Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/15/20230835

Homography transform enhance CNN prediction accuracy on image classification

Chang Li

University of Illinois Urbana champaign, Lincoln Hall, 702 S Wright St, Urbana, IL 61801, USA

changli8@illinois.edu

Abstract. Convolutional Neural Network (CNN) image classification is a well-established algorithm that has been implemented in many fields. Benefit from the digitalization process and the exponential increase in the base of smart devices, this algorithm can be applied to even more traditional or casual contexts in the future driven by the trend of Internet of Things. Thus, the needs for optimizing image classification in specific domains may have turned out to be ongoing valuable research direction. This paper focuses on providing optimization under one example context which is using traffic signs as the experimental target. For the randomly selected traffic sign samples in the experiment, the accuracy obtained from the samples treated by the homography transform compared to control group passed all three statistical tests: Two Sample t-test, McNemar's test, and Fisher's exact test. Therefore, research direction has achieved small-scale validation and presents optimism for large-sample experiments and further research in optimization using the introduced strategy in the paper.

Keywords: CNN prediction, homography, image classification, autopilot.

1. Introduction

Machine learning algorithms have been utilized for image classification in various fields [1-4]. For instance, in agriculture, the work of Dingle Robertson and King, who employed the k-Nearest Neighbor (kNN) algorithm to classify various types of agricultural land cover using Landsat-5 TM imagery [5]. Furthermore, Zhang et al. conducted a study that aimed to classify crops in precision agriculture applications using hyperspectral imagery. The study utilized a Convolutional Neural Network (CNN) and achieved high accuracy in crop classification [6]. The authors concluded that CNNs hold potential in the field of crop classification for precision agriculture. This study highlights the effectiveness of deep learning algorithms in image classification tasks, especially in the agricultural sector where accurate classification can have significant economic and environmental impacts. On the other hand, autopilot and smart car have been very popular in the past few years, and the capital markets have given recognition by substantial investing, such as Tesla's stock price surge and the fund investment in Autopilot &smart cars by various large volume & established car manufacturing companies. In fact, engineers were already working on cameras in cars before autonomous driving became popular [7-9]. Tracing back to 1956 there were already applications for back-up cameras on cars [9]. However most rear-facing cameras installed in vehicles only provide perspective projection images, which has the incompetence of providing the distance between the vehicle and obstacles behind it. For example, Lin,

C. C., & Wang, M. S. in their paper have studied on using utilizing homography transform to warp bird's eye view captured by car camera; they called TVTM approach [8].

However, autopilot is still undergoing some obstacles. According to Liu, Y, current driverless cars can only support simple traffic conditions and that it is difficult to make correct judgments in more complex traffic situations, which could lead to accidents [7]. Under this context, the author foresees that car recognition of road conditions may have long-term optimizing needs. Autopilot is designed to collect information in real-time from the surrounding area while driving and turn it into data that matches traffic rules, just like a human driver, so correctly identifying traffic signs is a suitable entry point for the purpose of this paper.

The author finds that the homography transform share some characteristic that may help classic CNN image classification algorithm with its shortfall. Thus, this paper will apply this strategy by feeding CNN with homography transformed image and evaluate if it is a promising optimization related to autopilot technique.

2. Method

This paper mainly studies and explores whether using homography transform can directly improve the accuracy of the original classical CNN image classification.

2.1. Convolutional Neural Network

CNN is the most adopted algorithm on solving image classification problem. The crucial property of CNNs is that it automatically learns and extracts features from raw input data through a series of convolutional layers, pooling layers, and fully connected layers. The two main factors of this process are filters and pooling layers. The filter step for CNN is used to compress the input image matrix:

1) Define a window matrix with size less than the input image matrix, normally 5x5 or 7x7.
2) Fill the window matrix with certain value based on the needs for current filter.
3) Starting from the upper left corner of the input matrix, multiple the original matrix that is under the window matrix with the filter matrix result in one value.
4) Iteratively pivot the window matrix until all possible window cut of original matrix has met.

Based on the mechanism of the filter step, it is noticeable that while window pivots, the data at the edges of the matrix will be used significantly less times than the data in the middle ones. Consider the following sample diagram, black matrix is the 5x5 input image matrix and colored rectangles are the 3x3 window matrix. It is clear the upper left position is only used to calculate the upper left one box for the result by the red upper left window matrix (Figure 1). Any other pivots will not use upper left position.

