Multispectral Imaging for Fine-Grained Recognition of Powders on Complex Backgrounds

Tiancheng Zhi, Bernardo R. Pires, Martial Hebert and Srinivasa G. Narasimhan
Carnegie Mellon University
{tzhi,bpires,hebert,srinivas}@cs.cmu.edu

Abstract

Hundreds of materials, such as drugs, explosives, makeup, food additives, are in the form of powder. Recognizing such powders is important for security checks, criminal identification, drug control, and quality assessment. However, powder recognition has drawn little attention in the computer vision community. Powders are hard to distinguish: they are amorphous, appear matte, have little color or texture variation and blend with surfaces they are deposited on in complex ways. To address these challenges, we present the first comprehensive dataset and approach for powder recognition using multi-spectral imaging. By using Shortwave Infrared (SWIR) multi-spectral imaging together with visible light (RGB) and Near Infrared (NIR), powders can be discriminated with reasonable accuracy. We present a method to select discriminative spectral bands to significantly reduce acquisition time while improving recognition accuracy. We propose a blending model to synthesize images of powders of various thickness deposited on a wide range of surfaces. Incorporating band selection and image synthesis, we conduct fine-grained recognition of 100 powders on complex backgrounds, and achieve 60%~70% accuracy on recognition with known powder location, and over 40% mean IoU without known location.

1. Introduction

In the influential paper "on seeing stuff" [1], Adelson argues about the importance of recognizing materials that are ubiquitous around us. The paper explains how humans visually perceive materials using a combination of many factors including shape, texture, shading, context, lighting, configuration and habits. This has since lead to many computer vision approaches to recognize materials [3, 10, 17, 32, 39, 41, 44, 45]. Similarly, this work has inspired methods for fine-grained recognition of "things" [2, 18, 22, 26, 40, 42] that exhibit subtle appearance variations, which only field experts could achieve before.
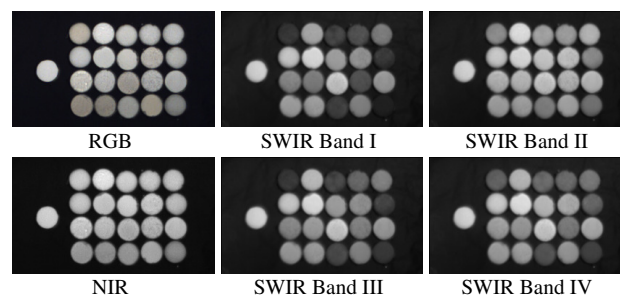

Figure 1. White powders that are not distinguishable in visible light (RGB) and Near Infrared (NIR) show significantly different appearances in Shortwave Infrared (SWIR). The leftmost sample is a white patch for white balance while the others are powders. Row 1 (left to right): Cream of Rice, Baking Soda, Borax Detergent, Ajinomoto, Aspirin; Row 2: Iodized Salt, Talcum, Stevia, Sodium Alginate, Cane Sugar; Row 3: Corn Starch, Cream of Tartar, Blackboard Chalk, Boric Acid, Smelly Foot Powder; Row 4: Fungicide, Calcium Carbonate, Vitamin C, Meringue, Citric Acid.

But there is a large class of materials — powders — that humans (even experts) cannot visually perceive without further testing by other sensory means (taste, smell, touch). We often wonder: "Is the dried red smudge ketchup or blood? Is the powder in this container sugar or salt?" In fact, hundreds of materials such as drugs, explosives, makeup, food or other chemicals are in the form of powder. It is important to detect and recognize such powders for security checks, drug control, criminal identification, and quality assessment. Despite their importance, however, powder recognition has received little attention in the computer vision community.

Visual powder recognition is challenging for many reasons. Powders have deceptively simple appearances — they are amorphous and matte with little texture. Figure 1 shows 20 powders that exhibit little color or texture variation in the Visible (RGB, 400-700nm) or Near-Infrared (NIR, 700-1000nm) spectra but are very different chemically (food ingredients to poisonous cleaning supplies). Unlike materials like grass and asphalt, powders can be present anywhere (smudges on keyboards, kitchens, bathrooms, out-

doors, etc.) and hence scene context is of little use for accurate recognition. To make matters worse, powders can be deposited on other surfaces with various thicknesses (and hence, translucencies), ranging from a smudge to a heap. Capturing such data is not only time consuming but also consumes powders and degrades surfaces.

We present the first comprehensive dataset and approach for powder recognition using multispectral imaging. We show that a broad range of spectral wavelengths (from visible RGB to Short-Wave Infrared: 400-1700nm) can discriminate powders with reasonable accuracy. For example, Figure 1 shows that SWIR (1000-1700nm) can discriminate powders with little color information in RGB or NIR spectra. While hyperspectral imaging can provide hundreds of spectral bands, this results in challenges related to acquisition, storage and computation, especially in time-sensitive applications. The high dimensionality also hurts the performance of machine learning [14] and hence recognition. We thus present a greedy band selection approach using nearest neighbor cross validation as the optimization score. This method significantly reduces acquisition time and improves recognition accuracy as compared to previous hyperspectral band selection approaches [6, 30].

Even with fewer spectral bands, data collection for powder recognition is hard because of the aforementioned variations in the thicknesses and the surfaces on which powders could be deposited. To overcome this challenge, we present a blending model to faithfully render powders of various thicknesses (and translucencies) against known background materials. The model assumes that thin powder appearance is a per-channel alpha blending between thick powder (no background is visible) and background, where α follows the Beer-Lambert law. This model can be deduced from the more accurate Kubelka-Munk model [23] via approximation, but with parameters that are practical to calibrate. The data rendered using this model is crucial to achieve strong recognition performance on real data.

Our multi-spectral dataset for powder recognition is captured using a co-located RGB-NIR-SWIR imaging system. While the RGB and NIR cameras (RGBN) are used as-is, the spectral response of the SWIR camera is controlled by two voltages. The wide-band nature of the SWIR spectral response (Figure 6) is more light efficient while retaining the discriminating ability of the traditional narrow-band hyper-spectral data [5, 43]. The dataset has two parts: *Patches* contains images of powders and common materials and *Scenes* contains images of real scenes with or without powder. For *Patches*, we imaged 100 thin and thick powders (food, colorants, skincare, dust, cleaning supplies, etc.) and 100 common materials (plastics, fabrics, wood, metal, paper, etc.) under different light sources. *Scenes* includes 256 cluttered backgrounds with or without powders on them. We incorporate band selection and data synthesis

in two recognition tasks: (1) 100-class powder classification when the location of the powder is known, achieving top-1 accuracy of 60%∼70% and (2) 101-class semantic segmentation (include background class) when the powder location is unknown, achieving mean IoU of over 40%.

