

PAPER • OPEN ACCESS

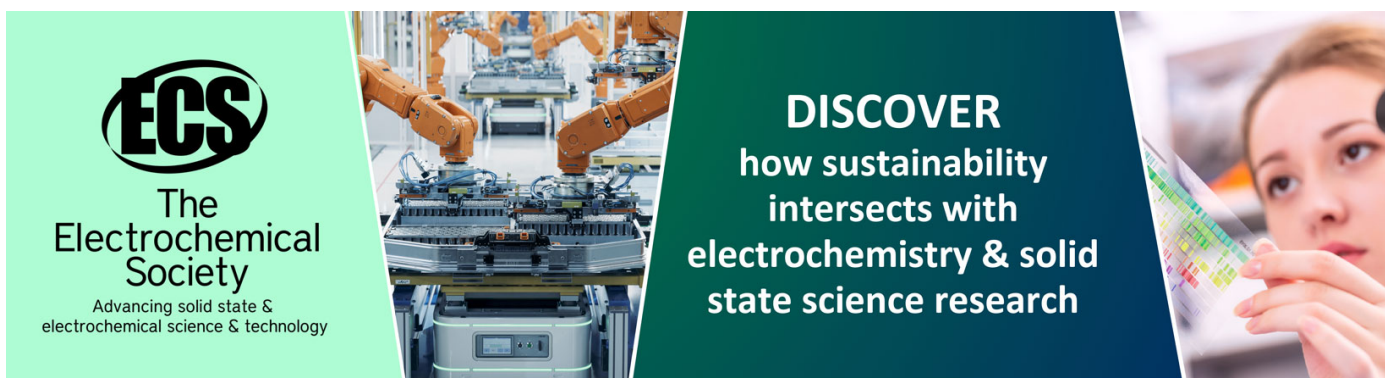
Fusion of Deep Feature and Hand-Crafted Features for Terrain Recognition

To cite this article: Xue Li *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **646** 012052

View the [article online](#) for updates and enhancements.

You may also like

- [A Local Update and Refinement method of the Global Terrain Nature Grid](#)
Hongbin Wang, Junwu Zhao, Yuanlin Liu et al.
- [Analysis of terrain navigability in underwater terrain aided navigation](#)
Miao Zhao, Zhenyang Xu and MA Weiqiang
- [Study on 3D Terrain Generation and Lightweighting Method of Highway](#)
Yinghao Chen, Leisheng Zhou and Jianping Xia



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Fusion of Deep Feature and Hand-Crafted Features for Terrain Recognition

Xue Li^a, Xin Ma^b and Peng Song

School of Control Science and Engineering, Shandong University, Jinan, China

^alx_sdu@163.com; ^bmaxin@sdu.edu.cn

Abstract. Terrain recognition is one of the key problems of mobile robots. It can help the robots understand the surrounding environment. With terrain prediction, the robots could realize autonomous navigation and path planning. This paper focuses on image feature selection for terrain recognition with visions. For terrain recognition tasks, feature is used to represent image information. Traditional visual features can be targeted to express the low-level information like color or texture. The deep feature is extracted by self-learning of neural network, containing richer semantic information than low-level features. There is a complementary relationship between the two. The efficiency and accuracy of terrain recognition is remarkably raised by the fusion of two above features. In the course of algorithm, combination of the off-line training model and on-line recognition model is used to identify the terrain type of the sample, which is to ensure the real-time performance. The corresponding terrain dataset--SDUterrain is established. The algorithm achieves 96% or higher classification accuracy in the experiments based on the SDUterrain Dataset, which is much higher than the single feature classification algorithm.

1. Introduction

Terrain recognition is an important research in the theory of autonomous navigation of mobile robot. Its ability to perceive the environment largely determines its ability to move autonomously in an unknown environment [1]. Visual sensors contain far more information than others and not affected by themselves. All the while, terrain recognition based on machine vision is a research hotspot of robot.

Image classification is a typical problem in machine learning. Feature is the key point. The quality of features largely determines the accuracy of classification. The complexity and variability of the field environment bring considerable challenges to terrain identification. Features which performs well in one environment is not suitable in the others. In this case, the selection of terrain features is particularly important.

