The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-2/W2-2022
Optical 3D Metrology (O3DM), 15–16 December 2022, Würzburg, Germany

IMAGE ORIENTATION BY EMBEDDING IN A GAN LATENT SPACE

V. V. Kniaz[a,b]*, V. A. Knyaz[a,b], V. Mizginov[a], A. Bordodymov[a], P. Moshkantsev[a], S. Barylnik[a], D. Novikov[a]

[a] State Res. Institute of Aviation Systems (GosNIIAS), 125319, 7, Victorenko str., Moscow, Russia
(vl.kniaz, knyaz, vl.mizginov, bordodymov, moshkantsev, semyon_barylnik, daninov)@gosniias.ru
[b] Moscow Institute of Physics and Technology (MIPT), Russia

KEY WORDS: exterior orientation, deep learning, generative models

ABSTRACT:

Estimation of an image exterior orientation is required in multiple tasks including 3D reconstruction, autonomous driving and navigation, and single-photo 3D reconstruction. The problem can be easily solved if some reference points (keypoints) with known coordinates in the reference frame are detected in the image. While multiple robust keypoint detectors were developed, estimation of an image exterior orientation from a single image remains challenging in many cases. For example, repeating structures in the scene or absence of textures can reduce the performance of keypoint detectors. In this paper, we propose an algorithm for estimation of an image exterior orientation that leverages the latent space of Generative Adversarial Network (GAN). We propose a modification of the `StyleGAN2` model that we term `ExteriorGAN`. Unlike the `StyleGAN2` that generates random images from a random noise vector z, we aim training a mapping from a random vector z and a given image exterior orientation p $G : (z, p) \rightarrow A$. Our model generates random images for a constant exterior orientation p and random z that have constant geometry but differs in the scene appearance (e.g. different light direction or intensity). We perform embedding of the image A into the latent space w and reconstruct the input noise vector z and exterior orientation parameters w using a stochastic gradient descent. We developed a dedicated dataset with 50k images and corresponding orientation parameters to train and validate our `ExteriorGAN` model. The results of evaluation demonstrate that our algorithm allows estimation of the exterior orientation of an image with respect to a known 3D scene. The accuracy of the exterior orientation is comparable with modern state-of-the-art methods. The camera pose can be recovered with a mean error of 50 mm for a working space of 5 by 5 meters.

1. INTRODUCTION

Estimation of camera orientation is one of the key problems of photogrammetry usually it is required to estimate the camera pose and rotation from a given image or a set of images. Such task is commonly called estimation of the camera external orientation parameters in contrast to estimation of camera lens parameters that are commonly called interior orientation parameters. Also in computer vision estimation of camera exterior orientation is sometimes referred as camera pose estimation. The complexity orientation from a given image is related to the number of keypoints visible in the image.

Multiple methods have been proposed in the literature focused om estimation of the camera exterior orientation. These methods can be broadly divided into two groups: keypoint-based and scene-based Estimation of an image exterior orientation is the process of recovering the spatial pose and rotation parameters of a camera with respect to a given reference system Image external orientation is required in multiple tasks including 3D reconstruction, autonomous driving and navigation, and single-photo 3D reconstruction. The problem can be easily solved if some reference points (keypoints) with known coordinates in the reference frame are detected in the image. Hence, most of modern external orientation algorithms such as DLT (Abdel-Aziz and Karara, 1971) and EPnP (Lepetit et al., 2009) require n known 3D points to be detected in the image. This problem is also sometimes referenced as Perspective-n-Point (PnP), where n is the number of required points. While multiple robust keypoint detectors were developed, estimation of an image exterior orientation from a single image remains challenging in many cases. For example, repeating structures in the scene or absence of textures can reduce the performance of keypoint detectors. Some exterior orientation methods leverage general assumptions regarding the perspective projection and do not require any keypoints. Usually this methods are based on the vanishing point detection (Verykokou and Ioannidis, 2016a). Recently a number of neural networks (Kehl et al., 2017) have been proposed for the task of the camera pose estimation that is similar to estimation of exterior orientation. While such methods demonstrate encouraging results for challenging scenarios, usually they can estimate exterior orientation only with respect to a limited number of objects that neural model have observed during the training stage.