2. Related Work

Powder Detection and Recognition: Terahertz imaging is used for the detection of powders [38], drugs [19, 20] and explosives [33]. Nelson *et al.* [29] uses SWIR hyperspectral imaging to detect threat materials and to decide whether a powder is edible. However, none of them studied on a large dataset with powders on various backgrounds.

Hyperspectral Band Selection: Band selection [6, 7, 12, 15, 27, 30, 37] is a common technique in remote sensing. MVPCA [6] maximizes variances, which is subject to noise. A rough set based method [30] assumes two samples can be separated by a set of bands only if they can be separated by one of the bands, which ignores the cross-band information.

Blending Model: Alpha Blending [31] is a linear model assuming all channels share the same transparency, which is not true for real powders. Physics based models [4, 13, 16, 23, 28, 35] usually include parameters hard to calibrate. The Kubelka-Munk model [23] models scattering media on background via a two-flux approach. However, it models absolute reflectances rather than intensities, requiring precise instruments for calibration and costing time.

3. RGBN-SWIR Powder Recognition Database

We build the first comprehensive RGBN-SWIR Multispectral Database for powder recognition. We first introduce the acquisition system in Section 3.1. In Section 3.2, we describe the dataset—*Patches* providing resources for image based rendering, and *Scenes* providing cluttered backgrounds with or without powder. To reduce the acquisition time, we present a band selection method in Section 3.3, and use selected bands to extend the dataset.

3.1. Image Acquisition System

The SWIR camera is a ChemImage DP-CF model [29], with a liquid crystal tunable filter set installed. The spectral transmittance (1000-1700nm) of the filter set is controlled by two voltages ($1.5V \leq V_0$, $V_1 \leq 4.5V$). We call each spectral setting a band or a channel, corresponding to a broad band spectrum (Figure 6). It takes 12min to scan the voltage space at 0.1V step to obtain a 961-band image. The 961 values of a pixel (or mean patch values) can be visualized as a 31×31 SWIR signature image on the 2D voltage space.

We co-locate the three cameras (RGB, NIR, SWIR) using beamsplitters (Figure 2), and register images via homography transformations. The setup is bulky to mount vertically, hence a target on a flat surface is imaged through
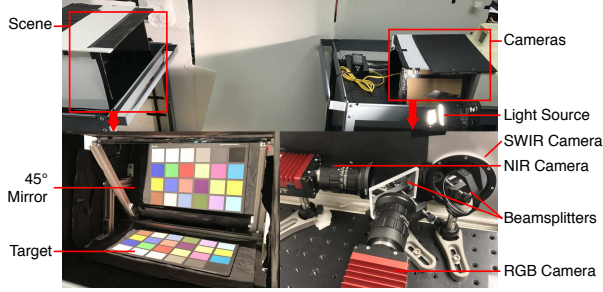
Figure 2. **Image Acquisition System**. RGB, NIR, and SWIR cameras are co-located using beamsplitters. The target is imaged through a $45°$ mirror.
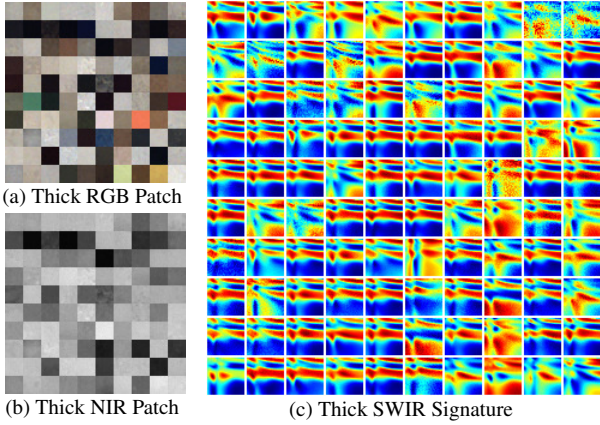

(a) Thick RGB Patch

(b) Thick NIR Patch

(c) Thick SWIR Signature

Figure 3. **Hundred powders**. Thick RGB patches, NIR patches and normalized SWIR signatures are shown.

a $45°$ mirror. A single light source is placed towards the mirror. We use 4 different light sources for training or validation (**Set A**), and 2 others for testing (**Set B**).

3.2. Patches and Scenes

The dataset includes two parts: **Patches** provides patches (size $14{\times}14$) to use for image based rendering; **Scenes** provides scenes (size $280{\times}160$) with or without powder. White balance is done with a white patch in each scene.

Patches (Table 1) includes 100 powders and 100 common materials that will be used to synthesize appearance on complex backgrounds. Powders are chosen from multiple common groups - food, colorants, skincare, dust, cleaning supplies, etc. Examples include Potato Starch (food), Cyan Toner (colorant), BB Powder (skincare), Beach Sand (dust), Tide Detergent (cleansing), and Urea (other). See supplementary for the full list. The RGBN images and SWIR signatures of the 100 powder patches are shown in Figure 3. Common materials (surfaces) on which the powders can be deposited include plastic, fabrics, wood, paper, metal, etc. All patches are imaged 4 times under different light sources (Set A). To study thin powder appearances, we also imaged thin powder samples on a constant background. As shown in Figure 4 (a), thick powders, thin powders, and a bare
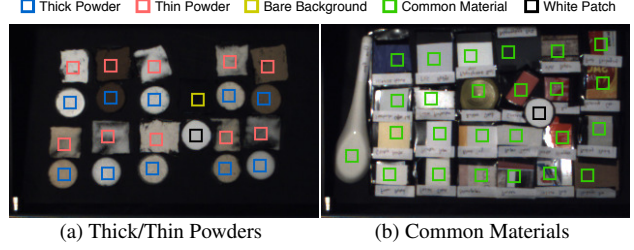

(a) Thick/Thin Powders   (b) Common Materials

Figure 4. **Patches** example. Thin powders are put on the same black background material. Patches are manually cropped for thick powders, thin powders, bare background, common materials, and white patch.

