be measured if the points are mapped between the correct points. The test points where aligned according to the depth in the same manor as the training

points. In this way the depth is used to minimize the input translation problem of LLP.

3.1 Results

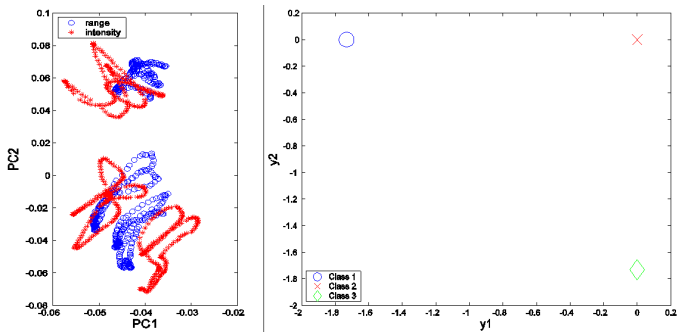

Fig. 3. A:) Two first principal components of the box data. B:)The the "Class" dimensions of the LLE mapping of the data.

The two first principal components of both datasets in figure 3A shows that the classes can be separated but the method completely fails to capture the regularity of the objects change in orientation. Still these two components contain 95% of the variance of the points.

The plot in figure 3B shows that the LLE embedding in 2D gives a perfect separation by mapping all the 180 data points of each class to three points. The three 2D local dimensional embeddings are shown in figure 4.

Figure 4 shows the local dimensions of each class using the range data on one hand and intensity data on the other. Both models model the 360° rotation of the object with a circle and in the case of range data the circles are almost perfectly smooth while the intensity model's curves are less smooth.

Experiments on LLP by projecting small subsets of the image data gave 100% correct pose estimations. Figure 5 shows the results when only half of the data is used to find the embedding dimensions and the other half is projected one by one into these dimensions. This sparser model results in less smooth curves than before but the test points are still projected to correct positions in all cases. The intensity model in figure 6 shows poorer performance with not as smooth curves and in the case of class 3 it has problems with connecting the starting and final points of the rotation, resulting in almost erratic projections i.e. points have a third close neighbor on the embedding.

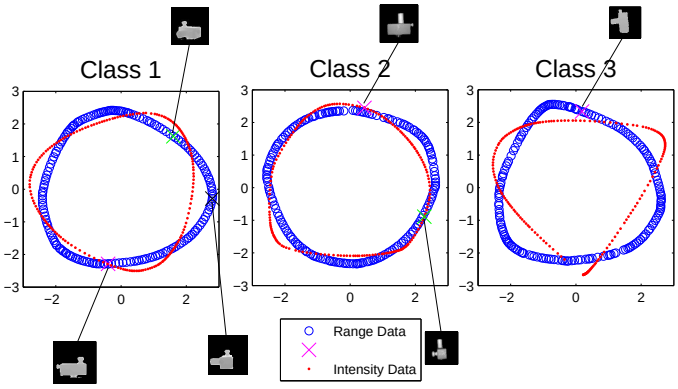

Fig. 4. The local intrinsic dimensions for each class. Six examples of image to intrinsic dimension mappings are shown.

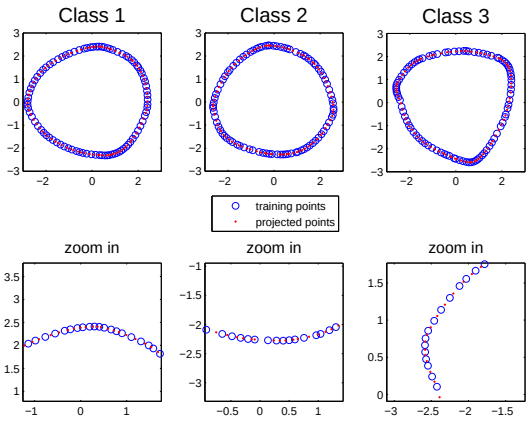

Fig. 5. *The Range Model.* An embedding made from 270 range images, the remaining 270 are then projected onto the embedding one by one.

4 Summary

LLE proved to capture the nature of the change between the points in a very efficient manner; both finding on which side the object is lying and the planar angle. LLP is also a very efficient way to quickly project an image to a pre-calculated model. Creating a model from range images also showed advantages over intensity images; i.e. smoother curves better capturing the nature of the points.

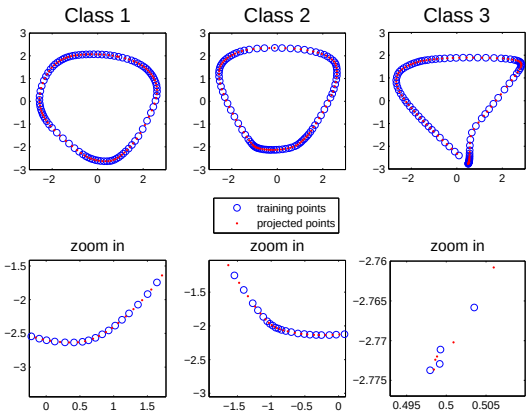

Fig. 6. *The Intensity Model.* An embedding made from 270 intensity images, the remaining 270 are then projected onto the embedding one by one.

References

1. S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000.
2. S. Roweis and L. Saul. Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *Journal of Machine Learning Research*, 4:119–155, 2003.
3. J.B. Tenenbaum, V. de Silva, and J.C. Langford. A global geometric framework for nonlinear dimension reduction. *Science*, 290:2319–2323, 2000.
4. M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *MIT Press, Neural Computation*, 15:1373–1396, 2003.
5. Y. Bengio, J. Paiement, and P. Vincent. Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering. Technical Report 1238, Université de Montréal, 2004.
6. J. Ham, Y. Lin, and D. Lee. Learning nonlinear appearance manifolds for robot localization. *IEEE Pacific Rim Conference on Communications, Computers and signal Processing*, pages 2971–2976, 2005.
7. D. de Ridder and R. Duin. Locally linear embedding for classification. *TUDelft, Pattern Recognition Group Technical Report Series*, PH-2002-01, 2002.
8. B. K. P. Horn. *Robot Vision (MIT Electrical Engineering and Computer Science)*. The MIT Press, 1986.
9. I. Balslev and R. Larsen. Scape vision, a vision system for flexible binpicking in industry. *IMM Industrial Visiondays, DTU*, 2006.
10. M. Turk and A. Pentland. Eigenfaces for recognition. *Journal for Cognitive Neuroscience*, 3:71–86, 1991.
11. N. Blanc, F. Lustenberger and T. Oggier. Miniature 3D TOF Camera for Real-Time Imaging. *Springer-Verlag, Lecture Notes in Computer Science*, 4021:212-216, 2006.