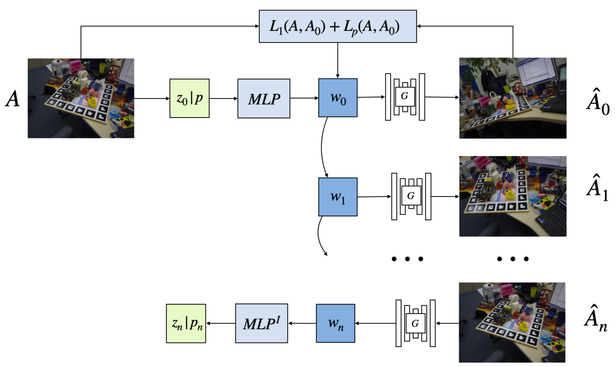

Figure 1. The `ExteriorGAN` neural network model.

In this paper, we propose an algorithm for estimation of an image exterior orientation that leverages the latent space of Generative Neural Network (GAN). GANs (Goodfellow et al., 2014)

* Corresponding author

Figure 2. Examples of color images from our *ExteriorViews* dataset.

have been proposed recently for the task of synthesizing random images from an input random noise vector z. All generated images remain similar to the domain of real images that a GAN model have observed during the training stage. Recently the `StyleGAN2` (Karras et al., 2019) generative model have been proposed that produce a continuous latent space. If two images are visually similar their corresponding latent vectors are close to each other. This allowed to develop robust methods (Abdal et al., 2019, Brooks and Efros, 2022) for embedding images into the GAN latent space. During the process of embedding the algorithm receives an input image A and estimates the corresponding latent vector w in the model's latent space.

We propose a modification of the `StyleGAN2` model that we term `ExteriorGAN` (Figure 1). Unlike the `StyleGAN2` that generates random images from a random noise vector z, we aim training a mapping from a random vector z and a given image exterior orientation p $G : (z, p) \rightarrow A$. Our model generates random images for a constant exterior orientation p and random z that have constant geometry but differs in the scene appearance (e.g. different light direction or intensity). Inspired by (Nie et al., 2020), we propose a supervised training approach that allows us to keep the mapping between exterior orientation parameters and the generated image A. We estimate the exterior orientation parameters vector p using a trained model. We perform embedding of the image A into the latent space w and reconstruct the input noise vector z and exterior orientation parameters w using a stochastic gradient descent.

We developed a dedicated dataset with 50k images and corresponding orientation parameters to train and validate our model. All images were generated using a 3D scene of an industrial room and represent scenarios that a typical for navigation of a mobile robot. We evaluated our algorithm using an independent test split of the dataset. The results of evaluation demonstrate that our algorithm allows estimation of the exterior orientation of an image with respect to a known 3D scene. The accuracy of the exterior orientation is comparable with modern state-of-the-art methods. The camera pose can be recovered with a mean error of 50 mm for a working space of 5 by 5 meters.

2. RELATED WORKS

2.1 Image Exterior Orientation

Image exterior orientation is one of the fundamental problems in photogrammetry that is needed for an estimation of accurate 3D coordinates of imaged scene. Photogrammetric approach is based on consideration of exterior orientation task in Cartesian systems of coordinates, thus having a deal with non-linear problem. As a fundamental problem of photogrammetry, exterior orientation task have been studying from the early years of photogrammetry.

One of the first approaches was Direct Linear Transformation (Abdel-Aziz and Karara, 1971), proposed in 1971. It is based on the assumption that relation between comparator coordinates and object coordinates can be described by direct linear transformation. For estimation of eleven unknown parameters more than six control points are needed, these points not belonging to the same plane.