(a) Background Image   (b) Image with Powder   (c) GT Powder Mask

Figure 5. **Scenes** example. The ground truth mask is obtained by background subtraction and manual annotation.

Dataset ID	Target	Light Sources	Num Patches
Patch-thick	100 thick powders	Set A	400
Patch-thin	100 thin powders	Set A	400
Patch-common	100 common materials	Set A	400

Table 1. **Patches**. 100 thick and thin powders, and 100 common materials are imaged under light sources Set A.

Dataset ID	Light Sources	Num SWIR Bands	Num Scenes	N Powder Instances
Scene-bg	Set A	961	64	0
Scene-val	Set A	961	32	200
Scene-test	Set B	961	32	200
Scene-sl-train	Set A	34	64	400
Scene-sl-test	Set B	34	64	400

Table 2. **Scenes**. Each powder appears 12 times. *Scene-sl-train* and *Scene-sl-test* include bands selected by NNCV, Grid Sampling, MVPCA [6], and Rough Set [30].

background patch are captured in the same field of view.

Scenes (Table 2) includes cluttered backgrounds with or without powder. Ground truth powder masks are obtained via background subtraction and manual editing (Figure 5). Each powder in *Patches* appears 12 times in *Scenes*. In Table 2, scenes captured with light sources Set A are for training or validation, while the others are for testing. *Scene-bg* only has background images, while the others have both backgrounds and images with powder. *Scene-sl-train* and *Scene-sl-test* are larger datasets of scenes with powder that include only selected bands (explained in Section 3.3).

3.3. Nearest Neighbor Based Band Selection

Capturing all 961 bands costs 12min, forcing us to select a few bands for capturing a larger variation of powders/backgrounds. Band selection can be formulated as se-
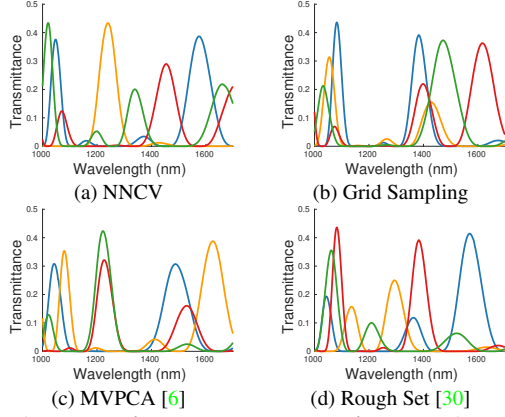
Figure 6. **Theoretical spectral transmittance** of 4 selected bands (different colors). NNCV has a good band coverage.

lecting a subset B_s from all bands B_a, optimizing a pre-defined score. We present a greedy method optimizing a Nearest Neighbor Cross Validation (NNCV) score. Let N_s be the number of bands to be selected. Starting from $B_s = \emptyset$, we apply the same selection procedure N_s times. In each iteration, we compute the NNCV score of $B_s \cup b$ for each band $b \notin B_s$. The band b maximizing the score is selected and added to B_s. Pseudocode is in supplementary.

To calculate the NNCV score, we compute the mean value of each patch in *Patch-thick* and *Patch-common* (Table 1) to build a dataset with 101 classes (background and 100 powders), and perform leave-one-out cross validation. Specifically, for each data point x in the database, we find its nearest neighbor $NN(x)$ in the database with x removed, and treat the class label of $NN(x)$ as the prediction of x. The score is the mean class accuracy.

The distance in nearest neighbor search is calculated on RGBN bands and SWIR bands in $B_s \cup b$. Because the number of SWIR bands changes during selection, after selecting 2 bands, we propose to compute cosine distances for RGBN and SWIR bands separately and use the mean value as the final distance. We call this the **Split Cosine Distance**.

We extend the *Scenes* dataset by capturing only the selected bands. *Scene-sl-train* and *Scene-sl-test* in Table 2 include 34 bands selected by 4 methods (9 bands per method, dropping duplicates): (1) NNCV (ours) as described above, (2) Grid Sampling uniformly samples the 2D voltage space, (3) MVPCA [6] maximizes band variances, and (4) Rough Set [30] optimizes a separability criterion based on rough set theory. See Figure 6 for theoretical spectral transmittances of the selected bands. Experiments in Section 5.2 and 6.2 will show that selecting 4 bands reduces acquisition time to 3s while also improving recognition accuracy.

4. The Beer-Lambert Blending Model

Powder appearance varies across different backgrounds and thicknesses. Even with fewer selected bands, capturing
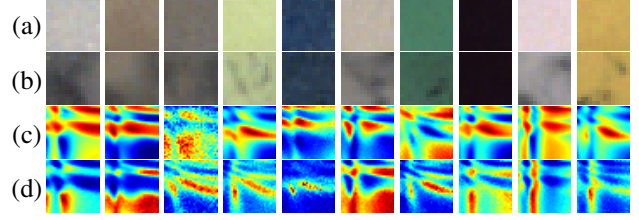

Figure 7. Examples of (a) thick powder RGB, (b) thin powder RGB, (c) SWIR signature, and (d) κ signature. The two signatures of many powders are negatively correlated.

such data is hard. Thus, we propose a simple yet effective blending model for data synthesis.

4.1. Model Description

The model is a per-channel alpha blending where α follows the Beer-Lambert law. Let I_c, A_c and B_c be the intensity of channel c of thin powder, infinitely thick powder (no background visible), and background, respectively. Let x be the powder thickness, and κ_c be the attenuation coefficient related to the powder rather than the background. Then:

$$I_c = (1 - e^{-\kappa_c x})A_c + e^{-\kappa_c x}B_c \tag{1}$$

Letting $\eta = e^{-x}$, the model can be rewritten as:

$$I_c = (1 - \eta^{\kappa_c})A_c + \eta^{\kappa_c}B_c \tag{2}$$

Equation 1 can be deduced as an approximation of the Kubelka-Munk model [23] (See supplementary material). The deduction indicates that κ is negatively correlated to A if the powder scattering coefficient is constant across channels. If we define the κ signature as a 31×31 image formed by the κ values of the 961 channels, similar to the SWIR signature defined in Section 3.1, the two signatures should show negative correlation if the scattering coefficient is constant across bands. In practice, 63% of the powders show a Pearson correlation less than -0.5. (Examples in Figure 7)

