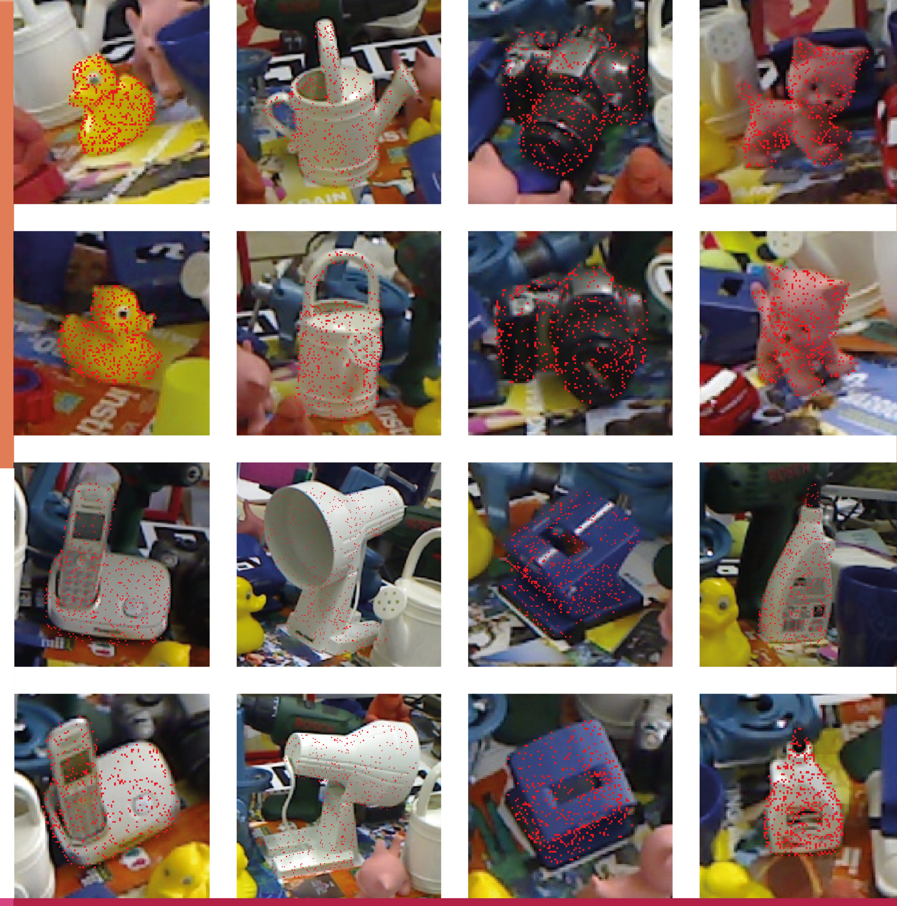
NOVALINCS
LABORATORY FOR COMPUTER SCIENCE AND INFORMATICS

26th Portuguese Conference on Pattern Recognition RECPAD



EXPLORING THE IMPACT OF COLOR SPACE IN 6D OBJECT PODE ESTIMATION

Nuno Pereira
Luís A. Alexandre
Universidade da Beira Interior
nuno.pereira@ubi.pt
luis.alexandre@ubi.pt

Objects	RGB	Grayscale	HSV	H	S	V
ape	74.29	86.67	**97.14**	67.62	34.29	82.86
bench vi.	99.03	**100.00**	99.03	88.35	89.32	99.03
camera	96.08	**98.04**	**98.04**	87.25	75.49	97.06
can	94.06	97.03	**98.02**	80.20	91.09	93.07
cat	**97.00**	95.00	**97.00**	81.00	86.00	**97.00**
driller	96.00	95.00	**99.00**	91.00	88.00	94.00
duck	62.26	93.40	**96.23**	51.89	35.85	88.68
eggbox	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
glue	**100.00**	99.03	**100.00**	**100.00**	**100.00**	**100.00**
hole p.	91.43	95.24	**99.05**	89.52	74.29	**99.05**
iron	75.26	97.94	95.88	94.85	91.75	**98.97**
lamp	**100.00**	**100.00**	98.08	95.19	97.12	**100.00**
phone	**100.00**	99.04	**100.00**	93.27	94.23	99.04
Average	91.17	96.63	**98.28**	86.08	81.14	96.03

Table 1: Results presented in this table were obtained through the training of MaskedFusion with its weights initialized as random. Italic names represent the symmetric objects. Bold values are the higher values in each line.

Objects	RGB	Grayscale	HSV	H	S	V
ape	88.57	90.48	96.19	92.38	**97.14**	94.29
bench vi.	**99.03**	97.09	**99.03**	97.09	**99.03**	**99.03**
camera	**99.02**	**99.02**	**99.02**	98.04	95.10	**99.02**
can	96.04	**98.02**	97.03	**98.02**	93.07	96.04
cat	99.00	99.00	**100.00**	99.00	**100.00**	99.00
driller	96.00	95.00	92.00	**98.00**	**98.00**	94.00
duck	95.28	91.51	94.34	**96.23**	93.40	91.51
eggbox	99.06	**100.00**	**100.00**	99.06	**100.00**	**100.00**
glue	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
hole p.	99.05	**100.00**	**100.00**	**100.00**	98.10	98.10
iron	97.94	96.91	**100.00**	97.94	93.81	95.88
lamp	99.04	99.04	99.04	**100.00**	98.08	**100.00**
phone	94.23	94.23	97.12	**98.08**	97.12	94.23
Average	97.08	96.93	**97.98**	**97.98**	97.16	97.01

Table 2: Results presented in this table were obtained by fine-tuning. Italic names represent the symmetric objects. Bold values are the higher values in each line.

