MVA2015 IAPR International Conference on Machine Vision Applications, May 18-22, 2015, Tokyo, JAPAN

Three-DoF Pose Estimation of Asteroids by Appearance-based Linear Regression with Divided Parameter Space

Naoki Kobayashi[1], Yuji Oyamada[1,2], Yoshihiko Mochizuki[1,2], and Hiroshi Ishikawa[1,2]

[1]Department of Computer Science and Engineering, Waseda University
[2]JST CREST

Abstract

We present an appearance-based linear regression method for pose estimation from a single image of an asteroid, which can have any pose in the full space of three degree-of-freedom rotation parameters. The method is characterized by its division of the parameter space into multiple regions. Given a large number of training images with known pose parameters, we learn the relationship between the images and the pose parameters, separately for each parameter region, using the standard linear pose estimation. We also create a common subspace such that, when projected to it, the difference between images in the same parameter region tends to collapse. In estimating the pose of an input image, we project it onto the common subspace to determine the parameter region. We apply the method for pose estimation from asteroid images and report the experimental results.

Figure 1. **Top row:** Input images of asteroid *Itokawa* taken by the *Hayabusa* (MUSES-C) space probe [9]. **Bottom row:** Rendering of the 3D model using the estimated pose from the input image above in the same column.

1 Introduction

Estimating the pose of three-dimensional objects from a single image has many practical applications and thus has been the focus of many research work. Our objective and motivation for this work is to determine the pose of an asteroid from a single image taken by the planned *Hayabusa 2* space probe [1], after it has obtained enough images of the asteroid from a distance to construct a 3D model.

There have been two main approaches to the problem of object pose estimation. In one approach, there are model-based methods [2] that extract structural properties of the object from the image and compare them with a known 3D model of the same object to estimate the pose. For our purpose, we found that asteroids lack convenient structural properties, unlike artificial objects. Another approach is appearance-based, where the image itself is treated as a vector and a large number of training images of the object with known pose parameters are compared with it [3, 4, 5, 6, 7, 8]. More directly, a regression map that gives the pose parameters when applied to an image can be obtained from the training images. Images of asteroids are especially suited for the appearance-based approaches, since the background is the blackness of space..

If the regression map is global, in the sense that there is a single map covering the whole of the space of possible pose parameters, the map tends to be highly non-linear, requiring more complex methods such as support vector regression [3]. Conversely, when the regression map is linear, it becomes impossible to globally map the images to pose parameters as the manifold of images with all possible poses embedded in the space of possible images becomes more complex. To overcome this limitation, there are locally linear embedded regression methods that takes training images near the input image to create the regression map on the fly [7, 8].

Contribution. However, to our knowledge there has been no method using appearance-based linear regression that is capable of estimating pose in the full space of three degree-of-freedom rotation parameters. In this paper, we overcome this limit by a simpler approach than support vector regression or locally linear embedded regression: dividing the parameter space into fixed multiple regions and creating a separate regression matrix for each region. By allowing the regression matrices to differ for different parameter regions, we

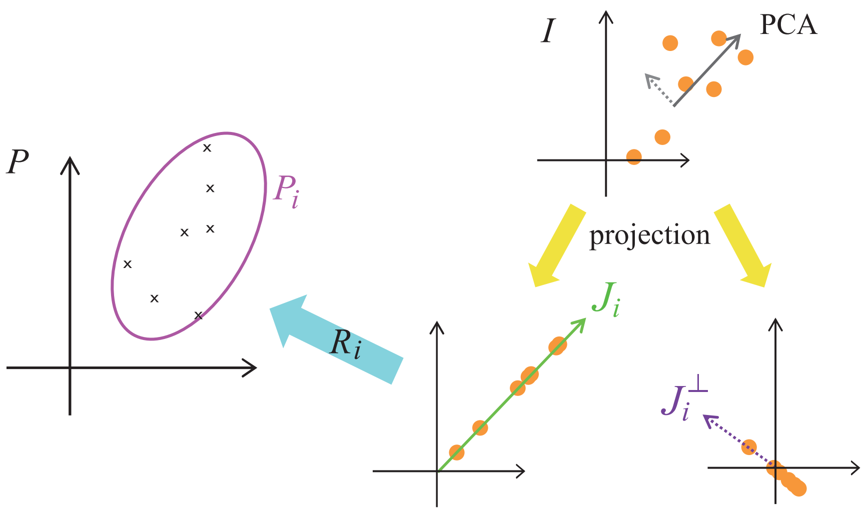

Figure 2. PCA is performed on the set of training images with pose parameters in each subregion of the parameter space. The eigenvectors with large eigenvalues form the subspace J_i while those with small eigenvalues form the subspace $J_i^\perp$. The linear regression map R_i gives the parameter values for images projected onto J_i. For an input image, its parameter region P_i is first determined by projecting it to the intersection of all $J_i^\perp$'s and finding the nearest training images. Then R_i is used to obtain the parameters.