4.2. Parameter Calibration

The parameter κ_c can be calibrated by a simple procedure using a small constantly shaded thick powder patch, a thin powder patch, and a bare background patch. The calibration is done by calculating $\kappa_c x$ for each thin powder pixel, and normalizing it across pixels and channels (see Algorithm 1). Let P be the set of pixels in the thin powder patch, C_1 be the set of RGBN channels (RGB + NIR), and C_2 be the set of SWIR channels. Let $p \in P$ be a thin powder pixel and $c \in C_1 \cup C_2$ be a channel. Let $I_{p,c}$ be the thin powder intensity, and x_p be the powder thickness. Let A_c and B_c be the average intensity of the thick powder patch and the background patch. Then, we first compute $\kappa_c x_p = -\ln(\frac{I_{p,c} - A_c}{B_c - A_c})$ for each pixel $p \in P$ according to Equation 1. Then we calculate $\kappa_c median\{x_p\} = median_p\{\kappa_c x_p\}$,

Algorithm 1 Beer-Lambert Parameter Calibration

Input: Set of thin powder pixels P; Set of RGBN channels C_1; Set of SWIR channels C_2; Thin powder intensity $I_{p,c}$ of each pixel p and channel c; Mean thick powder intensity A_c; Mean background intensity B_c

Output: Attenuation coefficients κ_c for each channel c

   for each $c \in C_1 \cup C_2$ **do**
      for each $p \in P$ **do**
         $t_{p,c} \leftarrow -\ln\left(\frac{I_{p,c}-A_c}{B_c-A_c}\right)$ # compute $\kappa_c x_p$
      end for
      $\kappa_c \leftarrow median_{p \in P}\{t_{p,c}\}$ # compute $\kappa_c median\{x_p\}$
   end for
   $r \leftarrow \left(\frac{1}{|C_1|}\sum_{c \in C_1}\kappa_c + \frac{1}{|C_2|}\sum_{c \in C_2}\kappa_c\right)/2$
   for each $c \in C_1 \cup C_2$ **do**
      $\kappa_c \leftarrow \kappa_c/r$ # channel normalization
   end for

Blending	RMSE (mean±std)	
	RGBN	SWIR
Alpha	0.028±0.018	0.028±0.020
Beer-Lambert	**0.018±0.016**	**0.016±0.016**

Table 3. **Fitting Error on *Patch-thin*.** Beer-Lambert Blending shows a smaller error than Alpha Blending.

assuming κ_c is the same for each pixel. Since the scale of κ does not matter, we simply let $\kappa_c = \kappa_c median\{x_p\}$. To make κ_c be in a convenient range, we compute the mean κ_c values for RGBN and SWIR channels separately, and normalize κ_c by dividing it by the average of the two values.

We compare the fitting error of Beer-Lambert and Alpha Blending in Table 3. For a thin patch, we search for the best thickness for each pixel, and render the intensity using thick powder intensity, background, thickness and κ. We evaluate RMSE=$\sqrt{\frac{1}{nPixels \times nChannels}\sum\left(\frac{Rendered-Real}{WhitePatch}\right)^2}$ for each patch in *Patch-thin*. Table 3 shows that Beer-Lambert Blending fits better than Alpha Blending.

5. Recognition with Known Powder Location

To validate band selection and the blending model, we conduct a 100-class classification with known powder location (mask). We use nearest neighbor classifier to obtain thorough experimental results without long training times.

5.1. Nearest Neighbor in the Synthetic Dataset

As in Algorithm 2, we recognize each pixel in the mask by finding its nearest neighbor in a thin powder dataset rendered for that pixel, and vote for the majority prediction.

To build such a dataset, we estimate the background by inpainting the mask using fast marching [36], and render thin powders using Beer-Lambert Blending. Concretely, for each pixel p to be recognized, let I_p be its intensity, and B_p

(a) Scene   (b) Ground Truth BG   (c) Inpainting BG

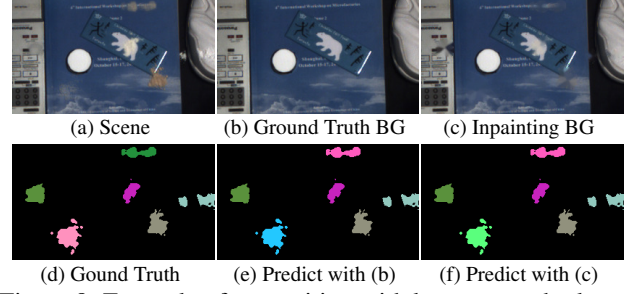(d) Gound Truth   (e) Predict with (b)   (f) Predict with (c)

Figure 8. Example of recognition with known powder location (powder mask) using ground truth and inpainting backgrounds. The results of two backgrounds are comparable.

be the intensity of the inpainting background. Let A be the mean pixel value of a thick powder patch from *Patch-thick* with calibrated κ. The channel subscript c is ignored. We iterate $\eta = 0.0, 0.1, ..., 0.9$ to render thin powder pixels of different thicknesses using Equation 2. We classify pixel p by finding the nearest neighbor for I_p with the Split Cosine Distance (Section 3.3) in the rendered dataset.

Algorithm 2 Recognition with Known Powder Mask

Input: Observed powder intensity I_p of each pixel p in the mask; Estimated background B_p

Output: Prediction $pred$

   $votes \leftarrow \emptyset$
   for each pixel p in the powder mask **do**
      $D \leftarrow \emptyset$
      for each patch $T \in$ *Patch-thick* dataset **do**
         $A \leftarrow$ mean value of T across pixels
         for $\eta = 0.0 : 0.1 : 0.9$ **do**
            $I' \leftarrow$ rendered thin powder intensity with A, B_p,
            η using Equation 2
            $D \leftarrow D \cup \{I'\}$
         end for
      end for
      $y \leftarrow$ the powder class of I_p's nearest neighbor in D
      $votes \leftarrow votes \cup \{y\}$
   end for
   $pred \leftarrow$ the mode value in $votes$

5.2. Experimental Results

We conduct experiments to analyze whether inpainting background, Beer-Lambert Blending, the three cameras, and the band selection are useful. We report the mean class accuracy on *Scene-val*, *Scene-test*, *Scene-sl-train* and *Scene-sl-test*, since the training data is from *Patches* only. If not specially stated, RGBN (RGB and NIR) bands, SWIR bands selected by NNCV, Beer-Lambert Blending, and inpainting background are used. This default setting achieves 60%∼70% top-1 accuracy, and about 90% top-7 accuracy.