Traditional machine learning relies on hand-crafted features heavily such as colors, textures, edges or shapes. Manduchi [2] proposed a color-based classification system, which recognizes the detected objects according to the preset terrain types. But color features is sensitive to illumination. Local binary patterns [3, 4] extracts the texture features from the grayscale image, which can eliminates the influence of illumination to some extent. Oliva [5] used spatial envelop feature to represent the content of images, and used Bayesian classifier to accomplish terrain recognition. Filitchkin and Byl [6] proposed the terrain classification method of SURF [7] combined with SVM [8], and apply it to a real quadruped robot-littledog, as the basis of gait planning and adjustment.



However, hand-crafted features only work in certain environment. This means that the generalization is insufficient in the unknown environment. The changing environment requires high descriptive and distinguishable features to represent the terrain image. It is not possible to do so solely on the hand-crafted features. In recent years, with the development of deep learning, it has become a new trend to complete terrain classification.

Hinton [9] first proposed the concept of deep network in 2006 and developed explosively in 2012 [10]. Deep learning has abandoned artificial design features. It can automatically learn more abstract feature at a higher level and describe the inner structure of abundant data. Deep learning shows its advantages in the terrain recognition in field environment. Liu [11] first introduced deep learning into scene terrain recognition, and proposed a terrain classification method based on deep sparse filtering network. Zhang [12] learned from [11], combining the deep convolutional neural network with the near distance learning strategy to improve the accuracy of terrain classification and make it more robust to the change of field environment. Xue [13] discusses the significance of imaging angle in terrain recognition, and proposes a differential angle imaging network, which uses both the original image and the differential image input into the convolutional neural network. And in [14], deep encoding pooling network for terrain recognition is proposed, which integrates random texture information and local spatial information in terrain images, outperforming the most advanced methods.

Moreover, the deep features in middle layers of neural network has also appeal to scholars. Donahue [15] extracted the deep features for terrain recognition in dynamic scene and aerial scene classification. Cimpoi [16, 17] improved the pooling layer and extracted the FV-CNN features. Results show that the deep features can integrate multi-scale information, and describe arbitrary shape and size of the region. It is suitable for material recognition and classification.

Inspired by [15]-[17], this paper proposed a terrain recognition method based on feature fusion. We extracted manual features, like color, LBP, and deep feature from trained convolutional neural network to represent terrain images. Finally inputting the fusion features into SVM to complete the terrain recognition.

2. Algorithm of terrain recognition

The proposed method in this paper is to solve the classification problem of several single flat terrain types with machine learning. The algorithm innovates in feature selection. We fused the hand-crafted features with deep features to improve the semantic information of the terrain features.

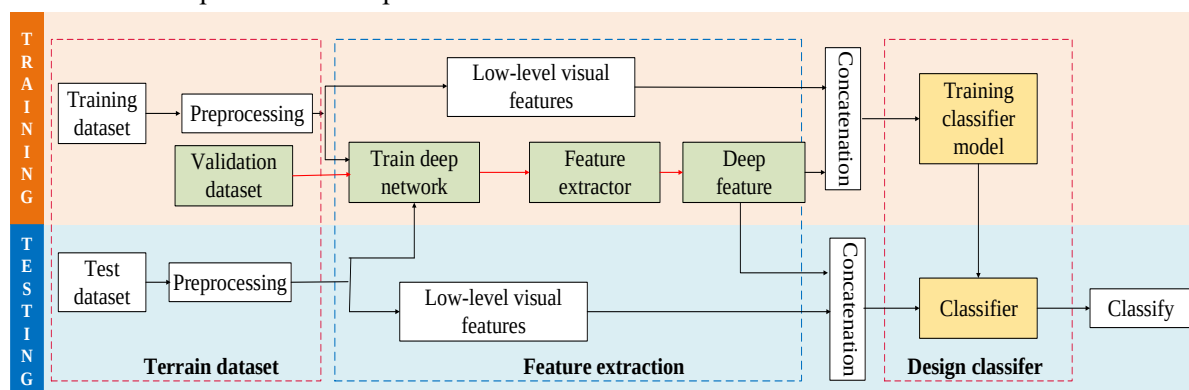


Figure 1. Terrain recognition algorithm based feature fusion.