improve the accuracy of pose estimation in each region. Also, since the regression maps are created in the learning phase, it is faster than creating it on the fly in the estimation phase. Of course, this gives rise to a new problem of determining which regression matrix to use for a given input image. For this, we create a vector subspace such that, when projected to it, the difference between images in the same parameter region tends to collapse. The input image is first projected onto this "common orthogonal subspace", and the nearest training image neighbors are found. Then the regression matrix that belongs to the parameter region corresponding to the majority of the neighbors is applied to the input image to yield the estimation of the rotation parameters.

We explain the method in more detail in the next section, report the result of pose estimation experiments in Section 3, and conclude in Section 4.

2 The Method

In the following, we assume all images, training and input alike, are normalized as much as possible, *e.g.*, they are of the same size, with the object centered. Let us denote the vector space of images by I and its dimension by d. Vectors in I are considered column vectors in the following. We divide the vector space P of pose parameters into N subsets, which we call *parameter regions* and denote by $P_1, \dots, P_N$, so that $P = P_1 \cup \cdots \cup P_N$.

2.1 Training Phase

We assume that training images are provided, showing the same object from different views, each with known pose parameters. Let I_i be the set of training images with pose parameters in P_i. We denote the training images in I_i by $v^i_1, \dots, v^i_{n_i} \in I_i$ and their corresponding pose parameters by $x^i_1, \dots, x^i_{n_i} \in P_i$. Define two matrices

$$V_i = (v^i_1, \dots, v^i_{n_i})$$

and

$$X_i = (x^i_1, \dots, x^i_{n_i})$$

that have the training images and their corresponding parameters as columns, respectively.

We also fix two paremeters d_1 and d_2 as the dimensions of the vector subspaces $J_i (i = 1, \dots, N)$ and $J^\perp$, respectively. The steps in the training phase is as follows.

1. Perform the principal component analysis on each $I_i (i = 1, \dots, N)$.

2. Obtain the d_1-dimensional subspace J_i of I spanned by the eigenvectors with the largest eigenvalues.

3. Fix an orthonormal basis

$$E_i = (e^i_1, \dots, e^i_{d_1})$$

of J_i. We consider E_i as a matrix with d_1 columns and d rows. The matrix $A_i = E_i^{\mathrm{T}} V_i$ consists of the coefficient vectors for the training images with respect to the basis E_i.

4. For each $i = 1, \dots, N$, let

$$R_i = X_i A_i^+,$$

where

$$A_i^+ = A_i^{\mathrm{T}} (A_i A_i^{\mathrm{T}})^{-1}$$

is the Moore-Penrose pseudo inverse of A_i. Then the pose parameter vector of image $v \in I$ can be obtained as $R_i E_i^{\mathrm{T}} v$.

5. Obtain the $\frac{(N-1)d+d_2}{N}$-dimensional subspace $J_i^\perp$ of I spanned by the eigenvectors with the least eigenvalues.

6. Take the intersection of $J_i^\perp$'s to obtain the common orthogonal subspace
$$J^\perp = \cap_i J_i^\perp.$$
Then, under the assumption of genericity, the dimension of $J^\perp$ will be d_2. We also orthogonally project all the training images onto $J^\perp$. Let us denote the training image v_j^i projected onto $J^\perp$ by u_j^i.

Because $J^\perp$ consists of vectors with small eigenvalues with respect to the PCA on all of I_i, difference between projections to $J^\perp$ of images that are in the same I_i tends to be minimized (Fig. 2). Hence we use it to classify the input image into the parameter regions.

2.2 Estimation Phase

Having thus prepared, the pose estimation can be done as follows. For an input image, we first determine its parameter region P_i by projecting the image to $J^\perp$ and then finding the k nearest vectors that are training images projected onto $J^\perp$. Let m_i be the number of u_j^i's in the k vectors. We assume that the pose of the input image also belongs to P_i with maximum m_i and use R_i to obtain the pose parameters.

