Comparison of 3 DOF Pose Representations for Pose Estimations

Kengo Harada† Satoko Tanaka†
Toru Tamaki† Bisser Raytchev† Kazufumi Kaneda† Toshiyuki Amano‡

†: Hiroshima University, Japan ‡: NAIST, Japan

Abstract We report a result of an experimental study on properties of pose representations for 3 DOF linear pose estimations with 100 CG objects. We use linear regression as a pose estimation method. First, we explain a method of linear pose estimation and two properties of pose representations. Next, we use four pose representations (rotation matrix, ZYX Euler angle, exponential map, and unit quaternions), for pose estimation experiments, and compare estimation errors. We show that estimation error of rotation matrix is significantly smaller than other representations by using pairwise t–test.

1 Introduction

In this paper we show experimental evaluation of properties of representations of 3 degrees-of-freedom (DOF) rotation for global appearance-based (view-based) pose estimation, which estimate a pose of a test image based on a training of relations between training sample images and corresponding pose parameters.

The pose parameters should be taken care [1] because the parameters should represent pose continuously and bijectively as the change of views of an object in an image. However, pose parameters have been used carelessly for linear and non-linear 3 DOF pose estimation methods.

In this paper, we focus how the representations affect on results of a linear pose estimation method, and experimentally evaluate the properties — continuity and bijection — of four major representations of 3 DOF rotation pose: a rotation matrix, Euler angles (or roll–pitch–yaw), Exponential map (or angle–axis) and unit quaternions. We show experimental results and comparisons of pose representations by using 100 objects with a linear pose estimation.

2 Properties of pose representations

In this section, we describe briefly a linear pose estimation method with global appearance of an object in an image, and then properties of pose representations that should be satisfied.

2.1 A linear pose estimation method

Linear pose estimation methods [7, 4, 8, 9] learn relations between images and poses. Here we focus on the most simple one; a liner regression method [4]. Given a training sample set that consists of images $\boldsymbol{x}_j$ and poses $\boldsymbol{p}_j$. Then, the following linear map (that is, a matrix) is estimated:

$$\boldsymbol{p}_j = F\boldsymbol{x}_j, \quad j = 1, 2, \ldots, n. \tag{1}$$

For a test image $\boldsymbol{x}$, an estimate of its pose $\boldsymbol{p}$ is obtained by $\boldsymbol{p} = F\boldsymbol{x}$.

2.2 Pose estimate as approximation

The pose estimation described above can be seen as an approximation with training poses. An image $\boldsymbol{x}$ can be approximated with training images $\boldsymbol{x}_j$ by their linear combination

$$\boldsymbol{x} \cong \sum_j b_j \boldsymbol{x}_j, \tag{2}$$

when $\boldsymbol{x}$ belongs to the same class that $\boldsymbol{x}_j$ describe. This linearity of images is well known and used for many applications such as eigenfaces.

We can rewrite the equation above with Eq.(1) as follows;

$$\boldsymbol{p} = F\boldsymbol{x} \cong \sum_j b_j F\boldsymbol{x}_j = \sum_j b_j \boldsymbol{p}_j. \tag{3}$$

This means that a pose estimate $\boldsymbol{p}$ is represented by a linear combination of training poses $\boldsymbol{p}_j$.

Table 1: Pose representations and properties.

representation	parameters	bijection	continuity
rotation matrix	$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{33} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$	√	√
ZXY Euler angles	$[\theta_x, \theta_y, \theta_z]^T, \quad -\pi \leq \theta_{x,y,z} < \pi$	×	×
Exponential map	$\boldsymbol{\omega} = [\omega_1, \omega_2, \omega_3]^T, \quad 0 \leq \|\|\boldsymbol{\omega}\|\| \leq \pi$	×	×
unit quaternions	$\boldsymbol{q} = [q_0, q_1, q_2, q_3]^T$	×	√

2.3 Continuity and bijection

Eq.(3) requires to pose representations two properties — continuity and bijection [1].