Inpainting *vs*. Ground Truth Background: Table 4 and Figure 8 show similar performances of the inpainting background and the captured ground truth background.

Background	nSWIR	val	test	sl-train	sl-test
Ground Truth	961	62.5	60.0	-	-
Inpainting	961	63.5	59.5	-	-
Ground Truth	4	68.0	**65.5**	**63.00**	**63.50**
Inpainting	4	**72.0**	64.0	62.50	62.50

Table 4. **Inpainting** *vs.* **Ground Truth Background**. Inpainting does not significantly decrease performance.

Blending	nSWIR	val	test	sl-train	sl-test
No Blending	961	40.5	40.0	-	-
Alpha Blend	961	59.5	55.0	-	-
Beet-Lambert	961	63.5	59.5	-	-
No Blending	4	41.0	42.0	36.25	43.00
Alpha Blend	4	61.5	58.0	58.25	60.00
Beet-Lambert	4	**72.0**	**64.0**	**62.50**	**62.50**

Table 5. **Beer-Lambert** *vs.* **Alpha** *vs.* **No Blending**. Alpha Blending is better than No Blending, while Beer-Lambert Blending outperforms Alpha Blending.

RGB	NIR	SWIR	Scene-val	Scene-test
✓			20.5	18.0
✓	✓		31.5	29.0
		✓	28.0	30.5
	✓	✓	33.0	35.5
✓		✓	51.5	49.5
✓	✓	✓	**63.5**	**59.5**

Table 6. **Camera Ablation**. All 961 SWIR bands are used if "SWIR" is checked. Normal cosine distance are used for row 1∼4. All three cameras (RGB, NIR, SWIR) are useful.

Beer-Lambert *vs.* **Alpha** *vs.* **No Blending:** No blending means using thick powder intensity as the blended intensity. Table 5 shows that Beer-Lambert Blending is better.

Camera Ablation: All three cameras are useful (Table 6).

Band Selection: We compare NNCV with Grid Sampling, MVPCA [6], and Rough Set [30] in Figure 9. The performance of NNCV saturates after 4 SWIR bands, better than other methods and than using all 961 bands.

6. Recognition with Unknown Powder Mask

In real situations, the powder location is usually unknown. The algorithm should distinguish between backgrounds and powders, leading to a 101-class semantic segmentation task (background+100 powders). We train a deep net using synthetic data and limited real data for this task.

6.1. Synthesizing Powder against Background Data

Since real data are limited or hard to capture, we propose to render powder against background images. We synthesize a thick powder image with thickness map, and combine it with a real or synthetic background via Beer-Lambert Blending. We use the NYUDv2 [34] dataset and *Patches* for image based rendering. Illustration is in Figure 10.
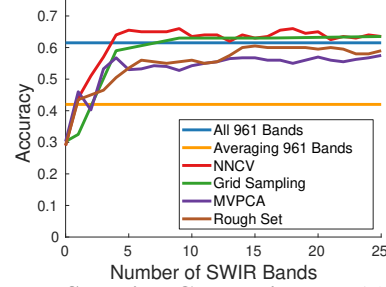

Figure 9. **Band Selection Comparison**. Grid Sampling is tested on square numbers only. We report the average results of *Scene-val* and *Scene-test*. See supplementary material for separate figures. The accuracy of NNCV saturates after four bands, outperforming the other methods. Selecting a few bands is even better than using all bands.
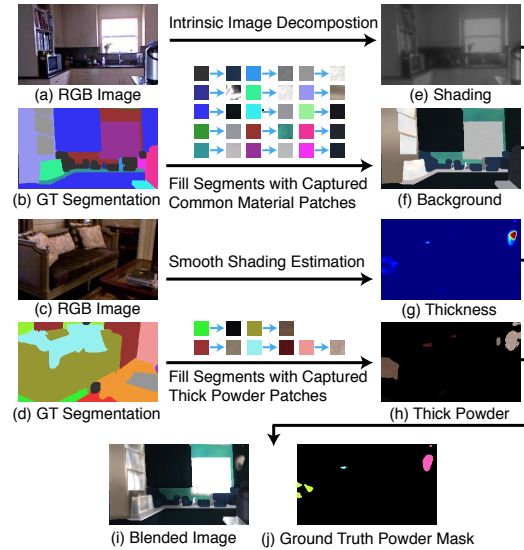

Figure 10. **Powder against background data synthesis**. (a)(c) are RGB regions from NYUDv2 [34], and (b)(d) are their segmentation labels. We obtain the shading (e) via intrinsic image decomposition, and the background image (f) by filling segments in (b) with patches from *Patch-common*. We obtain the powder thickness map (g) via smooth shading estimation, and the thick powder image (h) by filling segments in (d) with patches from *Patch-thick*, only for pixels with positive thicknesses. The final image (i) is obtained by blending background (f) and thick powder (h) using Equation 2 with (g) as $1 - \eta$, and applying shading (e). The ground truth (j) is obtained by thresholding thickness (g).