Then, a number of methods for exterior orientation was proposed, that used three fundamental conditions, namely collinearity, coplanarity and coangularity conditions. Among these methods are approximate methods, such as Church method (Slama et al., 1980), based on coangularity condition; 3D conformal transformation (Dewitt, 1996), providing the absolute orientation of stereomodel; approximate solution for spatial transformation (Kraus, 1997); a number of methods that use line feature or vanishing point (Van den Heuvel, 1998, Kniaz, 2016). These methods allowed finding exterior orientation approximation by linear processing of this non-linear problem.

To find accurate estimation of exterior orientation parameters three fundamental photogrammetric conditions are used. They establish relation between spatial coordinates of object control points and their image coordinates. Then estimation of unknown parameters is determined as root mean squares estimate, considering image coordinates of the control points as observations. These approaches require knowing of coordinates control points and some initial approximations for estimating parameters.

A number of studies addressed to find a closed-form solution without initialization. A solution for exterior orientation determination was proposed, that uses orthonormal matrices for closed-formsolution (Horn et al., 1988). The drawback of this algorithm is possibility of finding incorrect rotation matrix. For the solution of this problem another closed-form method, based on singular value decomposition technique was proposed (Arun et al., 1987). This approach allows reliable determining of the transformation parameters.

So, for automation of exterior orientation procedure various methods were proposed, that aim at automatically detecting of control points and determining initial approximation for unknown parameters. The approach for finding approximate values of exterior orientation parameters, based on Particle Swarm Optimization (Li and Li, 2012), begins from a large domain of possible values of external orientation parameters, and then finds best estimates, giving the minimum of image coordinates residual errors.

The technique for finding approximate exterior orientation parameters of multiple large-scale overlapping oblique aerial images for case of unavailable information from GPS (Verykokou and Ioannidis, 2016b) involves five-steps procedure. It includes: determination of the overlapping image pairs; computation of the transformation from image to the reference system of coordinates; rough estimate of the interior orientation of the camera; true horizon line and nadir point estimate for each image; estimation of approximate exterior orientation for every image.

The technique for improving the transformation accuracy (Yan et al., 2016) uses various sets of local similarities. The technique

Figure 3. Random images produced by our `ExteriorGAN` model.

construct triangular irregular network from the available ground control points, thus obtaining a set of triangles. Three vertices of each triangle are used for determining the local similarities. The new transformation combines these weighted local similarities, that allow to reduce errors.

Last decades exterior orientation problem also received a lot attention in computer vision, with the term camera pose estimation usually being used for the task of finding camera location and its angle orientation. Computer vision approach usually search for direct solution of pose estimation problem, trying to find a solution with a minimum amount information about the object. The problem is usually solved in homogeneous coordinates that provides linear solutions based on concepts of algebraic projective geometry.

Methods for camera localization and pose estimation usually exploit feature detectors such as Local Binary Patterns(LBP) and its modifications (Ojala et al., 1996, Knyaz et al., 2016), Speeded Up Robust Features (SURF) (Bay et al., 2006) or Scale-Invariant Feature Transform (SIFT) (Lowe, 1999) for detecting control points, with further solving so-called Perspective-n-Point (PnP) problem of estimating the pose of a camera given a set of n 3D control object points and their corresponding 2D image points (Wu et al., 2018).

With impressive progress in deep learning these techniques were successfully applied for solving the camera pose estimation problem. The PoseNet (Kendall et al., 2015) was one of the first convolutional neural networks for estimating absolute pose of a camera from an image. It was modified truncated GoogLeNet (Szegedy et al., 2015) architecture with softmax classification layer replaced by a sequence of fully connected layers.

Bayesian PoseNet (Kendall and Cipolla, 2016), that improves the performance of PoseNet, leverages the notion of Bayesian CNNs with Bernoulli distributions. For input image the pretrained PoseNet model generates samples by dropping out activation units of convolutional layers with a given probability. Then the pose is calculated as average of the individual samples' predictions.