A representation of pose $\boldsymbol{p}$ should be continuous with appearance. An example of a pose representation that is not continuous is an angle in radian; there is a discontinuity at $2n\pi$ for $n = 0, 1, lodts..$ This causes a problem because the pose change can not be handled by the linear approximation of poses at the discontinuity.

Also $\boldsymbol{p}$ should be bijective to images: a correspondence between a pose and an image is a one-to-one mapping. If two poses $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ could correspond to an image $\boldsymbol{x}$, then the following inconsistent equations might hold:

$$\boldsymbol{p}_1 = F\boldsymbol{x}, \quad \boldsymbol{p}_2 = F\boldsymbol{x}. \tag{4}$$

Even if we fix the equations by using one of them (for example, use only $\boldsymbol{p}_1$ and discard $\boldsymbol{p}_2$), the linear combination of poses may produce incorrect results because of unbalanced training samples.

2.4 Pose representations and properties

Here we describe four major representations of 3 DOF pose and their properties from the view point of continuity and bijection (see Tab. 1).

A rotation matrix R is an element of the special orthogonal group $SO(3) = \{R \in \mathbb{R}^{3\times3} \mid RR^T = R^TR = I, \det(R) = 1\}$. R is continuous and bijective because $SO(3)$ is a smooth closed manifold.

ZYX Euler angles are one of famous representation in vision and robotics. θ_x, θ_y and θ_z are angles about x, y and z axes. They are sometimes called fixed angles (but with a definition different fro Euler angles) or roll–pitch–yaw angles, and have several variants: the order of angles matters and we can use other orders such as ZYX, XYZ, and so on. Euler angles are frequently used, however, they are not continuous nor bijective because of the gimbal lock problem [5]: angles suddenly change as a pose changes smoothly.

Exponential map is a mapping from a rotation to an exponential form with an axis and angle of the rotation. A three-dimensional vector $\boldsymbol{\omega}$ represents the rotation axis by its direction and the angle by its norm. This means that a negative direction with a negative angle of $\boldsymbol{\omega}$ coincides to $\boldsymbol{\omega}$ itself (see fig. 1), and then it is not bijective.

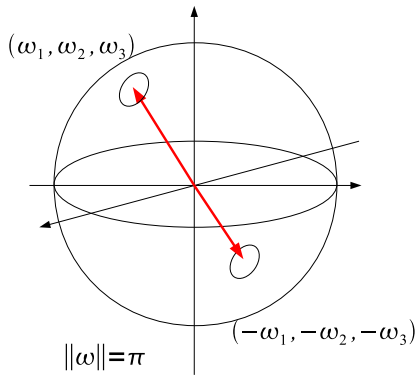


Fig. 1: Exponential map. $\boldsymbol{\omega}$ and $-\boldsymbol{\omega}$ are the same pose.

Unit quaternions are widely used for graphics, vision and robotics. A quaternion has one real part and three different imaginary parts, and operations are defined on a quaternion field. However, a unit quaternion can be seen as an four-dimensional vector $\boldsymbol{q} = (q_0, q_1, q_2, q_3)^T$ with a unit norm, and its components correspond to real and imaginary parts. Usually, the first element q_0 is *cos* of a rotation angle and the other elements q_1, q_2, q_3 are the direction of the rotation axis. This results in two unit quaternions with opposite signs, $\boldsymbol{q}$ and $-\boldsymbol{q}$, represent the same pose, hence not bijective. If we use positive unit quaternions (that is, discard $-\boldsymbol{p}$), then discontinuity arises around the edge of the hyperhemisphere (as shown in fig. 2).