Background Synthesis: NYUDv2 provides RGB images with segmentation labels. We randomly crop a RGB region and its segmentation, and assign a random common material patch from *Patch-common* (Table 1) to each segmentation class. The synthetic background is obtained by filling the segments with the assigned patch, using image quilting [11] or resizing and cropping. The shading map of the RGB region is estimated via intrinsic image decomposition [24].
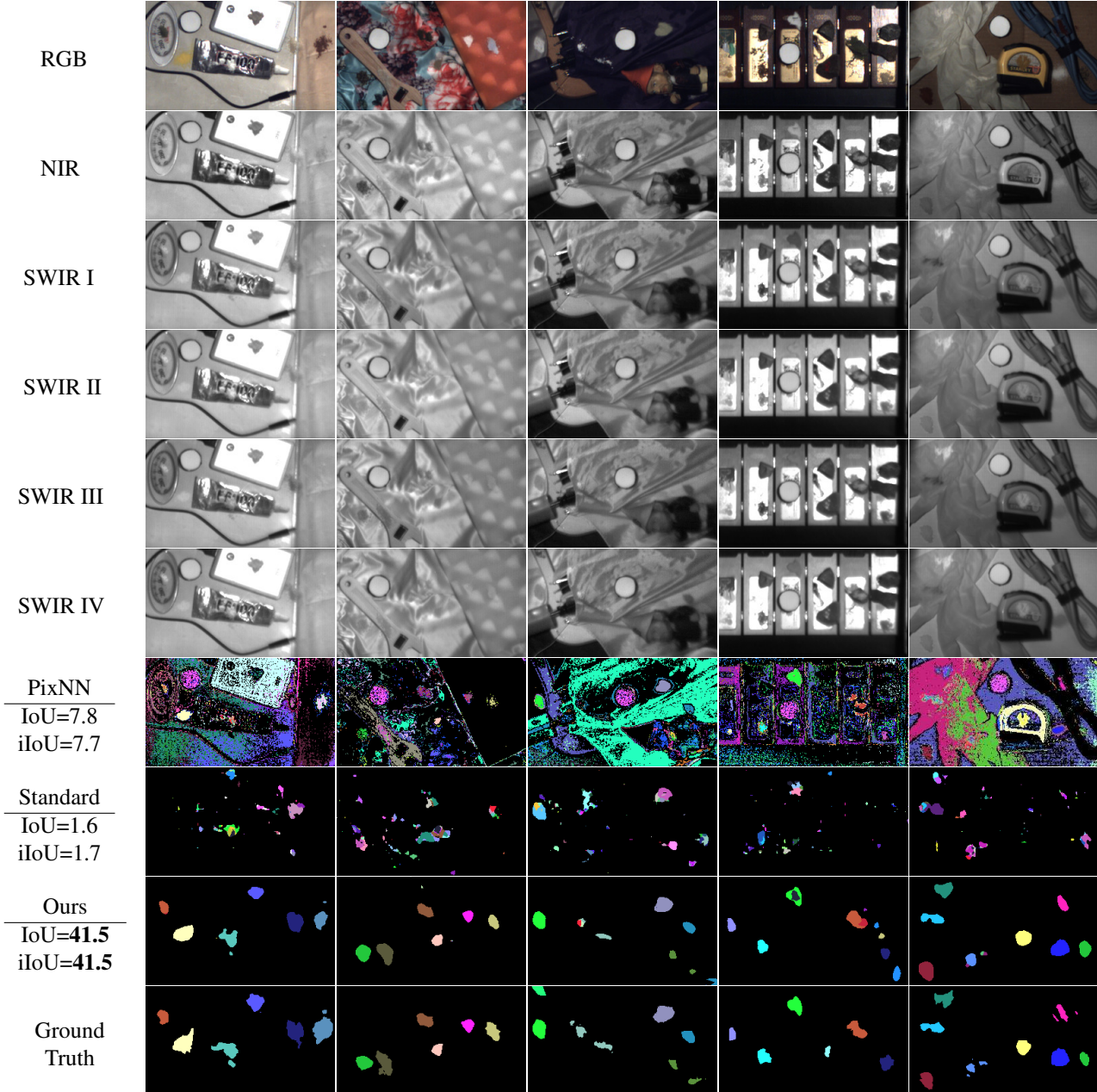
Figure 11. **Comparisons on *Scene-test*** with Per-pixel Nearest Neighbor (PixNN) and Standard Semantic Segmentation (Standard). Black color denotes background while the others denote different powders. Our method performs much better. Band selection and data synthesis lead to huge improvement over simply training on limited real data with all bands.

Powder Synthesis: Kovacs *et al.* [21] provides a method to estimate smooth shading probability. Its output heatmap looks similar to powder thickness map. We apply the method to images from NYUDv2 to obtain thickness maps. We treat the pixel values (between 0 and 1) in the heatmap as $1 - \eta$ in Equation 2. We use the same method as rendering backgrounds to render thick powder images for pixels with positive thicknesses, using patches from *Patch-thick*.

Finally, a random synthetic background and a synthetic powder mask are blended using Equation 2, with shading applied. The label is obtained by thresholding $1 - \eta$ at 0.1.

6.2. Experimental Results

We show that our method is superior by comparing with baselines, and that the Beer-Lambert Blending and NNCV band selection are necessary via ablation study.

Implementation Details: 1000 powder masks and 1000 backgrounds are rendered. We use the DeepLab v3+ [8] net, taking RGBN and 4 SWIR bands selected by NNCV as input. We train the model from scratch using AdamWR

Blending	Band Selection	Unextended		Extended	
		IoU	iIoU	IoU	iIoU
No Blending	NNCV	9.7	10.0	29.3	29.9
Alpha Blend	NNCV	30.2	30.2	39.3	39.5
Beer-Lambert	Grid Sampling	30.4	30.8	36.3	37.4
Beer-Lambert	MVPCA [6]	31.5	31.7	37.9	38.2
Beer-Lambert	Rough Set [30]	23.9	24.4	31.0	31.7
Beer-Lambert	NNCV	**36.8**	**37.0**	**42.7**	**42.2**

Table 7. **Ablation on Blending and Band Selection**. Beer-Lambert Blending with NNCV band selection is superior.

Number of SWIR Bands	Acquisition Time	Unextended		Extended	
		IoU	iIoU	IoU	iIoU
961	12min	29.1	28.9	-	-
16	12s	31.9	32.0	-	-
9	7s	34.5	34.9	42.5	**42.6**
4	3s	**36.8**	**37.0**	**42.7**	42.2
1	0.75s	20.4	20.4	26.6	26.9
1 (avg 961 bands)	12min	18.2	18.3	-	-
1 (avg 4 bands)	3s	16.5	17.2	24.7	24.8
0 (only RGBN)	0s	12.6	12.7	18.6	18.5

Table 8. **Number of SWIR Bands**. Selecting a few bands reduces acquisition time while improving IoU.