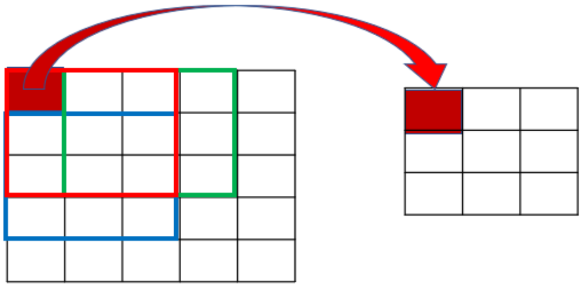


Figure 1. Filter for corner pixel.

On the contrary, the most middle position (center) of the input matrix will join the calculation by every pivot window matrix for this sample diagram. Therefore, every position of the result compressed matrix is related to the original center matrix (Figure 2).

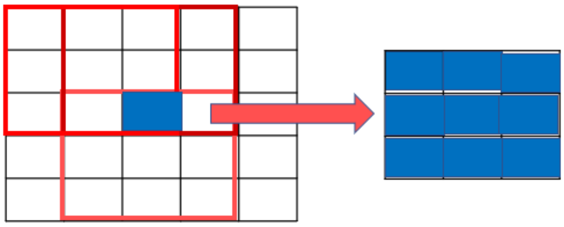


Figure 2. Filter for middle pixel.

From that observation, this paper deduces that after multiple filters when the model stacks up these layers together, the model would contain more information originally from the middle and less surrounding edges area of the input image matrix.

Although padding is an effective solution for this issue, the trade off to apply it may need be considered in the business operation aspect. Padding will expand extra pixels based on the size of filter and window size in order to offset the information loss after filters. However, this approach will increase computational cost and memory requirements of model and slow down the training speed. Even if the company manages to exclude the cost problem, it is also at risk of overfitting when the padding is too large.

Follow by this implication, the author states that it is possible to use homography transform to warp an image from a side angle to a front angle could make the CNN model better recognize the input since the front angle of a target provides more information around the center of the standard target image.

The author adopted a classic version of CNN filter strategy [10].

- The first convolutional layer has 10 filters of size 5x5 with a stride of 1 and a zero-padding of 'same'. The second convolutional layer has 40 filters of size 5x5 with a stride of 1 and a zero-padding of 'same'.
- After each convolutional layer, a ReLU activation function is applied to introduce non-linearity into the model. Then, a max pooling layer of size 2x2 with a stride of 2 is applied to reduce the spatial size of the feature maps.
- The output of the second max pooling layer is then flattened and fed into a fully connected layer with 500 neurons, which is again followed by a ReLU activation function.
- Finally, the output of the fully connected layer is passed through a softmax activation layer with the number of neurons equal to the number of classes in the problem.

2.2. Homography Transform

Homography transformations, also known as planar perspective transformations, is a fundamental concept in computer vision and image processing (Figure 3). This method serves the purpose of modeling the geometric relationships between two images of the same scene, taken from different viewpoints or with different camera configurations. Hartley and Zisserman, in their book Multiview Geometry in Computer Vision (2011), provide a definition of homographic transformations:

A projective transformation that maps corresponding points from one image to another can be described as a 3 by 3 matrix H [3].

$$[x', y' \; 1] = \mathrm{H}[x, y \; 1] \tag{1}$$

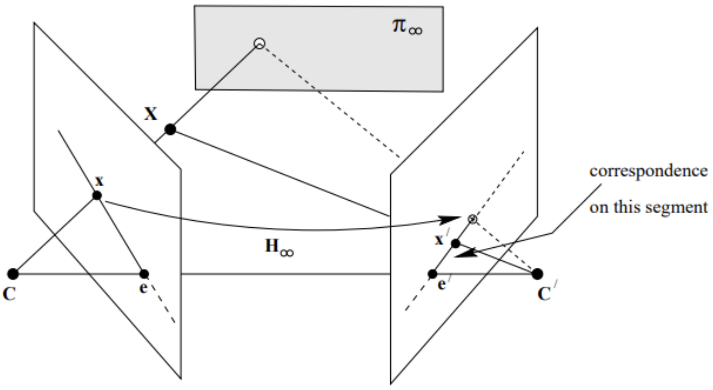


Figure 3. Scene planes and homographies [3].

2.2.1. Match points between image

SIFT: Scale Invariant Feature Transform, which is a feature point detection and matching method with good stability and invariance.

Gaussian kernel function: extract key points.