3 Experiments

In this section we report the result of an application of the method described in the previous section.

3.1 Experimental Setup

Our objective and motivation for this work is to determine the pose of an asteroid from a single image taken by the planned *Hayabusa 2* space probe [1]. However, since *Hayabusa 2* is yet to be launched, we tested our method with the images and 3D models of asteroid *Itokawa* obtained by its predecessor *Hayabusa* [9].

Training images. We used a 3D model [10] of asteroid *Itokawa* to generate training data. The images are all 40×40 pixel resolution, so the dimension $d = 1600$. We measure the pose by the x-z-x Euler angle (θ, ϕ, ψ) in the range of
$$[0°, 360°) \times [0°, 180°) \times [0°, 360°).$$
We rendered the total of $36 \times 18 \times 36 = 23328$ training images, corresponding to the poses in 10° intervals, *i.e.*
$$\begin{aligned}&(0°, 0°, 0°), (10°, 0°, 0°), (20°, 0°, 0°), \ldots\\ &\ldots, (350°, 0°, 0°), (0°, 10°, 0°), (10°, 10°, 0°), \ldots\\ &\ldots, (350°, 10°, 0°), (0°, 20°, 0°), (10°, 20°, 0°), \ldots\\ &\ldots, (350°, 170°, 0°), (0°, 0°, 10°), (10°, 0°, 10°), \ldots\\ &\ldots, (350°, 170°, 350°).\end{aligned}$$

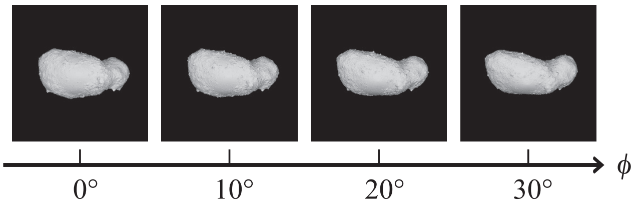

Figure 3. Sample training images.

Table 1. Mean absolute error of pose estimation angles for the linear regression method with and without the divided parameter space.

	Proposed method			Undivided method		
Angle	θ	ϕ	ψ	θ	ϕ	ψ
Error	5.81	2.83	5.48	34.07	29.47	59.27

Fig. 3 shows sample training images. We divided the training images into $N = 32$ non-overlapping sets, according to the combination of $4 \times 2 \times 4 = 32$ regions in which each angle has 90° range. Thus, each I_i contained 729 images.

Parameter space. As the parameter space P used in the method, we used the vector space of 3×3 real-valued matrices, rather than the three-parameter space of Euler angles. Thus, for each rendered image, we used the rotation matrix corresponding to the Euler angles as the corresponding parameter. In the estimation phase, the regression map R_i also gives a 3×3 matrix, which is not necessarily a rotation matrix; let us denote it by M. To recover the rotation parameters, we first find the rotation matrix M' nearest to M. This can be done by singular value decomposition:
$$M = UWV^T,$$
where U, V are orthogonal matrices and W is a diagonal matrix. By replacing W by the identity matrix, we obtain an orthogonal matrix
$$M' = UV^T.$$
The Euler angle can then be recovered from M' in a standard way.

Other parameters. The dimensions for the subspaces were $d_1 = 300$ and $d_2 = 320$. The number of nearest neighbors to search in the common orthogonal subspace $J^\perp$ was $k = 4$.

3.2 Results

The experiments were carried out on an AMD Opteron 6276 Processor with 2.3GHz clock speed. The learning phase took about 72 hours without any parallelization, most of which was spent finding the common orthogonal subspace $J^\perp$. The pose estimation takes about 10 seconds for each input image.