[25] on rendered data, and fine-tune on rendered powders against real backgrounds from *Scene-bg* and *Scene-sl-train* and pure real data from *Scene-sl-train*. *Scene-val* is for validation. See supplementary material for hyperparameters.
Evaluation Metrics: We report mean intersection over union (IoU) and mean instance-level intersection over union (iIoU) borrowed from Cityspaces [9]. We define the pixels with the same label in the same image as an instance.
Comparison with Baselines: We compare on *Scene-test* with two baselines: Per-pixel Nearest Neighbor (PixNN) finds per-pixel nearest neighbor in a database including mean patch values from *Patch-thick* and *Patch-common*. Standard Semantic Segmentation (Standard) trains DeepLab v3+ [8] on pure real data from *Scene-val* with RGBN and 961 SWIR bands. In Figure 11, our method significantly outperforms two baselines.
Ablation Study on blending, band selection, and the number of SWIR bands: Because *Scene-sl-train* and *Scene-sl-test* do not provide unselected SWIR bands, we conduct two types of experiments: (1) **Unextended** experiments do not include *Scene-sl-train* in training, and evaluate on *Scene-test* only. (2) **Extended** experiments include *Scene-sl-train* in training and evaluate on a dataset merging *Scene-test* and *Scene-sl-test*. Table 7 show that Beer-Lambert Blending and NNCV selection are better than other settings. Table 8 shows that 4 SWIR bands reach a high performance with a short acquisition time (3s), which could be used in time-sentitive applications (*e.g.* scenes with human in Figure 13).
ROC Curve: Security applications often care about the presence/absence of a specific powder rather than its exact mask. Thus, adjusting the confidence threshold, we plot
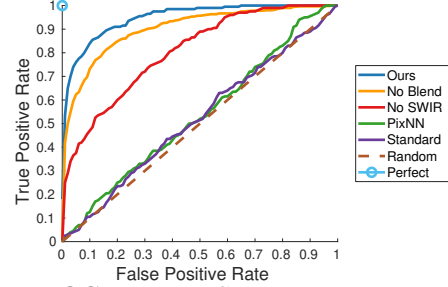

Figure 12. **ROC curve on *Scene-test***. Incorporating band selection and data synthesis, our method outperforms Per-pixel Nearest Neighbor (PixNN) and Standard Semantic Segmentation (Standard).
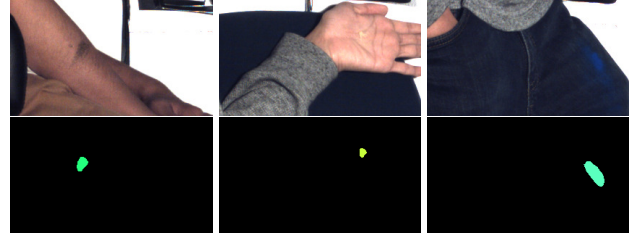

Figure 13. **Powder recognition on arm, palm and jeans**. The model is fine-tuned on ten human images with rendered powder. We vote for majority class in each connected component, and preserve components with confidence ≥ 0.95.

the ROC curve (and PR curve in supplementary material) for this 2-class classification task in Figure 12, showing the significant superiority of our method over the baselines.

7. Conclusion

The methods we present reach 60%~70% accuracy on recognition with known powder location, and over 40% mean IoU on recognition without known location (see supplementary for failure cases). We believe this performance is strong considering the fine-grained 100-class recognition problem at hand, especially one where large amounts of data are not readily available or very hard to collect.

Though this accuracy may not be sufficient for a safety application that demands near perfect detection of dangerous powders, it may be improved by (a) adding more data to reduce false positives on backgrounds and/or (b) considering a wider spectral range including mid-wave IR (2~5μm). Even if powder recognition may not achieve perfect accuracy using solely visual cues, a visual recognition system can eliminate most candidates, and the top-N retrievals can be further tested via other means (microscopic, chemical). This work is an initial attempt at solving powder recognition and we will address the above issues in the future.

References

[1] Edward H Adelson. On seeing stuff: the perception of materials by humans and machines. In *Human vision and electronic imaging VI*, volume 4299, pages 1–13. International Society for Optics and Photonics, 2001. 1

[2] Yan Bai, Yihang Lou, Feng Gao, Shiqi Wang, Yuwei Wu, and Ling-Yu Duan. Group-sensitive triplet embedding for vehicle reidentification. *IEEE Transactions on Multimedia*, 20(9):2385–2399, 2018. 1

[3] Sean Bell, Paul Upchurch, Noah Snavely, and Kavita Bala. Material recognition in the wild with the materials in context database. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3479–3487, 2015. 1

[4] Zalán Bodó. Some optical properties of luminescent powders. *Acta Physica Academiae Scientiarum Hungaricae*, 1(2):135–150, 1951. 2

[5] Ayan Chakrabarti and Todd Zickler. Statistics of real-world hyperspectral images. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 193–200. IEEE, 2011. 2

[6] Chein-I Chang, Qian Du, Tzu-Lung Sun, and Mark LG Althouse. A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. *IEEE transactions on geoscience and remote sensing*, 37(6):2631–2641, 1999. 2, 3, 4, 6, 8

[7] Chein-I Chang and Su Wang. Constrained band selection for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 44(6):1575–1585, 2006. 2

[8] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *European Conference on Computer Vision*. Springer, 2018. 7, 8

[9] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016. 8

[10] Joseph DeGol, Mani Golparvar-Fard, and Derek Hoiem. Geometry-informed material recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1554–1562, 2016. 1

[11] Alexei A Efros and William T Freeman. Image quilting for texture synthesis and transfer. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 341–346. ACM, 2001. 6

[12] Xiurui Geng, Kang Sun, Luyan Ji, and Yongchao Zhao. A fast volume-gradient-based band selection method for hyperspectral image. *IEEE Transactions on Geoscience and Remote Sensing*, 52(11):7111–7119, 2014. 2

[13] Harry G Hecht. The interpretation of diffuse reflectance spectra. In *Standardization in Spectrophotometry and Luminescence Measurements: Proceedings of a Workshop Seminar Held at the National Bureau of Standards, Gaithersburg, Maryland, November, November 19-20, 1975*, volume 466, page 57. US Department of Commerce, National Bureau of Standards, 1976. 2

[14] Gordon Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE transactions on information theory*, 14(1):55–63, 1968. 2

[15] Sen Jia, Guihua Tang, Jiasong Zhu, and Qingquan Li. A novel ranking-based clustering approach for hyperspectral band selection. *IEEE Transactions on Geoscience and Remote Sensing*, 54(1):88–102, 2016. 2

