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.

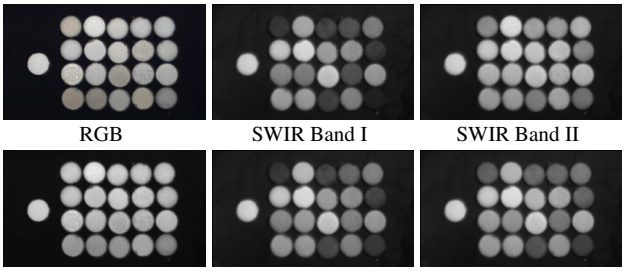


Figure 1. White powders that are not distinguishable in visible light (RGB) and Near Infrared (NIR) show significantly different appearances in Shortwave Infrared (SWIR). The leftmost sample is a white patch for white balance while the others are powders. Row 1 (left to right): Cream of Rice, Baking 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 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

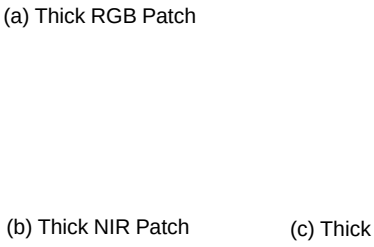
We build the first comprehensive RGBN-SWIR Multispectral Database for powder recognition. We first introduce the acquisition system in Section 3.1. In Section 3.2, we describe the dataset—*Patches* providing resources for image based rendering, and *Scenes* providing cluttered backgrounds with or without 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 ($1.5\text{V} \leq V_0, V_1 \leq 4.5\text{V}$). 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).

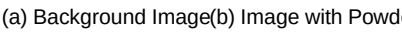
3.2. Patches and Scenes

The dataset includes two parts: Patches provides patches (size 14 × 14) to use for image based rendering; Scenes provides scenes (size 280 × 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 background patch are captured in the same field of view.

(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].

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-