3 Experiments

To see how representations affect the accuracy of pose estimation, we compared estimation errors of a linear regression method with four pose representations: rotation matrix, Euler angles, Exponential

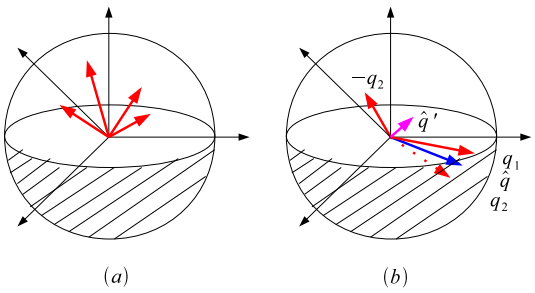


Fig. 2: Unit quaternions.

map, and unit quaternions.

3.1 Estimation method

First, we describe how we estimate a pose by using a linear regression.

Given a training image $\boldsymbol{x}_j$, a pose parameter is also given in a vector form. For a rotation matrix, $\boldsymbol{p} = (r_{11}, \ldots, r_{33})^T$ is a nine-dimensional vector consisting of nine elements in a rotation matrix. For Euler angles we use $\boldsymbol{p} = (\theta_x, \theta_y, \theta_z)^T$, for Exponential map $\boldsymbol{p} = \boldsymbol{\omega}$, and for unit quaternions $\boldsymbol{p} = \boldsymbol{q}$.

Then equations $\boldsymbol{p}_j = F\boldsymbol{x}_j$ are stacked to construct a system of equations. Usually this system is under-determined because the number of samples are smaller than the number of unknowns (that is, the same with the dimensionality of $\boldsymbol{x}$). Therefore, a pseudo-inverse is used to solve the system and obtain F.

A pose of an image $\boldsymbol{x}$ is then estimated by $\boldsymbol{p} = F\boldsymbol{x}$. However, $\boldsymbol{p}$ should be corrected to ensure that it is actually 3 DOF pose. For rotation matrix, first $\boldsymbol{p}$ is rearranged to a 3×3 matrix. Then, it is converted to a rotation matrix that is the most closest to it by using the polar decomposition [6]. For unit quaternions, $\boldsymbol{p}$ is normalized to have norm 1.

3.2 Experimental setup

We used 100 different 3D objects (Fig. 7) from a commercial dataset of 3D models with complex textures [10]. Because appearances of objects affect directly estimation results, we evaluated estimation errors for many objects and averaged them.

2500 images were created for training each object by specifying rotation matrices for a OpenGL renderer. Some images are shown in Fig. 3; 128×128 in size, gray background, and auto-adjustment of scale of an object. 100 images were created for testing. Errors are averaged for each object, and each pose representation. Note that 2500 training images and 100 test images are common for four representations in order to compare their errors.

Fig. 3: Sample images used for training and testing.

We focused poses around discontinuity: some pose representations have discontinuity at a rotation angle of π. Training and test images were created around the pose of π rotation as follows:

- Create a rotation matrix R_z with a rotation about z axis by π.
- Create a small random rotation R_s with a rotation about a random axis (a spherical random generator is used) by an angle uniformly distributed in $[0, \pi/6]$ [rad].
- Combine them: a (true) rotation matrix is $R_t = R_s R_z$.

3.3 Error metric

We evaluated errors of estimated poses.

Normalization As described in section 3.1, estimated poses of each representation were normalized to represent actually 3 DOF pose.

Conversion to rotation matrix Normalized poses are then converted to corresponding rotation matrix denoted by $\tilde{R}$.

Error as distance between rotation matrices We used a distance between rotation matrix [11] to evaluate errors between true and estimated rotations.

Let R_t be a true rotation matrix and $\tilde{R}$ an estimated rotation matrix. The distance $d_F(R_t, \tilde{R})$ between R_t and $\tilde{R}$ is defined as follows:

$$d_F(R_t, \tilde{R}) = \frac{1}{\sqrt{2}} \| \log R_t \tilde{R}^T \|, \tag{5}$$

$$\log R = \begin{cases} 0, & (\theta = 0) \\ \frac{\theta}{2 \sin \theta}(R - R^t), & (\theta \neq 0) \end{cases} \tag{6}$$

where $\theta = \cos^{-1}(\frac{trR-1}{2})$.