Quantitative evaluation. To quantitatively evaluate the regression error, we rendered 23328 test images of the same 3D model with the angle parameters

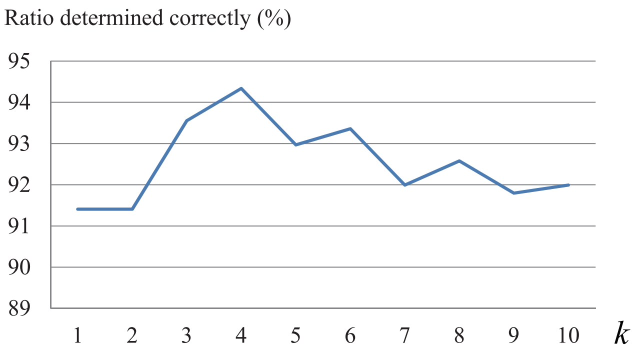

Figure 4. The ratio of the cases where the correct parameter region was determined for a test image. The horizontal axis is k, which is the number of neighboring training images to find in the common orthogonal subspace.

coming between those of training images, *i.e.*,

$$(5°, 5°, 5°), (15°, 5°, 5°), \ldots, (355°, 175°, 355°),$$

so that no training image is included in the test images. To measure the effect of the division of the parameter space, we compared the proposed method with the conventional pose estimation with a single linear regression matrix. Then we measured the absolute angle difference between the true value and the estimated value of the three angles θ, ϕ and ψ. Table 1 shows the results for the proposed method with divided parameter space, assuming the correct region is found, and the conventional method with a single linear regression map.

Determining the parameter region. Separately, we tried different value of k, the number of neighboring training images to find in the common orthogonal subspace. Fig. 4 shows the ratio of the cases where the correct parameter region was determined.

Evaluation with real images. We also tested on real images of asteroid *Itokawa* taken by the *Hayabusa* space probe [9]. Although we do not have the true pose parameters, the estimation results seem to be reasonable, as can be seen from the comparison of the input images and the rendering of the 3D model using the estimated pose parameters (Fig. 1).

4 Conclusion

In this paper, we present an appearance-based linear regression method for pose estimation from a single image of a 3D object, which can have any pose in the full space of three degree-of-freedom rotation parameters. We accomplish this by dividing the parameter space into fixed multiple regions and creating a separate regression matrix for each region, a linear approach simpler than support vector regression or locally linear embedded regression. To determine which regression matrix to use for a given input image, we create a "common orthogonal subspace" such that, when projected to it, the difference between images in the same parameter region tends to collapse. By allowing the regression matrices to differ for different regions, we improve the accuracy of pose estimation in each region. Also, since the regression maps are created in the learning phase, it is faster than creating it on the fly in the estimation phase. We report the result of experiments applying the method for pose estimation using actual asteroid images and 3D models.

Acknowledgement

This work was partially supported by KAKENHI 25287114 and 24300075 from JSPS as well as CREST from JST.

References

[1] Japan Aerospace Exploration Agency, "Asteroid Explorer "Hayabusa2"," `http://www.jaxa.jp/projects/sat/hayabusa2/index_e.html`.

[2] V. Lepetit and P. Fua, "Monocular model-based 3D tracking of rigid objects: A survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 1, no. 1, pp. 1–89, 2005.

[3] S. Ando, Y. Kusachi, A. Suzuki, and K. Arakawa, "Appearance based pose estimation of 3d object using support vector regression," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, Sept 2005, vol. 1, pp. I–341–4.

[4] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-D objects from appearance," *International Journal of Computer Vision*, vol. 14, no. 1, pp. 5–24, 1995.

[5] T. Okatani and K. Deguchi, "Yet another appearance-based method for pose estimation based on a linear model," in *IAPR Workshop on Machine Vision Applications*, 2000, pp. 258–261.

[6] T. Melzer, M. Reiter, and H. Bischof, "Appearance models based on kernel canonical correlation analysis," *Pattern Recognition*, vol. 36, no. 9, pp. 1961–1971, 2003.

[7] B. Raytchev, K. Terakado, T. Tamaki, and K. Kaneda, "Pose estimation by local procrustes regression," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, Sept 2011, pp. 3585–3588.

[8] B. Raytchev, K. Terakado, T. Tamaki, and K. Kaneda, "Object pose estimation by locally linearly embedded regression," in *ICONIP*. 2013, vol. 8228 of *Lecture Notes in Computer Science*, pp. 458–468, Springer Berlin Heidelberg.

[9] Japan Aerospace Exploration Agency, "Asteroid Explorer "Hayabusa" (MUSES-C)," `http://www.jaxa.jp/projects/sat/muses_c/index_e.html`.

[10] R. W. Gaskell, "Optical-only determination of small body shape and topography," in *International Society for Photogrammetry and Remote Sensing WG IV/9 Extraterrestrial Mapping Workshop Houston, TX, USA*, 2003.