[16] Peter D Johnson. Absolute optical absorption from diffuse reflectance. *JOSA*, 42(12):978–981, 1952. 2

[17] Christos Kampouris, Stefanos Zafeiriou, Abhijeet Ghosh, and Sotiris Malassiotis. Fine-grained material classification using micro-geometry and reflectance. In *European Conference on Computer Vision*, pages 778–792. Springer, 2016. 1

[18] Leonid Karlinsky, Joseph Shtok, Yochay Tzur, and Asaf Tzadok. Fine-grained recognition of thousands of object categories with single-example training. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4113–4122, 2017. 1

[19] Kodo Kawase. Terahertz imaging for drug detection and large-scale integrated circuit inspection. *Optics and photonics news*, 15(10):34–39, 2004. 2

[20] Kodo Kawase, Yuichi Ogawa, Yuuki Watanabe, and Hiroyuki Inoue. Non-destructive terahertz imaging of illicit drugs using spectral fingerprints. *Optics express*, 11(20):2549–2554, 2003. 2

[21] Balazs Kovacs, Sean Bell, Noah Snavely, and Kavita Bala. Shading annotations in the wild. In *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, pages 850–859. IEEE, 2017. 7

[22] Jonathan Krause, Hailin Jin, Jianchao Yang, and Li Fei-Fei. Fine-grained recognition without part annotations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5546–5555, 2015. 1

[23] Paul Kubelka and Franz Munk. An article on optics of paint layers. *Z. Tech. Phys*, 12(593-601), 1931. 2, 4

[24] Zhengqi Li and Noah Snavely. Learning intrinsic image decomposition from watching the world. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9039–9048, 2018. 6

[25] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. 8

[26] Yihang Lou, Yan Bai, Jun Liu, Shiqi Wang, and Ling-Yu Duan. Embedding adversarial learning for vehicle reidentification. *IEEE Transactions on Image Processing*, 2019. 1

[27] Adolfo MartÍnez-UsÓMartinez-Uso, Filiberto Pla, José Martínez Sotoca, and Pedro García-Sevilla. Clustering-based hyperspectral band selection using information measures. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12):4158–4171, 2007. 2

[28] NT Melamed. Optical properties of powders. part i. optical absorption coefficients and the absolute value of the diffuse reflectance. part ii. properties of luminescent powders. *Journal of Applied Physics*, 34(3):560–570, 1963. 2

[29] Matthew P Nelson, Shawna K Tazik, Patrick J Treado, Tiancheng Zhi, Srinivasa Narasimhan, Bernardo Pires, and Martial Hebert. Real-time, short-wave, infrared hyperspectral conforming imaging sensor for the detection of threat materials. In *Next-Generation Spectroscopic Technologies XI*, volume 10657, page 106570U. International Society for Optics and Photonics, 2018. 2

[30] Swarnajyoti Patra, Prahlad Modi, and Lorenzo Bruzzone. Hyperspectral band selection based on rough set. *IEEE Transactions on Geoscience and Remote Sensing*, 53(10):5495–5503, 2015. 2, 3, 4, 6, 8

[31] Thomas Porter and Tom Duff. Compositing digital images. In *ACM Siggraph Computer Graphics*, volume 18, pages 253–259. ACM, 1984. 2

[32] Gabriel Schwartz and Ko Nishino. Automatically discovering local visual material attributes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3565–3573, 2015. 1

[33] YC Shen, a T Lo, PF Taday, BE Cole, WR Tribe, and MC Kemp. Detection and identification of explosives using terahertz pulsed spectroscopic imaging. *Applied Physics Letters*, 86(24):241116, 2005. 2

[34] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *European Conference on Computer Vision*, pages 746–760. Springer, 2012. 6

[35] EL Simmons. An equation relating the diffuse reflectance of weakly absorbing powdered samples to the fundamental optical parameters. *Optica Acta: International Journal of Optics*, 18(1):59–68, 1971. 2

[36] Alexandru Telea. An image inpainting technique based on the fast marching method. *Journal of graphics tools*, 9(1):23–34, 2004. 5

[37] Lin Wang, Chein-I Chang, Li-Chien Lee, Yulei Wang, Bai Xue, Meiping Song, Chuanyan Yu, and Sen Li. Band subset selection for anomaly detection in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 55(9):4887–4898, 2017. 2

[38] Shaohong Wang, Bradley Ferguson, Carmen Mannella, Derek Abbott, and X-C Zhang. Powder detection using thz imaging. In *Lasers and Electro-Optics, 2002. CLEO'02. Technical Digest. Summaries of Papers Presented at the*, pages 131–vol. IEEE, 2002. 2

[39] Ting-Chun Wang, Jun-Yan Zhu, Ebi Hiroaki, Manmohan Chandraker, Alexei A Efros, and Ravi Ramamoorthi. A 4d light-field dataset and cnn architectures for material recognition. In *European Conference on Computer Vision*, pages 121–138. Springer, 2016. 1

[40] Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, Yi-Fan Chang, and Qiao-Liang Xiang. A leaf recognition algorithm for plant classification using probabilistic neural network. In *Signal Processing and Information Technology, 2007 IEEE International Symposium on*, pages 11–16. IEEE, 2007. 1

[41] Jia Xue, Hang Zhang, Kristin J Dana, and Ko Nishino. Differential angular imaging for material recognition. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6940–6949, 2017. 1

[42] Shulin Yang, Liefeng Bo, Jue Wang, and Linda G Shapiro. Unsupervised template learning for fine-grained object recognition. In *Advances in neural information processing systems*, pages 3122–3130, 2012. 1

[43] Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K Nayar. Generalized assorted pixel camera: post-capture control of resolution, dynamic range, and spectrum. *IEEE transactions on image processing*, 19(9):2241–2253, 2010. 2

[44] Hang Zhang, Kristin Dana, and Ko Nishino. Friction from reflectance: Deep reflectance codes for predicting physical surface properties from one-shot in-field reflectance. In *European Conference on Computer Vision*, pages 808–824. Springer, 2016. 1

[45] Hang Zhang, Jia Xue, and Kristin Dana. Deep ten: Texture encoding network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2896–2905. IEEE, 2017. 1