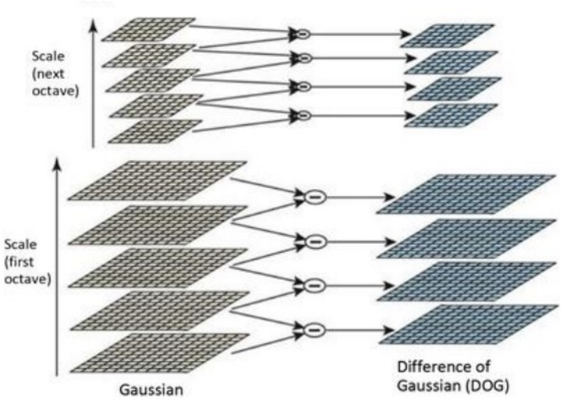


Figure 4. Gaussian pyramid and the dog pyramid.

This paper uses SIFT algorithm to extract feature points from the images, while the Gaussian kernel function is used to calculate the similarity between these points (Figure 4). It is an efficient approach to calculate the homography matrix between two images according to Wang, S who suggests that the SIFT algorithm works well for image matching and recognition, particularly when combined with other techniques such as Fast Library for Approximate Nearest Neighbors (FLANN) and Random sample consensus (RANSAC). The experimental data and analysis presented in the article demonstrate the effectiveness and accuracy of the SIFT algorithm in image points matching [7].

2.2.2. Calculate homography matrix

To find the homography matrix between two target images, just simply apply the Direct Linear Transformation where x and x' are the homogeneous coordinates of the points in the first and second images, respectively.

$$\boldsymbol{x'} = \boldsymbol{H} * \boldsymbol{x} \tag{2}$$

Solve for the homography matrix H by minimizing the reprojection error:

$$min||A * h - B||^2 \tag{3}$$

Normally it is suggested to use Singular Value Decomposition (SVD) to solve this equation.

3. Result

This section first presents the data set and relevant details, and then gives the analysis of the experimental results.

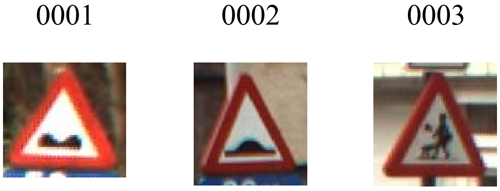


Figure 5. Traffic sign dataset.

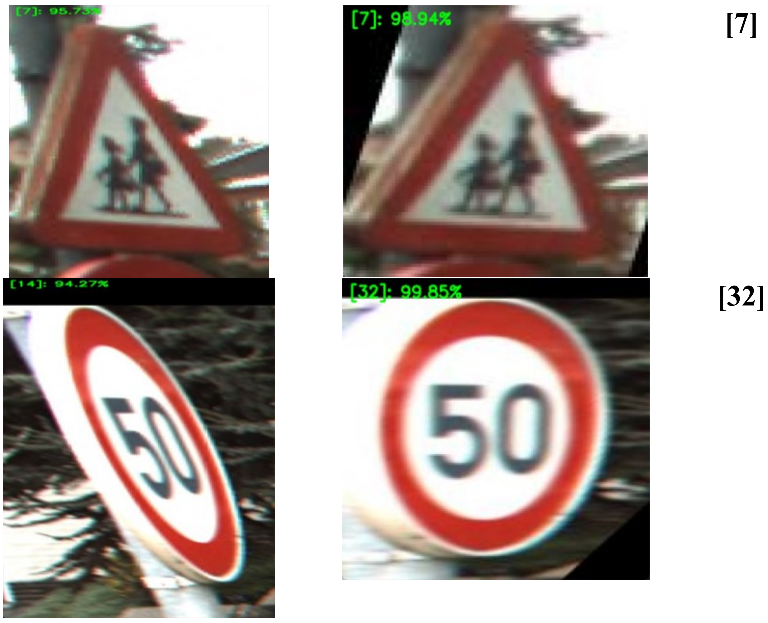
The author adopts a pre-categorized traffic sign dataset having around 1.5k images, the data source is described above where each folder with a unique number contains a certain category of traffic sign (Figure 5). For this experiment, the author inclines to sample images that don't have the front view of the sign which is not relevant with the paper argument and avoid inflation on accuracy. In general conditions, statistical random sampling will take 10 – 20 percent of the population.

