Combining Deep Feature and Handcrafted Features for Material Classification

Truong Phuc Anh
University of Information Technology
Ho Chi Minh, Vietnam
14520040@gm.uit.edu.vn

***Abstract*—Material classification is a challenging problem in robot and computer vision. The deep learning methods have achieved major success in object classification, but they does not conquer in material classification. One of the main reasons is different materials may yield very similar appearance. In this paper, we propose an new method combining deep feature with geometry and texture information as handcrafted local features to improve classification accuracy. Experimental results on the datasets such as GTOS and FMD show that our proposed method achieves the best performance among the methods.**

I. Introduction

This paper considers the problem of material classication whose goal is to classify an image belongs to one of the pre-dened material categories, such as brick, leaves, metal, glass, wood, marble, dirt and fabric. In recent years, material classification has become one of the active topics in robot and computer vision because of providing the detail of material information for variant applications such as Advanced Driver-Assistance Systems (ADAS) [1], robotic manipulation [2] and robotic navigation [3], [4].

In addition, material of surfaces contributes valuable information for computer to understand the whole image and interact with the physical world. For example, Figure 1 shows that with information about material which those bottles make of, computer could sort those bottles by weights, make a decision to choose which one is good to hold hot water or even know that it would be a risk to allow people bring a glass bottle which could be used as a weapon into a meeting between head of states.

Fig. 1: Bottles with similar shapes, are made of different materials which decides its physical properties, which could be extremely useful information in various situations.

Early studies are focused on using lab-based measurements to capture physical properties of object surface which are extremely helpful to predict the material. However, those method also limits their application due to the needed of special equipment. A different approach is image-based classification where surfaces are simply capture as single view images to train classifiers. In these methods, recognition is typically based more on context than physical properties of objects. Taking advantages of both approaches, in this paper, we suggest an image-based method with additional information about basic physical properties of object surface such as geometry and texture extracted as local features which similar to what human think when we try to predicting material of object from an image (see figure 2 and 3). The main idea is shown in figure 4.

Fig. 2: Human can easily predict the material of similar texture surfaces by using some of their basic geometry informations (like shape) [5].

Fig. 3: While the main object in both scenes can be considered as ”stone” because of its shape, we can also easily know which one is ”real stone” based on its texture [6].

The rest of the paper is organized as follows. In Sec 2, related works are presented. In Sec 3, the proposed method is describe. The experimental results are presented in Sec 4. Finally, Sec 5 concludes the paper.

II. Related Work

Some of existing techniques based on physical properties of object (lab-based) such as elasticity [7], water permeation [8]

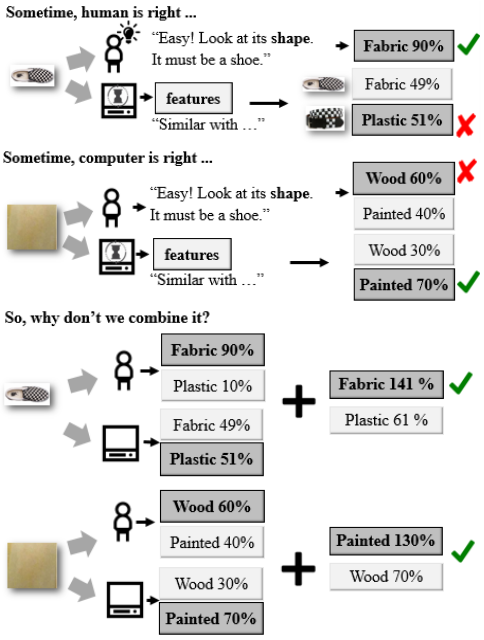

Fig. 4: We consider that deep feature extracted from a CNN can represent for the way human think and learn. Combine it with a suitable handcrafted feature would improve the prediction.

or optic response [9] give a high performance on distinguishing different types of similar materials because of rich information they provide. However, It generally requires a lot of special equipments to building those types of systems and datasets. Additional, with different approaches following this techniques we also usually need collect a new specialized dataset depends on which type of physical properties we want to get (Ex. GelFabric [10] - released in 2016 for Visual Perception on Fabrics Studies with color deep images, GelSight videos, and human labeling of the properties; Ground Terrain in Outdoor Scenes (GTOS) [4] - 2016 for multi-view points information; Reflectance Disk Database [11] - 2015 for reflectance of surfaces). Therefore, methods in this group can archive state-of-the-art on a specific dataset and also can easily fail on another (or can not be used on another because of missing informations).