Further development of deep absolute pose estimation techniques can be divided into two groups: end-to-end learning of pose estimation (Walch et al., 2017, Moreau et al., 2022) and hybrid pose learning, such as image retrieval with relative pose regression, structure-based with local learning and hierarchical pose estimation (Shavit and Ferens, 2019).

2.2 Generative Modeling

Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) have been proposed recently. GANs aim learning a given domain of samples (e.g., images) at the training stage. During the inference unconditional GANs generate new samples that are indistinguishable from the original samples from the training dataset. GANs can be broadly divided into two large groups: unconditional GANs and conditional GANs.

3. METHOD

3.1 Framework overview

Our method aims estimating camera external orientation $p \in \mathbb{R}^6$ using an input image A. We make a strong assumption that the geometry of the scene observed in the image A is known during the training stage. We consider two domains: the image domain $\mathcal{A} \in \mathbb{R}^{W\times H\times C}$ and the external orientation parameters domain $\mathcal{P} \in \mathbb{R}^6$. A parameter vector $p \in \mathcal{P}$ defines the camera pose $p = [x, y, z, \alpha, \omega, \kappa]^T$ with respect to the scene reference frame.

3.2 ExteriorGAN method

Our `ExteriorGAN` model is inspired by `StyleGAN2` model (Karras et al., 2019). The original `StyleGAN2` model leverages a multilayer perceptron (MLP) to project the input random vector $z \sim \mathcal{N}(\mu, \sigma^2)$ to the latent space w, $M : z \rightarrow w$. We made two main contributions to the original `StyleGAN2` model. Firstly, we concatenate an input exterior orientation parameter vector p to the random vector z during the training stage. Secondly, we train an additional MLP that learns a inverse mapping $M^{-1} : w \rightarrow (z, p)$. During the inference stage, we obtain the latent vector w using embedding in the latent space. After that, we use an inverse MLP to project latent vector w back to the exterior orientation parameters p.

3.3 Dataset Generation

We use our environment simulator to generate our *ExteriorViews* dataset. The dataset includes 18k samples consisting of pairs of images and the corresponding exterior orientation parameters. The images present the virtual scene with objects of six classes: wall, floor, window, furniture, door, sculpture. The resolution of color images is 512 by 512 pixels. The dataset is split into a training set with 16k samples and a test set with 2k samples. We use various augmentation techniques to increase the dataset diversity (Kniaz et al., 2021). Example images from the dataset are presented in Figure 2.

4. EVALUATION

We evaluate our `ExteriorGAN` model and baselines using two datasets: our *ExteriorViews* dataset and the *LINEMOD* dataset (Hinterstoisser et al., 2012). The *LINEMOD* dataset includes over 18000 real images with 15 different objects and ground truth exterior orientation parameters. For both datasets, we use the training split with 15k images, we perform evaluation using the remaining test split.

We consider four baseline methods and neural models: `DLT` (Abdel-Aziz and Karara, 1971), `EPnP` (Lepetit et al., 2009), `DPOD` (Zakharov et al., 2019), `RCVPose` (Wu et al., 2022).

4.1 Qualitative Evaluation

We evaluate our `ExteriorGAN` model and baselines in terms of image reconstruction accuracy and comparison between the ground truth trajectory and the reconstructed trajectory. Random images generated by our `ExteriorGAN` model for the *LINEMOD* dataset are presented in Figure 3. Qualitative comparison with original images from the *LINEMOD* dataset proves that our model successfully reproduces the foreground scene. Still objects in the background, e.g. humans, computer monitors remain blurry.

Ground truth and reconstructed trajectory are presented in Figure 4 and Figure 5. The reconstructed trajectory is smooth and generally follows the original motion of the camera.

4.2 Quantitative Evaluation

We evaluate our model qualitatively in terms of Root Mean Squared Error (RMSE) at the ground truth points of the scene. The results of root mean squared error estimate for the proposed method and baselines are given in Table 1.