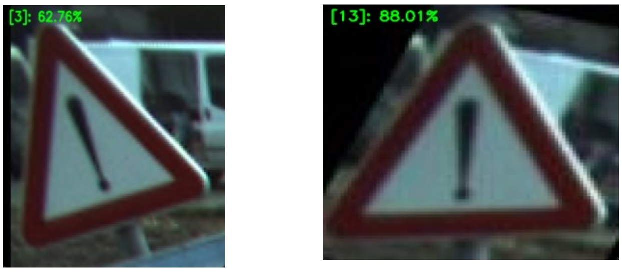
In addition, the author only considers traffic signs that has triangle or circle shape to control unintended interference factors when applying homography transform. After random sampling from each category, the test sample size in this paper is 60.

3.1. Analysis on the result

Images on figure 6 (Left) stand for the control group, and images on 6 (Right) stand for the homography transformed group.

Correct Answer





[13]

Figure 6. Control and homography transformed.

Author finds that the following situation will make the homography transform significantly helpful for classification accuracy:

Input image has an extreme camera angle toward the target, the second image on the left. Since this kind of image distributes useful information in areas with very few areas, it is more likely to mislead the CNN image classification in result of a wrong category. On the other hand, it is obvious that the homography transformed image (2th image) on the right has substantially flattened useful information in the middle, resulting in a correct classification. Input image has small size (less pixel) 3th image also tend to generate wrong classification by the algorithm since less base of pixels will cause low fault tolerance after applying multiple filter layers. Therefore, a small camera angle modification by holography transforms like the third image could correct a wrong classification.

The CNN classification result will generate two values:

Value in the represents the category predicted by the CNN model and the percentage value on the right is the confidence of that predicted by the model. Thus, the following passages will consider the probability of correctness. Observe that sample results by only applying the homography transformation for the target image could increase the certainty for classification when the original prediction is already high and correct the classification when the original prediction is uncertain and wrong.

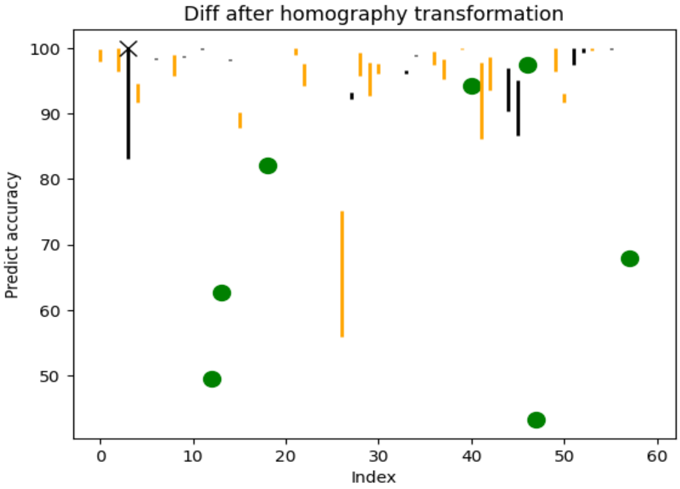


Figure 7. Diff after hormography transformation.

The y-axis is the probability that the CNN model predicts for the given traffic sign image belonging to a type of traffic sign. The x-axis is the index of each sample (Figure 7). For each index, the author draws a vertical line connect the two values for the same index represents the different of prediction accuracy before and after the homography warp of the given image. If the line is colored orange, it represents an increase in accuracy and black means a decrease.

There are two more situations demonstrated in this graph:

① The original prediction on classification is correct and the homography treated one is incorrect

② The original prediction on classification is incorrect and the homography treated one is correct

For context 1, the author marks a black cross sign on the value. For context 2, the author marks a green point.

In this graph, it is visible that majority of the homography transformed images gain reinforcement either on accuracy or even correct the original false prediction on category.

3.2. Statistical examine the result

In order to scientifically determine the effect of this strategy, the author in the following will use statical tests to examine. Since the sample size is > 50, the author will first assume the distribution for two outcomes are normal.

The null hypothesis (H0) is that there is no difference in probability (beta0 = beta1) between the two results, while the alternative hypothesis (H1) is that there is an improvement.

Consider there is prediction error in both groups, the author decides to assign zero accuracy value for incorrect prediction cases.

- Define the test statistic:

Author will use t-statistic:

$$t = \frac{mean_1 - mean_2}{\left(\frac{s}{n^{\frac{1}{2}}}\right)} \tag{4}$$

s is the pooled standard deviation of the two samples, calculated as:

$$s = \sqrt{\left\{\frac{\left((n_1-1)*s_1^2+(n_2-1)*s_2^2\right)}{n_1+n_2-2}\right\}} \tag{5}$$

- Set the significance level:

This paper will apply the most widely used 0.05 level of significance for all the statistical test in this paper.

