

MobiSpectral: Hyperspectral Imaging on Mobile Devices

Neha Sharma, M. Shahzaib Waseem, Shahrzad Mirzaei, Mohamed Hefeeda
Simon Fraser University, Burnaby, BC, Canada

ABSTRACT

Hyperspectral imaging systems capture information in multiple wavelength bands across the electromagnetic spectrum. These bands provide substantial details based on the optical properties of the materials present in the captured scene. The high cost of hyperspectral cameras and their strict illumination requirements make the technology out of reach for end-user and small-scale commercial applications. We propose MobiSpectral, which turns a low-cost phone into a simple hyperspectral imaging system, without any changes in the hardware. We design deep learning models that take regular RGB images and near-infrared (NIR) signals (which are used for face identification on recent phones) and reconstruct multiple hyperspectral bands in the visible and NIR ranges of the spectrum. Our experimental results show that MobiSpectral produces accurate bands that are comparable to ones captured by actual hyperspectral cameras. The availability of hyperspectral bands that reveal hidden information enables the development of novel mobile applications that are not currently possible. To demonstrate the potential of MobiSpectral, we use it to identify organic solid foods, which is a challenging food fraud problem that is currently partially addressed by laborious, unscalable, and expensive processes. We collect large datasets in real environments under diverse illumination conditions to evaluate MobiSpectral. Our results show that MobiSpectral can identify organic foods, e.g., apples, tomatoes, kiwis, strawberries, and blueberries, with an accuracy of up to 94% from images taken by phones.

CCS CONCEPTS

- Human-centered computing → *Ubiquitous and mobile computing*;
- Applied computing → Computer forensics.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ACM MobiCom '23, October 2–6, 2023, Madrid, Spain

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM

ACM ISBN 978-1-4503-9990-6/23/10...\$15.00

<https://doi.org/10.1145/3570361.3613296>

KEYWORDS

Hyperspectral Imaging, Mobile Applications, Food Fraud

ACM Reference Format:

Neha Sharma, M. Shahzaib Waseem, Shahrzad Mirzaei, Mohamed Hefeeda. 2023. MobiSpectral: Hyperspectral Imaging on Mobile Devices. In *The 29th Annual International Conference on Mobile Computing and Networking (ACM MobiCom '23), October 2–6, 2023, Madrid, Spain*. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3570361.3613296>

1 INTRODUCTION

The hardware and software resources on mobile phones are continually improving. For example, most recent phones have multiple camera modules, which enable capturing better quality photos and videos even in low light conditions. The goal of this paper is to push the boundaries of mobile phones even further, by bringing some of the *hyperspectral imaging* features to them. This will enable many novel applications that are currently not possible on mobile phones, such as food quality inspection, material identification, early detection of skin diseases, and artwork authentication. Spectral features can also improve the performance of current image/video processing applications, such as color correction [14], dehazing [13], scene analysis [29], face identification [3], and automatic skin enhancement in images [33].

Hyperspectral imaging systems capture information in multiple wavelengths across the electromagnetic spectrum, as illustrated in Figure 1. These bands provide substantial details based on the optical properties of the materials present in the captured scene, and thus they can be used to create *spectral signatures* for different materials. Figure 1 also shows sample spectral signatures. Hyperspectral imaging has been established for decades in domains such as remote sensing and military applications. However, its complexity and high cost make the technology out of reach for end-user applications. In addition, hyperspectral imaging requires strict illumination conditions that are not possible to satisfy in everyday environments such as homes and grocery stores.

We propose a new system, called MobiSpectral, that turns a low-cost phone into a simple hyperspectral imaging system. We design deep learning models that take regular RGB images and near-infrared (NIR) signals (which are used for face identification on recent phones) and reconstruct multiple hyperspectral bands in the visible and NIR ranges of the

spectrum. Based on these bands, we create spectral signatures that can differentiate various materials and detect *subtle* differences in their compositions and structures. To demonstrate the practicality of MobiSpectral, we use it to address a challenging problem: identifying organic *solid* foods, e.g., fruits, on mobile phones, which, to the best of our knowledge, has not been done before in the literature.

The contributions of this paper are as follows:

- We propose MobiSpectral, which includes multiple new ideas to bring hyperspectral features to phones and enable many new mobile applications.
- We design a robust deep-learning model to convert images captured by regular cameras to hyperspectral bands across the visible and NIR range, in §3.3. Our design builds on the state-of-the-art model in [9] and makes it practical.
- We present methods to mitigate the negative effects of diverse illumination sources on the reconstructed hyperspectral bands, which enable MobiSpectral to work in everyday environments, in §3.5.
- We analyze the spectral characteristics of organic foods and show that they can be identified by cameras operating in the visible and NIR range, in §4.2.
- We design and evaluate a mobile application to identify organic foods as a proof-of-concept of MobiSpectral.
- We collect datasets of hyperspectral, RGB, and NIR images, which can be useful for various mobile applications.
- We conduct extensive experimental studies to demonstrate the accuracy of our reconstruction model and mobile application, in §5. For example, our results show that MobiSpectral can identify organic kiwis, apples, tomatoes, strawberries, and blueberries with accuracy 94%, 93%, 92%, 90%, and 89%, respectively, from images taken by phones.

We summarize the related work in §6 and conclude in §7.