There are two reasons for this. As mentioned earlier, traditional visual features are used to express low-level information such as colors, textures, etc. It magnifies some visual characteristics of the image which is easy to classify, but simultaneously leads to the problem that the features are not fully expressed. Only when the image is similar to the training image has excellent performance. It is difficult to distinguish all terrain types. The deep feature is extracted from the trained neural network. Its semantic information is more comprehensive than the low-level features, and the generalization is stronger. Not that the highly aggregated features is ideal. It isn't. Theoretically, deep features are

highly abstract features obtained from multi-layer convolution. The performance of deep features is proportional to the size of dataset and the layers of network. On the premise of limited dataset size, high classification accuracy can only be improved by increasing layers of the network. By doing so, it must lead to the loss of some low-level information. The low-level features of the neural network are usually the basic visual information such as the color and the edge of the image, which are commonly used in traditional machine learning. Considering the above analysis, there is a semantic complementary relationship between the two. After feature fusion, more perfect feature expression can be obtained. The flow of terrain recognition algorithm in this paper is shown in figure 1.

2.1. Hand-crafted features

RGB. There are color differences between terrain types. Color features obtain the color histogram vector by counting the proportion of each color component in the whole image. By counting the RGB values of all pixels in a terrain image, the 768-dimensional color histogram features can be obtained.

LBP. Terrain images also have distinct texture properties. LBP is a global statistical texture descriptor with low computational complexity and high speed. This feature is obtained from grayscale images, which can avoid the effect of illumination changing.

CEDD. CEDD [18] combines the color and texture information of the image. It can be seen as the fusion feature of RGB and LBP. Compared with the fusion feature, the dimension is lower, and the classification accuracy is similar.

We tested the above visual features in SDUterrain dataset, which will be introduced later in this paper.

Results are showed in table 1.

Table 1. Results of composite features.

	SVM(%)	Dimension	Time(s)
CEDD	76.53	144	28.78
RGB-LBP	79.16	1024	41.02

2.2. Deep feature

The highly distinguishable features can significantly improve the accuracy of classification. Visual features such as color and texture obtain high accuracy only if the training data is similar to the test data. Images captured in the actual scene is much more complex than images in the dataset. We hope to extract more comprehensive features than hand-crafted features. As we all know, deep learning takes the lead in the field of image classification. It can be applied in two ways. One is to directly provide end-to-end classification results. Another is used to extract deep features from the middle layers of the deep learning network. Inspired by this, we extract deep features for terrain recognition.

The advantage of deep learning is that it can realize the self-learning of large-scale data, and it can get rid of the disadvantage of human intervention in traditional visual feature extraction. The structure of deep network makes it possible to extract features in layers. The deep feature of full connected layer is generally the preferred method. However, the spatial structure information of the terrain image is lost because the high dimensional feature is mapped to one directly in the full connected layer. The pooling layer is another way to extract deep feature. As shown in table 2, results of deep features from pooling layer is better than full connected layer. Therefore, the deep feature of max pooling layer is used to express the terrain image in this paper. We choose ResNet-34 [19] to train the feature extractor.

Table 2. Results of deep features.

	SVM(%)	Dimension
Full connected layers	54.5	512
Max pooling layers	94.6	512

2.3. Feature fusion of deep feature and low-level features

Lots of experiments have proved that the features extracted by trained neural network are indeed contains more semantic information than traditional visual features. Generally speaking, the larger the dataset, the better the network performs. Besides, from the premise of limited datasets, we can increase the layers of neural network to improve the performance of network [20]. Another drawback of having too many network layer is the loss of lower layer information. These portions usually are color, texture and edges as mention above. Considering this problem, we will concatenate the deep feature with the hand-crafted feature to express the terrain image. The experiment results prove that the classification ability of the feature is improved to some extent.

3. SDUterrain dataset

Dataset is the most important part of image classification. There are few published dataset in direction of terrain recognition and cannot be directly applied in this paper. We constituted a terrain images dataset captured in actual environment. Sample images are collected by Canon IXUS 175 consumer camera and Logitech Pro C920 Webcam hardware. In the field environment, the surface of terrain types is often influenced by illumination, which will frequently lead to the color or texture change. The above factors (illumination, weather, season, angle, etc.) are considered as much as possible in the process of image collection. Typical terrain in the SDUterrain is divided into six classes: sand, mud, asphalt, grass, gravel and mulch that includes 13,200 images. Samples are shown in figure 2.