Note that this error d_F is actually the angle of a rotation between two rotation matrices. Suppose R_d satisfies $R_t = R_d\tilde{R}$, then d_F is an angle of the rotation of R_d in radian.

3.4 Comparison of estimation errors

Figure 4 shows estimation error of each object with each representation of pose. Each bar shows an average of 100 estimation errors with standard deviations: for each representation, poses of 100 test images of each object were estimated.

Figure 5 shows estimation error of all objects with each representation of pose. Each bar shows an average of 10,000 estimation errors with standard deviations: for each representation, poses of 100 test images of 100 objects were estimated. The result shows that the error of rotation matrix is small rather than their other representations. Average errors of Euler angles and Exponential map is about 1.5 radian (85 degrees) which means that they are not appropriate for the use of this kind of applications. Examples of images are shown in Fig. 6 in which such representations have failed to estimate poses.

To see the difference of errors statistically, we performed pairwise t-test between rotation matrix and the others. There are 10,000 pairs for each test. This test showed significant differences for all test ($p < 0.01$).

4 Conclusions

We have shown experimental results of comparison of pose representations for global appearance-based pose estimation of an object. We used rotation matrix, Euler angles, Exponential map, and unit quaternions as representations of pose of 3 DOF rotation. Experiments with 100 objects demonstrated that the error with the use of rotation matrix is significantly smaller than the errors of other representations.

References

[1] T. Tamaki, T. Amano and K. Kaneda, "Yet another representation of $SO(3)$ pose by spherical functions,' *Proc. or FCV2010*, 2010.

[2] J. J. Craig, *Introduction to Robotics: Mechanics and Control*. Prentice Hall, 3 ed., 2004.

[3] Y. Ma, S. Soatto, J. Košecká, and S. S. Sastry, *An Invitation To 3-D Vision*. Springer, 2004.

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

[5] F. Nielsen, *Visual Computing: Geometry, Graphics, and Vision*. Charles River Media, 2005.

[6] Z. Zhang, "A flexible new technique for camera calibration", *Technical Report MSN–TR–98–71*, Microsoft Research, 1998.

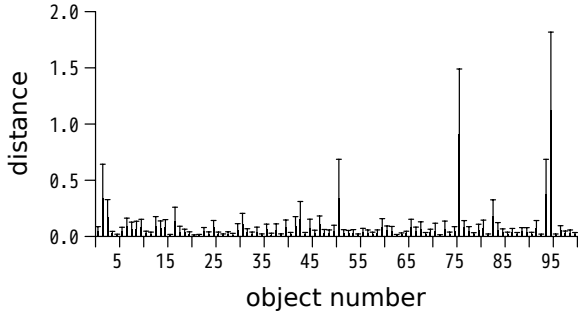
[7] T. Amano and T. Tamaki, "An appearance based fast linear pose estimation," *IAPR Conference on Machine Vision Applications 2009*, pp. 182–186, 2009.

[8] T. Melzer, M. Reiter, and H. Bischof, "Appearance models based on kernel canonical correlation analysis," *Pattern Recognition*, vol. 36, pp. 1961–1971, 2003.

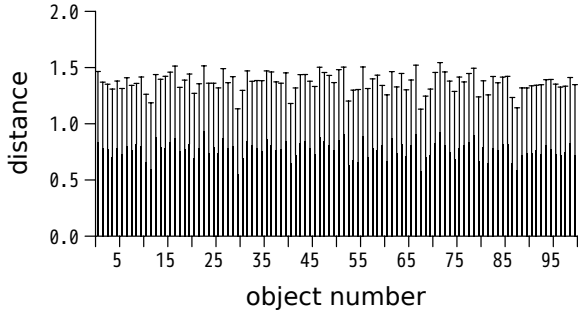
[9] S. Fidler, D. Skočaj, and A. Leonardis, "Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 337–350, 2006.

