

MaterialNet: Multi-scale Texture Hierarchy and Multi-view Surface Reflectance for Material Type Recognition

Dongjin Lee¹, Hyun-Cheol Kim², Jeongil Seo², Seungkyu Lee¹

¹Department of Computer Science and Engineering Kyunghee University, ²Electronics and Telecommunications Research Institute

Motivation

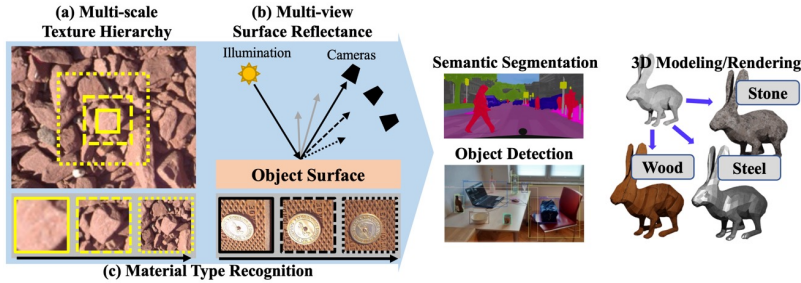


Figure 1. Multi-scale Texture Hierarchy and Multi-view Surface Reflectance

Certain material type is well categorized by its surface characteristics such as reflectance, stiffness, friction, roughness, and texture. While haptic properties are difficult to be estimated from visual data, texture could be easily observed from color image. Surface reflectance is another distinguishing property of a material that can be estimated from multiple viewpoint observations. We claim that texture features robust to environmental changes, their hierarchy along multiple scales (Figure 1. (a)), and surface reflectance (Figure 1. (b)) obtained from multi-view images can characterize material types comprehensively.

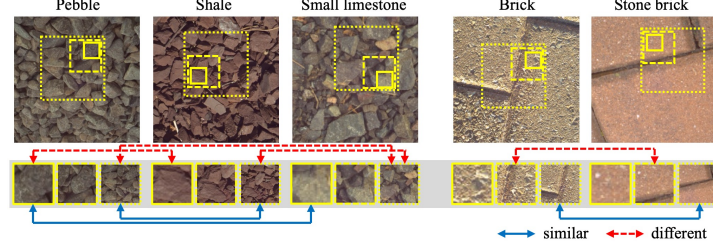


Figure 2. Samples of Multi-scale Texture Hierarchy



Figure 3. Samples of Multi-view Surface Reflectance

Experimental Results

Method	Backbone	Material (Single-color)					
		DTD[5]	KTH[3]	FMD[16]	MINC[2]	GTOS[22]	GTOS-mobile[23]
MAPNet[25]	VGGVD	74.10±0.6	82.70±1.5	82.90±0.9	NA	80.80±2.5	82.00±1.6
DSRNet[26]		74.90±0.7	83.50±1.5	84.00±0.8	NA	81.80±2.2	82.94±1.6
DeepTEN[28]	ResNet18	NA	NA	NA	NA	NA	76.12±x.x
DEPNet[23]		NA	NA	NA	NA	NA	82.18±x.x
MAPNet[25]		69.50±0.8	80.90±1.8	80.80±1.0	NA	80.30±2.6	82.98±1.6
DSRNet[26]		71.20 ±0.7	81.80±1.6	81.30±0.8	NA	81.00±2.1	83.65±1.5
CLASSNet[4]		71.50 ±0.4	85.40±1.1	82.50 ±0.7	80.50 ±0.6	84.30 ±2.2	85.25 ±1.3
FENet[21]		69.59±0.1	86.62 ±0.1	82.26±0.3	80.57 ±0.1	83.10±0.2	85.10 ±0.4
MSTH-Net		69.33 ±0.9	86.69 ±1.4	83.17 ±1.5	79.10 ±0.5	84.95 ±2.2	85.10 ±0.3
DeepTEN[28]		69.60±x.x	82.00±3.3	80.20±0.9	81.30±x.x	84.50±2.9	NA
DEPNet[23]		73.20±x.x	NA	NA	82.00±x.x	NA	NA
MAPNet[25]		76.10 ±0.6	84.50±1.3	85.20±0.7	NA	84.70±2.2	86.64 ±1.5
DSRNet[26]	ResNet50	77.60 ±0.6	85.90±1.3	86.00±0.8	NA	85.30±2.0	87.03 ±1.5
CLASSNet[4]		74.00±0.5	87.70±1.3	86.20 ±0.9	84.00 ±0.6	85.60±2.2	85.69±1.4
FENet[21]		74.20±0.1	88.24 ±0.2	86.74 ±0.2	83.98 ±0.1	85.71 ±0.1	85.20±0.4
MSTH-Net		71.45 ±0.6	87.72 ±1.0	85.65 ±1.4	81.47 ±0.6	85.73 ±2.6	87.45 ±0.8

Table 1. Single-color Material Recognition (MSTH-Net)