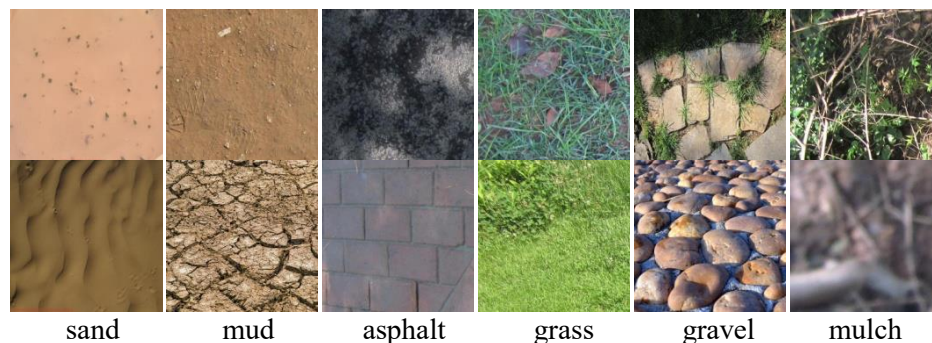


Figure 2. Samples of six types terrain in SDUterrain dataset.

4. Experiments

We evaluate the performance of our terrain algorithm on the SDUterrain dataset.

4.1. Dividing of the SDUterrain dataset

We divide the training dataset, the validation dataset, and the test dataset in proportion with 9:1:1. Images are collected under different environmental conditions. The number of images is not exactly the same under each condition, which should be extracted at random according to the proportion to avoid vary in quality. The training dataset is used to train the deep network parameters, and the validation dataset is used to verify the performance of the training network to avoid over-fitting.

4.2. Extracting of terrain features

We first extract the traditional visual features like color and texture features. Color histogram features are extracted from RGB space, each component is 256 dimensions, a total of 768 dimensions. LBP features are extracted from grayscale maps, which lost the color information. It eliminates the effect of illumination. The feature dimension is 256. CEDD features belong to color and texture composite features, feature dimension is 144, compared with the fusion feature of color histogram and LBP, it has lower dimension and higher computation speed. It is also an ideal visual feature in terrain recognition.

Then we extract the deep features. The residual network of 34 layers is used to train as a feature extractor. We replace the last fully connected layer with the max pooling layer. The dimension of the

deep feature is 512.

Last we fused the above feature in concatenation, as shown in figure 3. In this experiment, we choose two kinds of fusion feature, which are deep feature concatenated RGB color histogram feature and LBP texture feature, another is deep feature concatenated CEDD feature.

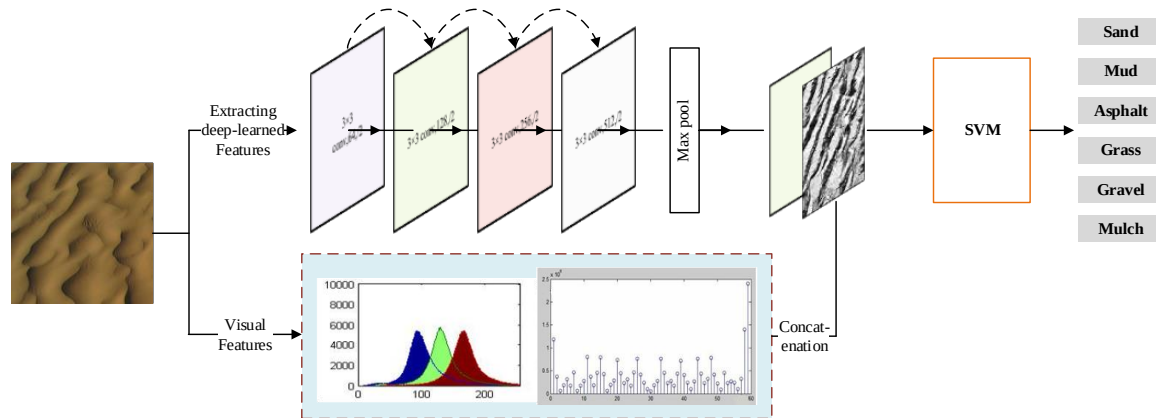


Figure 3. Low-level features and deep features extracted in terrain recognition algorithm.

4.3. Classification

Terrain features extracted from above process are inputted into the SVM for classification. Result is represented by confusion matrix. It is a kind of visualization matrix which expresses the performance of algorithm. It is usually used for supervised learning. The horizontal coordinate represents the true value of classification, and the vertical coordinate represents the expected value. The diagonal elements represent the classification accuracy of each type of terrain, and the average classification accuracy can be obtained by counting the diagonal elements. Figure 4(a) gives the average accuracy of 96.33% of the first fusion feature, and figure 4(b) gives the average accuracy of 97.00% of the other.