On the other hand, alternative image-based techniques typically only require a natural RGB image and therefore can be set up easily on any dataset. Basing on the visual appearance and often replying on object classification to predict material, the main problem of these methods is lacking information and easily fooled by similar visual appearance such as shape.

Recently, features learned from Deep Neural networks has proven its valuable in object recognition and also transformed to the material field including both material classification and material segmentation. Bell et al., achieved per-pixel material category labeling by using a state-of-the-art object recognition network [12]. We have also been motivated by the two-stream fusion network which achieves state-of-the-art on GTOS dataset.

III. Method

As mentioned above, 2D-image base methods often replying on object classification to predict material and easy fooled when the objects have the same shape or texture but different material. We develop a method that combine deep features extracted from a CNN represent for current modern approach for object classification with handcrafted features includes edges and textures information to avoid miss classification samples have different material from similar objects. The main idea of our method is shown in figure 4.

We used the ImageNet pre-trained VGG-16 model as a deep feature extraction (from the "fc2" layer) and SVM for training classifier. The first input branch is original 2D-image, the second branch is edges image which is generated by using some common edges detector and the third branch is texture image which also generated by using texture filters.

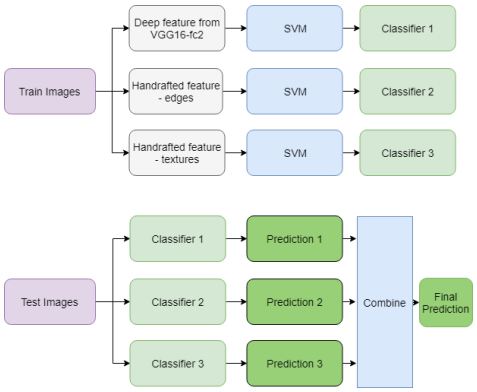

Fig. 5: Method 1: Combine probability predictions

In the first method, each branch will be trained particularly, then using three different classifiers to get probability predictions for each test sample. Finally, We combine the probability result from all branches with a simple averaging computation (see figure 5). Equation 1 shows how we combine predictions result from three classifier $P_1 = (p_1^1, p_2^1, ..., p_n^1)$, $P_2 = (p_1^2, p_2^2, ..., p_n^2)$ and $P_3 = (p_1^3, p_2^3, ..., p_n^3)$ with p_i^j is probability of class i th from classifier j th (n is number of classes).

$$P_{combined} = (\frac{p_1^1 + p_1^2 + p_1^3}{3}, ..., \frac{p_n^1 + p_n^2 + p_n^3}{3}) \quad (1)$$

The second method combines three branches after features have been extracted to get only one final feature vector for each sample, then use the combined features vector for training and predicting as usual. Features are combined by vector concatenation (see figure 6). The equation 2 shows how we combine features vector from different branch with f_i^j is the i th value in feature vector j th. m, n, p are the length of 1st, 2nd, 3rd feature vector.

$$F_{combined} = (f_1^1, f_2^1, ..., f_n^1, f_1^2, ..., f_m^2, f_1^3, ...f_p^3) \quad (2)$$

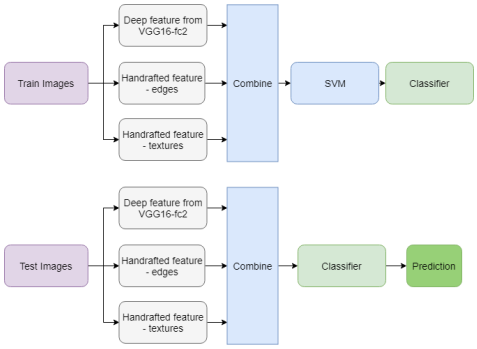

Fig. 6: Method 2: Combine features

The third method is a mixed architecture of two above. The original image will go by it own on the first branch and also combine its features with remain input images in the second branch. Final prediction would be an averaging combination of two branches. This approach gives the best performing in three methods.Both combination methods (for features combination, prediction combination) and variations of handcrafted features we may choose for particular task or dataset make the architecture more flexible and effective (see figure 7)..