Also three metrics adopted from the Benchmark for 6D Object Pose Estimation (Hodaň et al., 2020, Hodaň et al., 2018) were used for the proposed method evaluation: Visible Surface Discrepancy (VSD) (Hodaň et al., 2018), Maximum Symmetry-Aware Surface Distance (MSSD), and Maximum Symmetry-Aware Projection Distance (MSPD). These metrics are defined as:

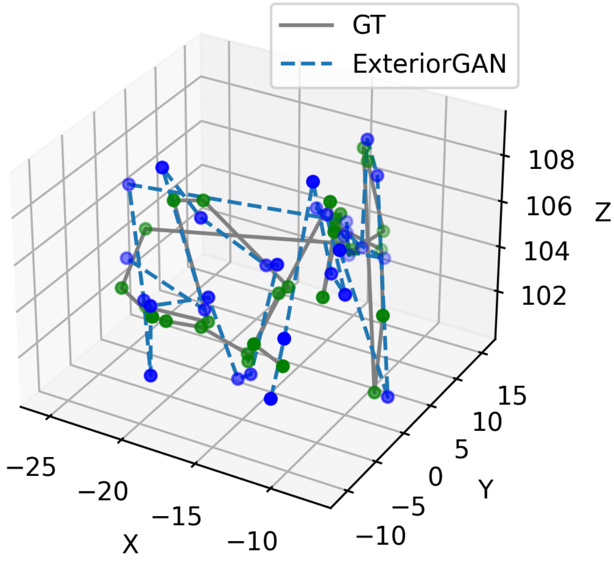

Figure 4. Qualitative comparison of ground-truth camera trajectory and trajectory estimated by our method for the first 20 points.

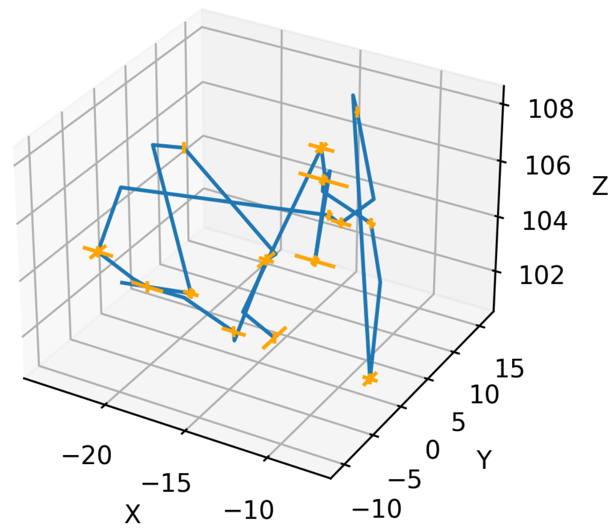

Figure 5. Error bars along X and Y axes for the first 30 points.

4.2.1 Visible Surface Discrepancy

Visible Surface Discrepancy treats poses that are indistinguishable in shape (color is not considered) as equivalent by measuring the misalignment of only the visible part of the object surface following:

$$e_{VSD}(\hat{D}, \bar{D}, \hat{V}, \bar{V}, \tau) =$$
$$avg_{p \in \hat{V} \cup \bar{V}} \begin{cases} 0 & if \quad p \in \hat{V} \cap \bar{V} \wedge |\hat{D}(p) - \bar{D}(p)| \\ 1 & \quad otherwise \end{cases}$$

The symbols $\hat{D}$ and $\bar{D}$ denote distance maps obtained by rendering the object model M in the estimated pose $\hat{\boldsymbol{P}}$ and the ground-truth pose $\bar{\boldsymbol{P}}$ respectively These distance maps are compared with the distance map D_I of the test image I to obtain the visibility masks $\hat{V}$ and $\bar{V}$, i.e. sets of pixels where the model

Method	**RMSE** mm	
	ExteriorViews	*LINEMOD*
`DLT`	64	101
`EPnP`	44	45
`DPOD`	39	48
`RCVPose`	**31**	**40**
`ExteriorGAN`	43	52

Table 1. Quantitative evaluation in terms of RMS error between estimated and ground truth exterior orientation parameters.