Abstract

6D pose estimation is an open challenge due to complex world objects and many possible problems when capturing data from the real world, e.g., occlusions, and truncations. Getting the best input data to the deep learning methods is critical, for example, light can alter the features that these methods extract from the objects. Not obtaining accurate poses of the objects can lead to poor experiences in augmented reality scenarios or can lead to a fail grasping task of a robot. To try to avoid these issues, we investigate the impact of color spaces in 6D object pose estimation. For that, we evaluated RGB, Grayscale, HSV, and the HSV individual channels to study which color space would perform better in the 6D pose estimation task. We increased the accuracy of a method in 7.11% by using the HSV color space instead of the frequently used RGB.

Introduction

In computer vision, the ambient light can be a notable problem. It can create artifacts, alter the colors or cause shadows in the captured scene therefore constituting a problem in many computer vision algorithms. The RGB color space is widely used, although it does not represent the color as humans perceive it. If we want to isolate an object just using color in the image, it is hard to do in RGB because there may be many similar colors in the image. The HSV color space has three channels similar to RGB but instead of Red, Green, and Blue we have Hue, Saturation and Value, or intensity. The Hue channel represents the color. For example, red is a color but light red or dark red is not. The saturation channel is the amount of color present. It differentiates the pale red from the pure red. Finally, the value or intensity represents the brightness of the color, light red or dark red. So in the Hue channel, each color has its own value the entire red is a particular value. The lightness or darkness of the color does not affect the hue channel, so this channel is useful to extract specific colors from images. In real photographs, you will obtain varied saturation throughout the images depending on the intensity of the color present in them. The intensity channel shows the brightness of the colors.

Methodology

We use MaskedFusion as a framework for 6D pose estimation (object rotation matrix and translation vector from the viewpoint). MaskedFusion consists of three sub-tasks that executed sequentially estimates the 6D pose of an object presented in the scene. Initially, it uses a semantic segmentation method to detect and generate masks for each object presented in the scene. Then for each object segmented it crops the RGB image, depth image and mask. To eliminate the background around the object, a bit-wise and operation is made between the images and the mask. These segmented images are fed to a fully convolution neural network so it can regress the 6D pose of that object. After the preliminary pose is estimated, it is possible to utilize another method to refine the pose of the object. The method used in MaskedFusion is a neural network that enables it to be executed in real-time instead of other methods that are resource-heavy. In our experiments, we did not use the first sub-task of the MaskedFusion. Our primary goal is to report the impact of the different color spaces and/or channels in the 6D pose estimation.

To perform our tests, we choose to compare the HSV color space and each of its channels with the RGB color space. We tested MaskedFusion using the RGB, HSV, Grayscale, H (Hue), S (Saturation), and V (Value). We evaluated the MaskedFusion method in two independent roundups. In the first series of tests executed we trained the method from scratch, this means, the neural network presented in the method started with random weights, and we trained it for 150 epochs. In the second series of tests, we trained the method with RGB for 350 epochs. Furthermore, we saved the best performing weights in the validation set and use these weights to start fine-tuning the neural network for the other color channels. We fine-tuned the neural network for 150 epochs.

The computer used for all the tests have a SSD NVME, 64GB of RAM, an NVIDIA GeForce GTX 1080 Ti and Intel Core i7-7700K CPU.

Results

In Table 1 and 2, we present the results of MaskedFusion in the LineMOD test set. These results were calculated using the ADD and ADD-S metric.

In Table 1, we present the results where the MaskedFusion neural network was trained for 150 epochs with weights initialized as random values. In Table 1, its shown that the best performing color space is the HSV, as it performed higher on average. Specially for the first object, it achieved less error overall. HSV color space also achieved the best accuracy in 10 out of 13 objects.

In Table 2, we present the results for the executed tests with fine-tuning. The results presented were obtained by using the best-performed weights in the evaluation set during 350 epochs and then we used these weights to fine-tune the MaskedFusion for the other color spaces. Fine-tuning took 150 epochs and then we evaluate the method in the test set. On average the HSV color space and the Hue color channel had the lowest average error in the LineMOD dataset. Both of these colors had seven objects in which they performed higher than the other color spaces/channels.

During inference, MaskedFusion took an average 0.014 seconds to estimate the 6D pose of an object. With our experiments MaskedFusion took on average more 0.002 seconds to estimate the 6D pose of an object.

Conclusion

Sometimes using different color spaces aid in specific computer vision tasks. In these evaluations we discovered that using the HSV color space can help MaskedFusion achieve less error overall, if the same number of training epochs are used as when training using RGB images. Training MaskedFusion for 150 epochs from the random weights in the RGB color space we achieved on average 91.17% accuracy and using the same setup but only changing to HSV color space we achieved 98.28%, a substantial improvement. We suspect that the advantage of HSV over RGB can be even greater when pose estimation is performed in uncontrolled environments, since, LineMOD dataset uses a controlled light environment.