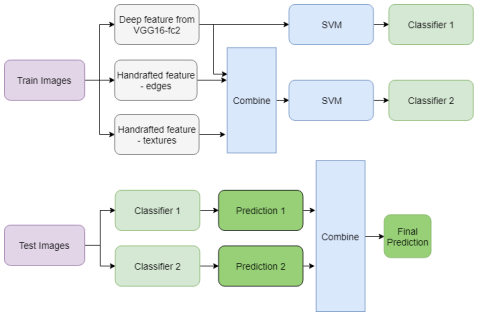

Fig. 7: Method 3: Mixed of method 1 and 2, the best performing architecture

IV. Experiments

We evaluate our method on newest material dataset Ground Terrain in Outdoor Scenes Dataset (GTOS) [4] and another common dataset from Flicker (Flicker Material Dataset - FMD) [13] and compare them with other state-of-the-art methods to see how handcrafted affect on different dataset. For each dataset, we firstly evaluate with different architectures to determine which method has best performance, then compare results with different features combination to show that suitable handcrafted features could be used to improve prediction result using deep feature from CNN.

Datasets: As mentioned above, our evaluation considers two datasets: GTOS and FMD. GTOS is the newest material dataset published in 2017 by the Rutgers ECE Vision Lab. Consisting over 34.000 images covering 40 classes of outdoor ground terrain under varying weather and lighting conditions, this is one of the most extensive dataset for the task. In this paper, we use pre-processed images (includes crop size, subtracting a per color channel mean and and normalizing for unit variance - for detail, you can find it at [4]) for training and testing. On the other side, FMD has been one of the most common dataset for material classification from the very beginning of the task (2009).

Method	Deep feature	Deep+edges	Deep+edges+textures
Combine prediction	75 ± 1.8	75.5 ± 3.0	76.3 ± 1.9
Combine features	75 ± 1.8	77.8 ± 2.5	79.5 ± 2.5
Mixed two above	75 ± 1.8	78.9 ± 2.2	$\mathbf{82.2 \pm 2.3}$

TABLE I: Evaluation on GTOS dataset. The table compares results on different method of combination and different features (also deep feature only vs deep feature combines with other handcrafted features).

Method	Deep feature	Deep+edges	Deep+edges+textures
Combine prediction	74.2 ± 1.4	75.1 ± 2.2	75.5 ± 2.3
Combine features	74.2 ± 1.4	75.3 ± 1.5	76.3 ± 1.7
Mixed two above	74.2 ± 1.4	75.5 ± 1.2	$\mathbf{77.2 \pm 0.9}$

TABLE II: Evaluation on FMD dataset

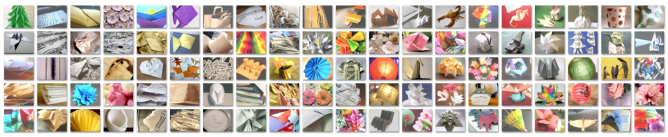

Fig. 8: Examples from FMD dataset, ensure a variety of illumination conditions, compositions, colors, texture and material sub-types

Note that the size of all images in both datasets is 224 x 224.

Evaluation measures: In particular, for both dataset, evaluation uses average classification accuracy.

Training process: We use splits from the previous research which assign 70% for training - 30% for testing on GTOS and these rates are 80% - 20% for FMD. First, from the original images, we extract edges with Canny edges detection and textures with local range texture filter. For deep feature, we use Keras framework (with Tensorflow as back-end) to extract deep feature from layer "fc2" of a VGG16 Image-Net pre-trained network. The training has been done with SVM from Scikit-learn frame work. The combination for features is vector concatenation and for final probability prediction is averaging.

Evaluation: The table 1 and 2 shows the mean accuracy (%) of three approaches mentioned in Method section on GTOS dataset.

V. Conclusion

In this paper, we investigate how handcrafted features can be used with deep feature learned from a trained CNN to improve material classification on 2D color image which often

Method	GTOS	FMD
DAIN [4]	81.2 ± 1.7	
Reflectance [11]		65.5
SIFT IFV+fc7 [14]		69.6 ± 0.3
Ours	**82.2 ± 2.3**	**77.2 ± 0.9**