M is visible in the image I. The parameter τ is a misalignment tolerance.

4.2.2 Maximum Symmetry-Aware Surface Distance. The maximum distance between the model vertices is relevant for robotic manipulation, where the maximum surface deviation strongly indicates the chance of a successful grasp.

$$e_{MSSD}(\hat{\boldsymbol{P}}, \bar{\boldsymbol{P}}, S_M, V_M) = \\ min_{\boldsymbol{S} \in S_M} max_{\boldsymbol{x} \in V_M} ||\hat{\boldsymbol{P}} - \bar{\boldsymbol{P}}\boldsymbol{S}\boldsymbol{x}||_2$$

The set S_M contains global symmetry transformations of the object model M, identified as described in (Hodaň et al., 2020), and V_M is a set of the model vertices.

4.2.3 Maximum Symmetry-Aware Projection Distance considers global object symmetries and replaces the average by the maximum distance to increase robustness against the geometry and sampling of the object model.

$$e_{MSPD}(\hat{\boldsymbol{P}}, \bar{\boldsymbol{P}}, S_M, V_M) = \\ min_{\boldsymbol{S} \in S_M} max_{\boldsymbol{x} \in V_M} ||proj(\hat{\boldsymbol{P}}) - proj(\bar{\boldsymbol{P}}\boldsymbol{S}\boldsymbol{x})||_2$$

The function $proj(\cdot)$ is the 2D projection (the result is in pixels). The other symbols have the same meaning as in definition of Maximum Symmetry-Aware Surface Distance.

Table 2 shows the results of camera orientation estimates for the proposed method and baselines obtained on two considered datasets.

Method	*ExteriorViews*			*LINEMOD*		
	VSD	MSSD	MSPD	VSD	MSSD	MSPD
`DLT`	0.397	0.522	0.029	0.441	0.701	0.014
`EPnP`	0.481	0.501	0.020	0.455	0.701	0.034
`DPOD`	0.521	0.611	0.031	0.450	0.622	0.024
`RCVPose`	0.789	0.292	0.855	0.740	0.286	0.832
`ExteriorGAN`	0.492	0.512	0.029	0.481	0.519	0.017

Table 2. Quantitative evaluation in terms of Visible Surface Discrepancy (VSD), Maximum Symmetry-Aware Surface Distance (MSSD), and Maximum Symmetry-Aware Projection Distance (MSPD).

Tables 1 and 2 shows the `ExteriorGAN` model can estimate exterior orientation parameters of the camera with a mean error about 50 mm for a working space of 5× 5 meters and can perform at the level of the state-of-the-art methods, expressed in 6DOF metrics VSD, MSSD, MSPD.

5. CONCLUSION

We demonstrated that the latent space of a trained generative model can be used to encode the camera exterior orientation parameters with respect to a given scene. Furthermore, a mutual correspondence between a latent vector w and the camera pose p can be estimated. This allowed us to develop a method for estimating the camera exterior orientation parameters from a given image by embedding this image into the latent space. Specifically, we embed the input image A into the latent space to obtain the latent vector w. After that, we back-project the latent w into the camera exterior orientation parameters p using an inverse multilayer perceptron.

Our model estimates the exterior orientation parameters of the camera with a mean error of 50 mm for a working space of 5 by 5 meters. Evaluation proves that it is comparable with the modern methods. The main limitation of our model is that it can operate only in a constant environment. In further work, we are planning to estimate the camera pose with respect to a given object class. That will allow to estimate exterior orientation with respect to given objects in an arbitrary environment.

REFERENCES

Abdal, R., Qin, Y., Wonka, P., 2019. Image2stylegan: How to embed images into the stylegan latent space? *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, IEEE, 4431–4440.

Abdel-Aziz, Y. I., Karara, H. M., 1971. Direct Linear Transformation from Comparator Coordinates into Object Space Coordinates in Close-Range Photogrammetry. *Photogrammetric Engineering and Remote Sensing*, 81, 103-107.