Input	Method	GTOS[22]	Input	Method	GTOS[22]
Multi-Color	CNN[24]	82.50±2.8	Multi-Color + Diff.	DAIN[22, 24]	86.20±2.5
	DEP[23, 24]	85.80±1.9		TEAN[24]	87.60±2.0
	MVSR-Net (9 views)	85.54 ±2.7		MVSR-Net (9 views)	86.65 ±2.3
	MaterialNet (4 views)	86.20 ±2.5		Dual MaterialNets (4 views)	87.84 ±2.1
	MaterialNet (9 views)	86.71 ±2.1		Dual MaterialNets (9 views)	88.41 ±2.1

Table 2. Multi-color Material Recognition (MSVR-Net and MaterialNet)

MSTH-Net		GTOS	FMD	KTH
Attention	Hierarchy			
✓		83.82±2.3	82.34±1.5	85.31±1.7
	✓	84.64±2.1	82.78±1.6	84.93±2.1
✓	✓	84.95 ±2.2	83.17 ±1.5	86.69 ±1.4

Table 3. Ablation studies on the effectiveness of texture attention and texture hierarchy

Proposed Methods

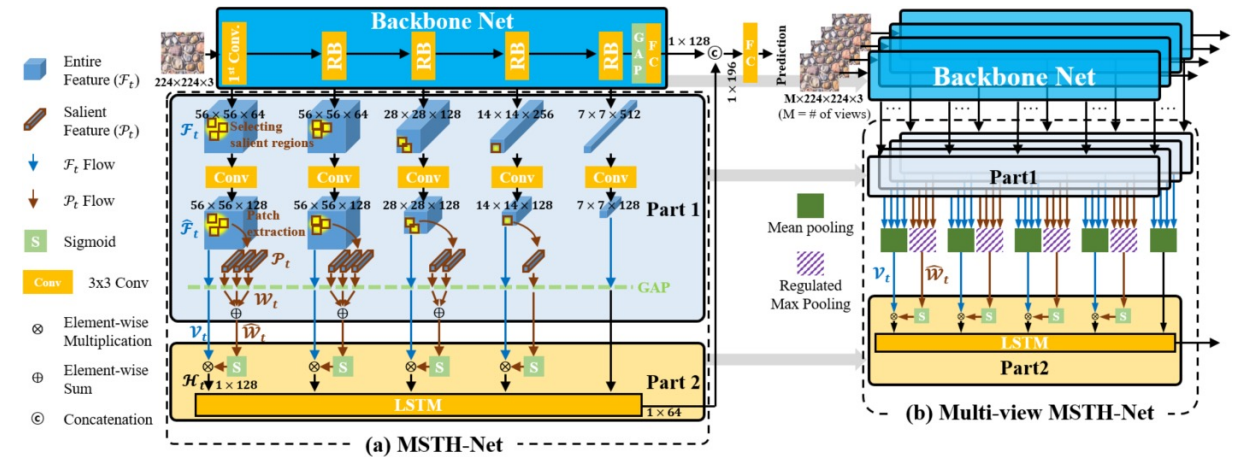


Figure 4. Multi-Scale Texture Hierarchy-Net (MSTH-Net)

- **MSTH-Net** encodes view-independent comprehensive multi-scale textures and their hierarchy.
 - **MSTH-Net Part1** : Take both entire and salient features from each layer.
 - **MSTH-Net part2** : Enhance the salient features and build texture hierarchy.
- For the multi-view environment, **multi-view MSTH-Net**(Figure 4. (b)) is constructed by collecting as many texture extractors (part 1) as the number of views MaterialNet accepts.

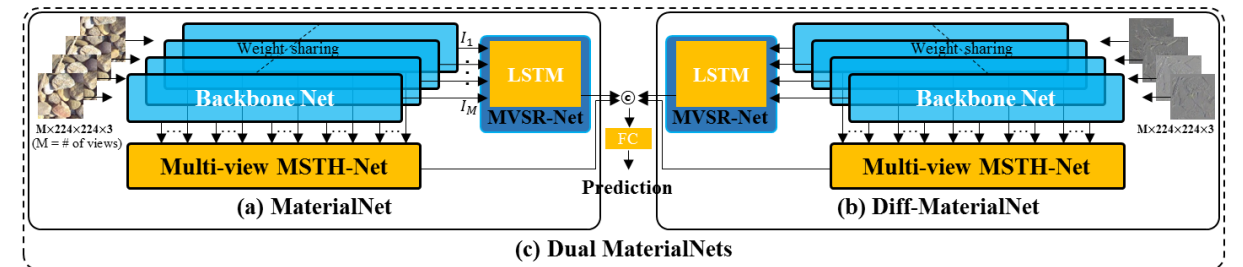


Figure 5. Multi-View Surface Reflectance-Net (MVSR-Net) and MaterialNets

- **MVSR-Net** encodes view-specific features revealing surface reflectance of a material type
- **Dual MaterialNet** (Figure 5. (c)) : combination of MaterialNet and diff-MaterialNet.
 - **MaterialNet** (Figure 5. (a)) : combination of MSTH-Net and MVSR-Net
 - **Diff-MaterialNet** (Figure 5. (b)) : network that has the same structure as MaterialNet, but receives a difference images of every two consecutive color images aligned by affine transformation before subtraction