[10] J. Wiedemann (ED.), *500 3D-OBJECTS*. vol. 1, Taschen, 2002.

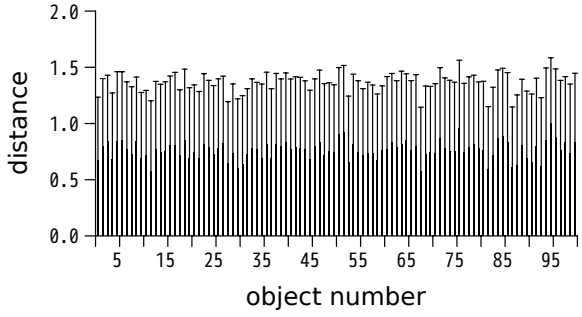
[11] M. Moakher, "Means and Averaging in The Group of Rotation", *SIAM Journal on Matrix Analysis and Applications*, vol. 24 , no. 1, pp. 1–16, 2002.



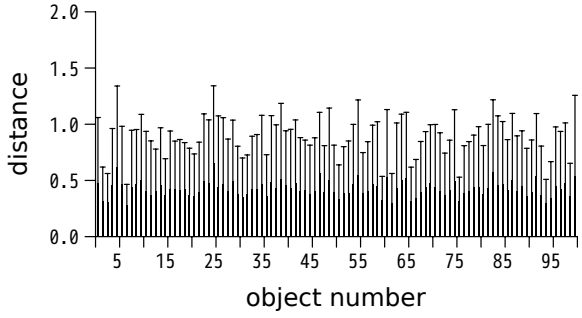
(a)



(b)



(c)



(d)

Fig. 4: Errors for each object with (a) rotation matrix, (b) Euler angles, (c) Exponential map, and (d) unit quaternions. Horizontal axis is object numbers shown in Fig. 7.

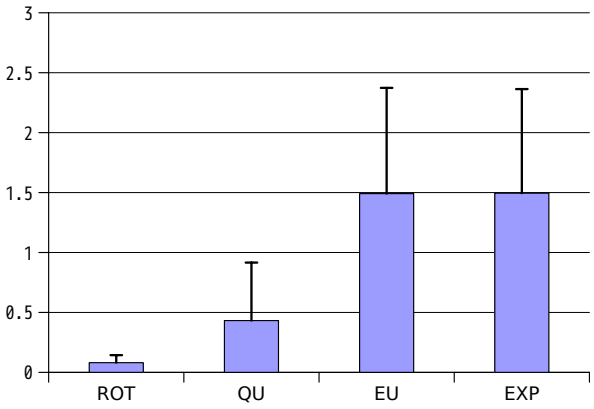
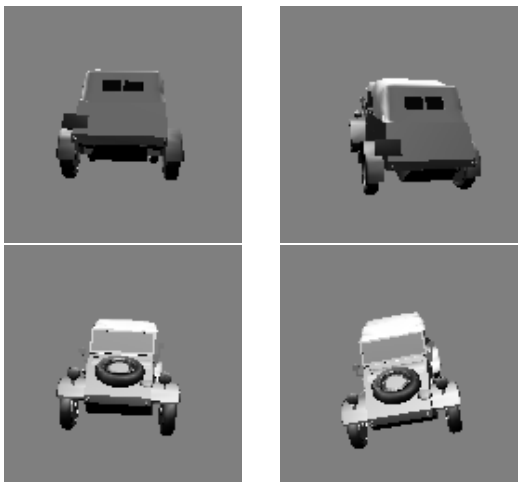


Fig. 5: Average errors of pose estimations. Vertical axis is the distance d_F (in radian) between true and estimated rotations.



(a) (b)

Fig. 6: Examples of erroneous estimation with (a) Euler angles and (b) Exponential map. Top row shows images of true poses and bottom row shows images of estimated poses.