Arun, K. S., Huang, T. S., Blostein, S. D., 1987. Least-Squares Fitting of Two 3-D Point Sets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9(5), 698-700. http://dblp.uni-trier.de/db/journals/pami/pami9.html.

Bay, H., Tuytelaars, T., Van Gool, L., 2006. Surf: Speeded up robust features. A. Leonardis, H. Bischof, A. Pinz (eds), *Computer Vision – ECCV 2006*, Springer Berlin Heidelberg, Berlin, Heidelberg, 404–417.

Brooks, T., Efros, A. A., 2022. Hallucinating pose-compatible scenes. S. Avidan, G. J. Brostow, M. Cissé, G. M. Farinella, T. Hassner (eds), *Computer Vision - ECCV 2022 - 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part XVI*, Lecture Notes in Computer Science, 13676, Springer, 510–528.

Dewitt, B., 1996. Initial approximations for the three-dimensional conformal coordinate transformation. *Photogrammetric Engineering and Remote Sensing*, 62, 79-83.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. C., Bengio, Y., 2014. Generative Adversarial Networks. *CoRR*.

Hinterstoisser, S., Lepetit, V., Ilic, S., Holzer, S., Bradski, G. R., Konolige, K., Navab, N., 2012. Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes. K. M. Lee, Y. Matsushita, J. M. Rehg, Z. Hu (eds), *Computer Vision - ACCV 2012 - 11th Asian Conference on Computer Vision, Daejeon, Korea, November 5-9, 2012, Revised Selected Papers, Part I*, Lecture Notes in Computer Science, 7724, Springer, 548–562.

Hodaň, T., Michel, F., Brachmann, E., Kehl, W., Glent Buch, A., Kraft, D., Drost, B., Vidal, J., Ihrke, S., Zabulis, X., Sahin, C., Manhardt, F., Tombari, F., Kim, T.-K., Matas, J., Rother, C., 2018. BOP: Benchmark for 6D Object Pose Estimation. *European Conference on Computer Vision (ECCV)*.

Hodaň, T., Sundermeyer, M., Drost, B., Labbé, Y., Brachmann, E., Michel, F., Rother, C., Matas, J., 2020. BOP Challenge 2020 on 6D Object Localization. *European Conference on Computer Vision Workshops (ECCVW)*.

Horn, B. K., Hilden, H. M., Negahdaripour, S., 1988. Closed-form solution of absolute orientation using orthonormal matrices. *J. Opt. Soc. Am A*, 5(7), 1127–1135.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T., 2019. Analyzing and Improving the Image Quality of StyleGAN. *CoRR*, abs/1912.04958. http://arxiv.org/abs/1912.04958.

Kehl, W., Manhardt, F., Tombari, F., Ilic, S., Navab, N., 2017. SSD-6D: making rgb-based 3d detection and 6d pose estimation great again. *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, IEEE Computer Society, 1530–1538.

Kendall, A., Cipolla, R., 2016. Modelling uncertainty in deep learning for camera relocalization. *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 4762-4769.

Kendall, A., Grimes, M., Cipolla, R., 2015. Posenet: A convolutional network for real-time 6-dof camera relocalization. *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*, IEEE Computer Society, 2938–2946.

Kniaz, V., 2016. ROBUST VISION-BASED POSE ESTIMATION ALGORITHM FOR AN UAV WITH KNOWN GRAVITY VECTOR. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B5, 63-68.

Kniaz, V. V., Knyaz, V. A., Mizginov, V. A., Papazyan, A., Fomin, N., Grodzitsky, L., 2021. Adversarial dataset augmentation using reinforcement learning and 3d modeling. B. Kryzhanovsky, W. Dunin-Barkowski, V. Redko, Y. Tiumentsev (eds), *Advances in Neural Computation, Machine Learning, and Cognitive Research IV. NEUROINFORMATICS 2020. Studies in Computational Intelligence*, 925, Springer International Publishing, Cham.