- Compute the p-value:

Result of applying Welch Two Sample t-test is at follow (Table 1).

Table 1. Result of applying welch two sample t-test.

t	-2.0777
mean of control = 85.63583	mean of treated = 94.76383
p-value	0.0206
95 percent confidence interval	-Inf -1.810377

Since the p-value is less than 0.05 level of significance, reject the null hypothesis and adopts the alternative that the homography transform does have an improvement on the prediction accuracy. P-value less than half of the significance level and the 95 percent confidence interval has an upper bound -1.81 which is highly distanced from zero indicate a strong support for the conclusion. The only concern is that the lower bound for 95 percent confidence interval is negative infinite which is a sign of unstable for changes in data. Furthermore, if examine the normality on the sample (Figure 8).

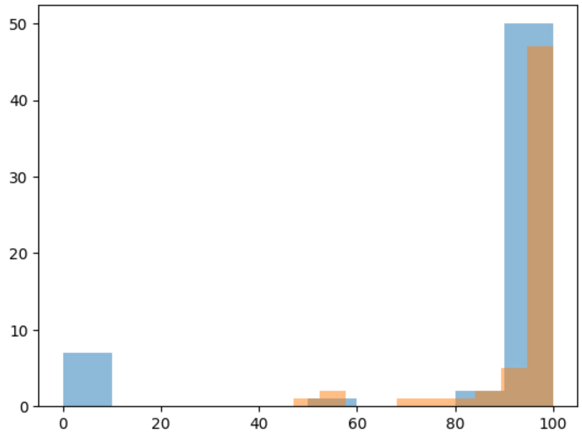


Figure 8. Frequency distribution.

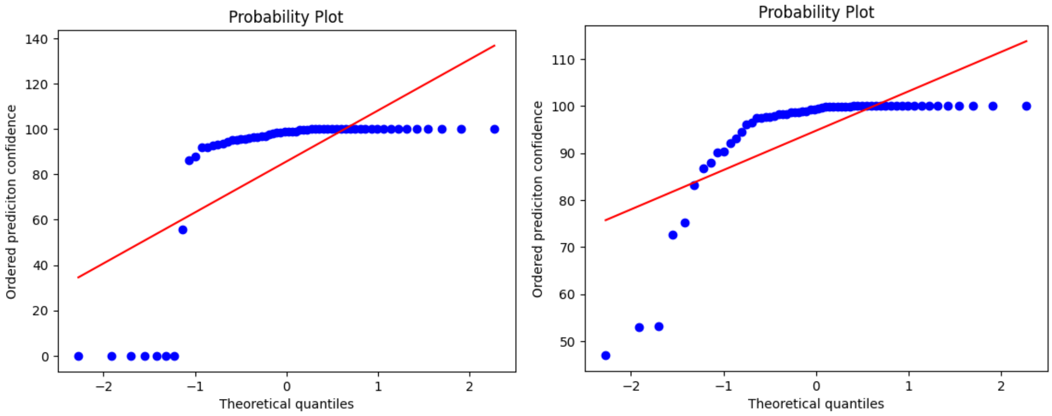


Figure 9. Q-Q plot for control and Q-Q Plot for treated.

Looking at the disparity of the sample, both box plot and Q-Q plot incline to imply that the sample is not likely to be normal distributed, though the homogrpahy prediction has some smooth effect at the lower tail (Figure 9).

Nevertheless, it does not necessarily mean that the sample is less trustworthy. CNN prediction commonly adopts a threshold strategy which makes a decision only when the confidence level passes a certain level. Therefore, it is by mechanism that small prediction confidence at the lower tail will more likely results in wrong prediction which counts to a zero.

This is not to say that the t-test is unreliable, but the author will apply further use additional statistical test to give multiple validation.

3.3. Contingency table test on correction

Since the data is in repeated measure and only two measurement points (raw prediction, homography prediciotn), McNemar's test is applicable which is a classic statistic approach to evaluate machine learning result (Table 2).

Table 2. McNemar's test.