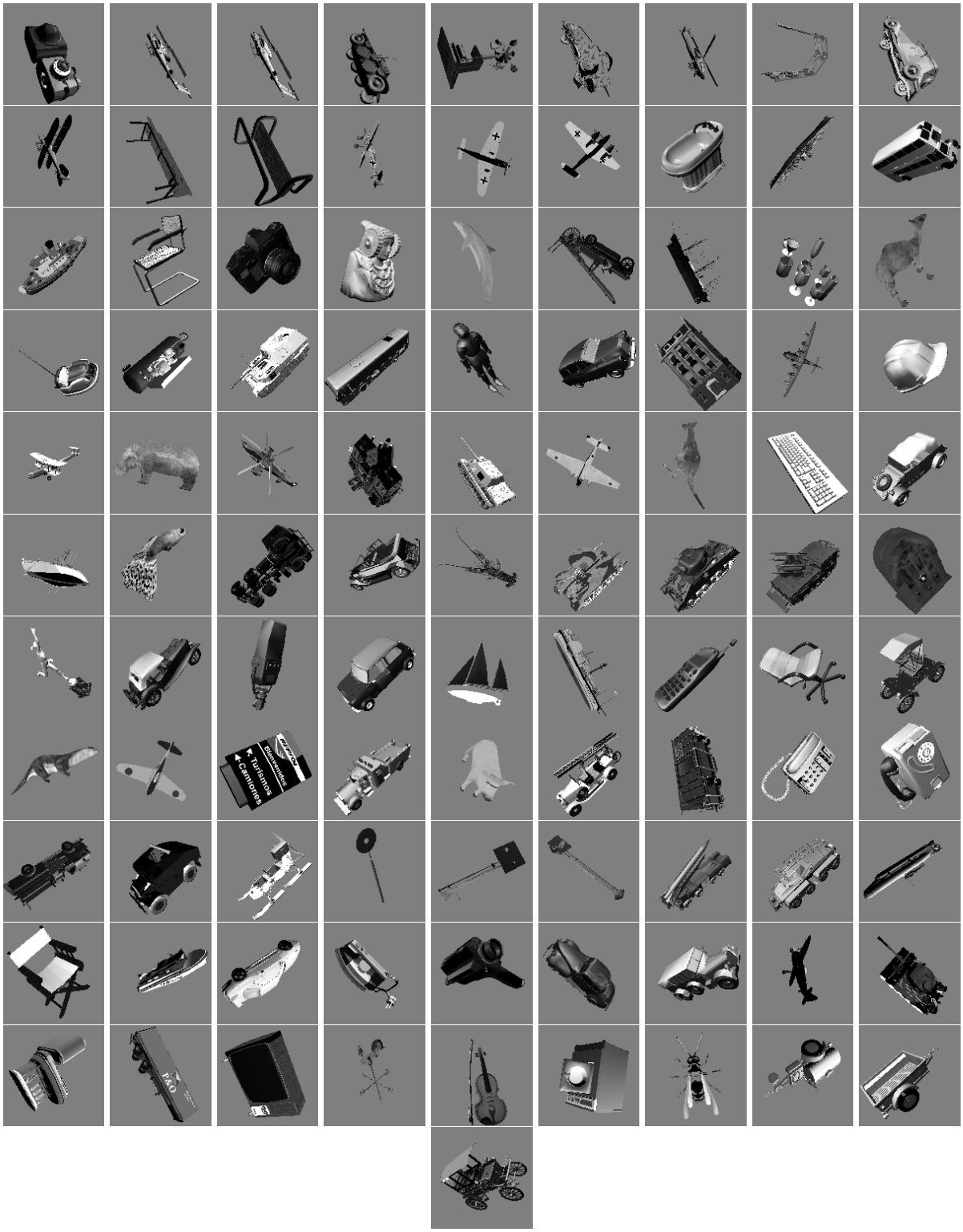


Fig. 7: 100 objects from a 3D model dataset [10] used in the experiment.