Sand	0.93	0.02	0.04	0.00	0.01	0.01
Mud	0.03	0.94	0.02	0.00	0.00	0.00
Asphalt	0.02	0.00	0.96	0.00	0.00	0.00
Grass	0.01	0.00	0.00	0.99	0.01	0.00
Gravel	0.00	0.00	0.00	0.01	0.98	0.01
Mulch	0.00	0.01	0.00	0.00	0.01	0.97
	Sand	Mud	Asphalt	Grass	Gravel	Mulch

Sand	0.93	0.02	0.04	0.00	0.01	0.00
Mud	0.03	0.96	0.01	0.00	0.00	0.01
Asphalt	0.01	0.01	0.97	0.00	0.00	0.00
Grass	0.01	0.00	0.00	0.99	0.01	0.00
Gravel	0.01	0.00	0.01	0.01	0.97	0.01
Mulch	0.00	0.01	0.00	0.00	0.01	0.97
	Sand	Mud	Asphalt	Grass	Gravel	Mulch

(a)Deep + RGB + LBP mean accuracy: 0.9633

(b)Deep + CEDD mean accuracy: 0.9700

Figure 4. Results based different features of the terrain recognition algorithm in SDUterrain dataset.

5. Conclusion

The terrain recognition algorithm in this paper combined the advantages of deep features and hand-crafted features. On the premise of limited dataset, we improve the performance of deep network by increasing the number of layers, and fuse the hand-crafted features to compensate the loss of low-level information. Compared with the method of traditional machine learning and deep learning, terrain recognition algorithm in this paper can further improve the classification accuracy.

Acknowledgments

This work was supported in part by the Chinese National Key Research and Development Plan (2018YFB1305803), Chinese National Natural Science Foundation (61673245), Chinese National Programs for High Technology Research and Development (2015AA042307).

References

- [1] Wang T H, Sung D G, Lin P C. Terrain classification, navigation, and gait selection in a leg-wheel transformable robot by using environmental RGBD information[C]. In Proceedings of the IEEE Conference on International Automatic Control Conference. 2018: 1-1.
- [2] Manduchi R, Castano A, Talukder A, et al. Obstacle detection and terrain classification for autonomous off-road navigation[J]. *Autonomous Robots*. 2005, 18(1): 81-102.
- [3] Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. 2002, 24(7): 971-987.
- [4] Mäenpää T, Pietikäinen M. Texture analysis with local binary patterns[M]. *Handbook of Pattern Recognition and Computer Vision*. 2005: 197-216.
- [5] Oliva A, Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope[J]. *International Journal of Computer Vision*. 2001, 42(3): 145-175.
- [6] Filitchkin P, Byl K. Feature-based terrain classification for littledog[C]. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. 2012: 1387-1392.
- [7] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*. 2004, 60(2): 91-110.
- [8] Cortes C, Vapnik V. Support-vector networks[J]. *Machine Learning*. 1995, 20(3): 273-297.
- [9] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. *Neural Computation*. 2006, 18(7): 1527-1554.
- [10] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]. *Advances in Neural Information Processing Systems*. 2012: 1097-1105.
- [11] Liu H, Min Q, Sun C, et al. Terrain classification with Polarimetric SAR based on Deep Sparse Filtering Network[C]. In Proceedings of the IEEE Conference on Geoscience and Remote Sensing Symposium. 2016: 64-67.
- [12] Zhang W, Chen Q, Zhang W, et al. Long-range terrain perception using convolutional neural networks[J]. *Neurocomputing*. 2018, 275: 781-787.
- [13] Xue J, Zhang H, Dana K, et al. Differential angular imaging for material recognition[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:6940-6949.
- [14] Xue J, Zhang H, Dana K. Deep texture manifold for ground terrain recognition[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 558-567.
- [15] Donahue J, Jia Y, Vinyals O, et al. Decaf: A deep convolutional activation feature for generic visual recognition[C]. In Proceedings of the International Conference on Machine Learning. 2014: 647-655.
- [16] Cimpoi M, Maji S, Vedaldi A. Deep filter banks for texture recognition and segmentation[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3828-3836.
- [17] Cimpoi M, Maji S, Kokkinos I, et al. Describing textures in the wild[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 3606-3613.
- [18] Zou Y, Chen W, Xie L, et al. Comparison of different approaches to visual terrain classification for outdoor mobile robots[J]. *Pattern Recognition Letters*. 2014, 38(38):54-62.
- [19] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
- [20] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1-9.