Knyaz, V. A., Zheltov, S., Vishnyakov, B., 2016. Robust object tracking techniques for vision-based 3D motion analysis applications. *Proc.SPIE*, 9896, 9896 - 9896 - 9. https://doi.org/10.1117/12.2227879.

Kraus, K., 1997. *Photogrammetry*. Dümmler Verlag, Bonn.

Lepetit, V., Moreno-Noguer, F., Fua, P., 2009. An Accurate O(n) Solution to the PnP Problem. *Int J Comput Vis*, 81, 155–166. https://doi.org/10.1007/s11263-008-0152-6.

Li, X., Li, S. W., 2012. OBTAINING APPROXIMATE VALUES OF EXTERIOR ORIENTATION ELEMENTS OF MULTI-INTERSECTION IMAGES USING PARTICLE SWARM OPTIMIZATION. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX-B3, 35–40. https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XXXIX-B3/35/2012/.

Lowe, D. G., 1999. Object recognition from local scale-invariant features. *Proceedings of the seventh IEEE international conference on computer vision*, 2, Ieee, 1150–1157.

Moreau, A., Piasco, N., Tsishkou, D., Stanciulescu, B., de La Fortelle, A., 2022. Coordinet: Uncertainty-aware pose regressor for reliable vehicle localization. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2229–2238.

Nie, W., Karras, T., Garg, A., Debnath, S., Patney, A., Patel, A. B., Anandkumar, A., 2020. Semi-supervised stylegan for disentanglement learning. *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, Proceedings of Machine Learning Research, 119, PMLR, 7360–7369.

Ojala, T., Pietikäinen, M., Harwood, D., 1996. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1), 51-59. https://www.sciencedirect.com/science/article/pii/0031320395000674.

Shavit, Y., Ferens, R., 2019. Introduction to Camera Pose Estimation with Deep Learning. *CoRR*, abs/1907.05272. https://arxiv.org/abs/1907.05272.

Slama, C., Theurer, C., Henriksen, S., 1980. *Manual of Photogrammetry*. American Society of Photogrammetry, Falls Church, Va., 1980.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S. E., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, IEEE Computer Society, 1–9.

Van den Heuvel, F. A., 1998. Vanishing point detection for architectural photogrammetry. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXII(5), 652-659.

Verykokou, S., Ioannidis, C., 2016a. EXTERIOR ORIENTATION ESTIMATION OF OBLIQUE AERIAL IMAGERY USING VANISHING POINTS. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 123-130.

Verykokou, S., Ioannidis, C., 2016b. EXTERIOR ORIENTATION ESTIMATION OF OBLIQUE AERIAL IMAGERY USING VANISHING POINTS. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B3, 123–130. https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLI-B3/123/2016/.

Walch, F., Hazirbas, C., Leal-Taixe, L., Sattler, T., Hilsenbeck, S., Cremers, D., 2017. Image-based localization using lstms for structured feature correlation. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

Wu, Y., Tang, F., Li, H., 2018. Image-based camera localization: an overview. *Vis. Comput. Ind. Biomed. Art*, 1(1). https://doi.org/10.1186/s42492-018-0008-z.

Wu, Y., Zand, M., Etemad, A., Greenspan, M., 2022. Vote from the center: 6 dof pose estimation in rgb-d images by radial keypoint voting. *European Conference on Computer Vision (ECCV)*, Springer.

Yan, L., Wan, J., Sun, Y., Fan, S., Yan, Y., Chen, R., 2016. A Novel Absolute Orientation Method Using Local Similarities Representation. *ISPRS International Journal of Geo-Information*, 5(8). https://www.mdpi.com/2220-9964/5/8/135.

Zakharov, S., Shugurov, I., Ilic, S., 2019. DPOD: Dense 6D Pose Object Detector in RGB images. *CoRR*, abs/1902.11020. http://arxiv.org/abs/1902.11020.