Raw prediction /Homography prediction	Correct	Incorrect
Correct	51	7
Incorrect	1	1

Consider the sample size is not very large and the value in two position is 1, the author decides to apply the adjusted version (constant term 0.5) of McNemar's test. The null hypothesis H0 is that the homography transform is ineffective.

$$X^2 = \frac{(|B-C|-0.5)^2}{B+C} = \frac{(7-1)^2}{8} = 4.5 \tag{6}$$

P-value resulted from this statistic is 0.03389485 which is less than the 0.05 level of significance, therefore reject the null hypothesis and concludes that homography transform does have improvements on the prediction.

Table 3. Fisher's exact test.

Method	Correct	Incorrect
Raw prediction	52	8
Homography prediction	58	2

Author additionally uses Fisher's exact test, Dual Verification, to test proportions for correction are different among raw prediction and Homography prediction (Table 3). Fisher's exact test is favourable in this case because author provides a contingency table and the sample size is not large (Fisher's exact test is more accurate than the chi-square test or G–test of independence when the expected numbers are small.).

The null hypothesis H_0: θ = 1 versus HA: θ < 1, where θ is the odds ratio for raw prediction versus Homography prediction. H0 stands for there is no difference between the two methods and HA stands for Homography prediction is better than raw prediction.

Table 4. P-value test.

p-value	0.04731
95 percent confidence interval	(0.0000000 0.9827049)
sample estimates: odds ratio	0.2266694

According to the test result (Table 4), since p-value is less than the common 0.05 significance level, reject the null hypothesis and favor the alternative hypothesis that Homography prediction is better than raw prediction. It is 95% confident that the true odds ratio is between 0 and 0. 9827049 is less than 1. The sample estimates odds ratio indicates that the odds of being correct for Raw prediction is 0. 2266694 times that for homography prediction which is a significance enhancement for the test case.

Therefore, Fisher's exact test supports that Homography prediction is statistically significantly better than raw prediction on correction. Although the p-value for this test approximates the threshold of 0.05, it is reasonable to keep observe this effect since the more refined and complex the CNN algorithm is, the less significant difference will appear in this test for having small prediction error base.

4. Conclusion

This paper concludes that applying homography transform could have accuracy improvements on CNN image classification. For the sample introduced in the paper, applying homogrpahy transform definitely improves the accuracy of prediction by passing Two Sample t-test, McNemar's test, and Fisher's exact test. For the context of traffic sign classification, further validation could be done by increasing the sample size, testing CNN algorithms that have different layer strategy, or source the sample image from real auto-driving cars' camera. Although the sample size in this paper is not generally large to give a robust conclusion for large scale, it is still reasonable to consider that this method could be plausible under other contexts and worthwhile for further practices.

References

[1] Tanzeel U. Rehmana, Md. Sultan Mahmudb, Young K. Changb, Jian Jina, Jaemyung Shinb. Current and future applications of statistical machine learning algorithms for agricultural machine vision systems, 2018, Computers and Electronics in Agriculture, 156:585-605.

[2] Wang, Y., Yu, M., Jiang, G., Pan, Z., & Lin, J. Image Registration Algorithm Based on Convolutional Neural Network and Local Homography Transformation. 2020 Applied Sciences, 10(3):732.

[3] Hartley, R., & Zisserman, A. Multiple View Geometry in Computer Vision. 2011 Cambridge Core. https://doi.org/10.1017/CBO9780511811685

[4] Liu, Y. Analysis of Key Technical Problems in Internet of Vehicles and Autopilot. 2020, Advances in Intelligent Systems and Computing, 1-10.

[5] Dingle Robertson, L., King, D.J. Comparison of pixel-and object-based classification in land cover change mapping.2011 Int. J. Remote Sens. 32 (6), 1505–1529.

[6] Zhang, S., Wu, X., You, Z., Zhang, L., Leaf image-based cucumber disease recognition using sparse representation classification. 2017, Comput. Electron. Agric. 134, 135–141.

[7] Wang, S., Guo, Z., & Liu, Y. An Image Matching Method Based on SIFT Feature Extraction and FLANN Search Algorithm Improvement. 2021 Journal of Physics: Conference Series. 201-213.

[8] Lin, C. C., & Wang, M. S. A Vision Based Top-View Transformation Model for a Vehicle Parking Assistant. 2012, Sensors, 12(4):4431-4446.

[9] The Car and the Camera. (n.d.). Google Books. https://books.google.com/books/about/The_Car_and_the_Camera.html?hl

[10] Jaemyung Shinb. Spatio-Temporal Anomaly Detection for Industrial Robots through Prediction in Unsupervised Feature Space. (n.d.) 2021 Advances in Intelligent Systems and Computing, 122-134.