TABLE III: Comparison with other methods on GTOS and FMD datasets

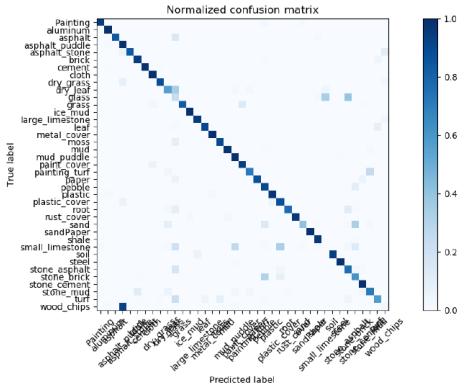

Fig. 9: Normalized confusion matrix of best result on GTOS dataset

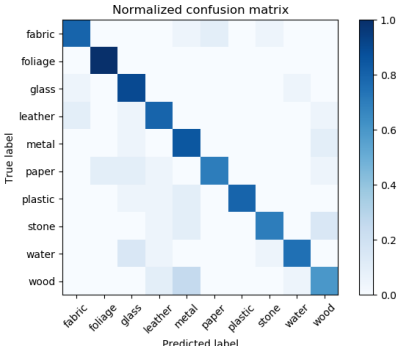

Fig. 10: Normalized confusion matrix of best result on FMD dataset

reply on object classification to predict material. We compare the performance of different architectures exploiting different types of input information as handcrafted features, and the results show that, the two streams architectures which combine both features and predictions is the best one for classification in GTOS and FMD.

Classes	Acc.	Classes	Acc.	Classes	Acc.	Classes	Acc.
painting	0.95	glass	0.83	painting turf	0.96	small limestone	0.87
aluminum	0.82	grass	0.98	paper	0.90	soil	0.77
asphalt	0.90	ice mud	0.84	pebble	0.91	steel	0.96
asphalt puddle	0.76	large limestone	0.95	plastic	0.89	stone asphalt	0.40
asphalt stone	0.60	leaf	0.99	plastic cover	0.98	stone brick	0.92
brick	0.99	metal cover	0.92	root	0.99	stone cement	0.97
cement	0.71	moss	0.88	rust cover	0.94	stone mud	0.97
cloth	0.57	mud	0.56	sand	0.71	turf	0.94
dry grass	0.93	mud puddle	0.23	sand paper	0.87	wood chips	0.98
dry leaf	0.95	paint cover	0.85	shale	0.97		

TABLE IV: Accuracy per class on GTOS dataset

REFERENCES

[1] H. Lay, "Toyota to add wrong way driving alert to navigation systems, autoguide.com news," May 2011. [Online]. Available: http://www.autoguide.com/auto-news/2011/05/toyota-to-add-wrong-way-driving-alert-to-navigation-systems.html

[2] M. W. Spong, S. Hutchinson, and M. Vidyasagar, *Robot modeling and control*. Wiley New York, 2006, vol. 3.

[3] J.-H. Kim, E. T. Matson, H. Myung, and P. Xu, *Robot Intelligence Technology and Applications 2012: An Edition of the Presented Papers from the 1st International Conference on Robot Intelligence Technology and Applications*. Springer Science & Business Media, 2013, vol. 208.

[4] J. Xue, H. Zhang, K. Dana, and K. Nishino, "Differential angular imaging for material recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 5, 2017.

[5] L. Sharan, C. Liu, R. Rosenholtz, and E. H. Adelson, "Recognizing materials using perceptually inspired features," *International journal of computer vision*, vol. 103, no. 3, pp. 348–371, 2013.

[6] P. Wieschollek and H. Lensch, "Transfer learning for material classification using convolutional networks," *arXiv preprint arXiv:1609.06188*, 2016.

[7] A. Davis, K. L. Bouman, J. G. Chen, M. Rubinstein, F. Durand, and W. T. Freeman, "Visual vibrometry: Estimating material properties from small motion in video," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5335–5343.

[8] P. Saponaro, S. Sorensen, A. Kolagunda, and C. Kambhamettu, "Material classification with thermal imagery." in *CVPR*, 2015, pp. 4649–4656.

[9] K. Tanaka, Y. Mukaigawa, T. Funatomi, H. Kubo, Y. Matsushita, and Y. Yagi, "Material classification using frequency-and depth-dependent time-of-flight distortion," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 79–88.

[10] W. Yuan, S. Wang, S. Dong, and E. Adelson, "Connecting look and feel: Associating the visual and tactile properties of physical materials," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR17), Honolulu, HI, USA*, 2017, pp. 21–26.

[11] H. Zhang, K. Dana, and K. Nishino, "Reflectance hashing for material recognition," in *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*. IEEE, 2015, pp. 3071–3080.

[12] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Advances in neural information processing systems*, 2014, pp. 568–576.

[13] L. Sharan, R. Rosenholtz, and E. Adelson, "Material perception: What can you see in a brief glance?" *Journal of Vision*, vol. 9, no. 8, pp. 784–784, 2009.

[14] S. Bell, P. Upchurch, N. Snavely, and K. Bala, "Material recognition in the wild with the materials in context database (supplemental material)," in *Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2015.