2 BACKGROUND AND CHALLENGES

2.1 Background

RGB Imaging. A simplified illustration of regular cameras, e.g., ones on phones, is shown in Figure 2.a, where a CMOS sensor converts the incoming light into three channels (RGB) using a color filter array (CFA). The most common CFA is the shown Bayer pattern: a green square means that the green color is allowed for this pixel, similarly for the red and blue squares. Each pixel captures only one of the three colors, and an interpolation (aka demosaicing) method implemented in hardware is used to estimate the other two. Notice that there are two green pixels for each blue and red pixel, as the human visual system is more sensitive to the green color.

Most sensors in current RGB cameras have a sensitivity of up to 1000 nm, whereas the visible light range is 400–700 nm. RGB cameras use a cut-off filter to truncate signals between 700 and 1000 nm, which is the NIR range. If not removed,

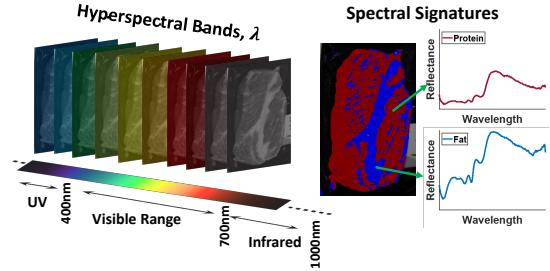


Figure 1: Hyperspectral bands and spectral signatures.

NIR signals may over-saturate the red channel and cause damage to the captured RGB image. Figure 2.b shows the sensitivity of the CMOS sensor to different color channels.

Hyperspectral Imaging. Hyperspectral imaging measures the intensity of diffusely reflected light from a surface at many wavelengths with relatively narrow band-pass filters. Figure 3.a illustrates the basic concepts of hyperspectral imaging. These systems require specialized hardware such as collimating optics and a wavelength dispersion component (diffraction grating or prism) [11]. Hyperspectral images are three-dimensional; thus, a scanning technique is required to collect them on the two-dimensional sensor. Line scanning is the commonly used method, which collects photons from a line of pixels in the spatial domain through a slit, and it splits the light into its components to produce many bands at the same time [11]. The complete data cube is obtained through a relative motion between the camera and the scene.

Hyperspectral cameras have much finer spectral sensitivity than regular RGB cameras, as illustrated in Figure 2.b. This enables obtaining more detailed information and creating spectral signatures for the different materials present in the captured scene. As illustrated in Figure 1, the spectral signature is created per pixel by tracking how the reflectance value changes across different wavelengths.

2.2 Challenges

The goal of our work is to bring some features of hyperspectral imaging to mobile phones. This, however, faces several research challenges, which we summarize in the following.

Lack of information beyond the visible range. Hyperspectral applications rely on the optical properties of materials, e.g., how they absorb and reflect various wavelengths of the incident light, especially ones in the 700–1000 nm range, as these wavelengths reveal hidden information that cannot be obtained from the visible range (400–700 nm). However, regular cameras use cutoff filters to truncate all signals in that range (Figure 2.b). A straightforward solution for this problem is to remove the IR cutoff filter. This, however, requires changing the phone hardware and may lead to damaging the quality of regular RGB images. The quality of regular

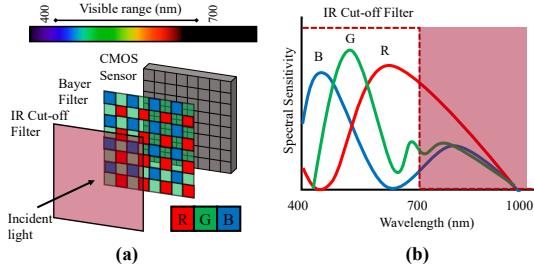


Figure 2: Basic principles of regular RGB cameras.

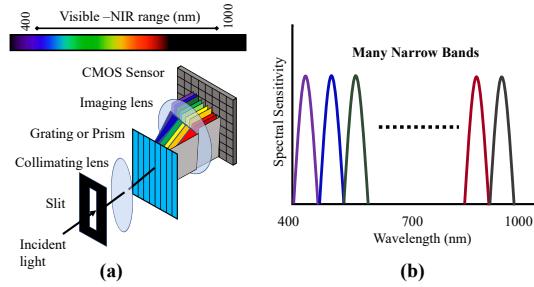


Figure 3: Basic principles of hyperspectral cameras.

images is critical for most users. In §3.2, we show how we address this problem without requiring any changes to the phone hardware or interfering with the RGB image pipeline, by utilizing the low-resolution infrared signal available in recent phones and used for face identification.

Very few captured bands. Regular cameras capture only three channels (or bands), which are sufficient for the human visual system to perceive the captured scene. In contrast, hyperspectral cameras capture many equally spaced and much narrower bands in the entire spectral range of the camera sensor (400–1000 nm). Commercial hyperspectral cameras, such as the one used in this study, can produce more than 200 bands. Unlike the three wide RGB bands, hyperspectral bands yield more accurate spectral signatures, which are essential for identifying various materials. To address this problem, in §3.3, we design a deep learning model to reconstruct multiple narrow bands from the input RGB ones.

Strict illumination requirements. *Indoor* hyperspectral imaging requires *strict* illumination conditions [23, 37]; otherwise, the captured bands will have noise and could even be completely damaged. Specifically, halogen bulbs are typically required for indoor hyperspectral applications because they cover most of the 400–1000 nm spectral range. However, such sources are not available in everyday environments such as grocery stores and shopping malls. Halogen bulbs are expensive, have a short lifetime, dissipate significant heat, and consume substantially more energy than common fluorescent and LED bulbs. They are mostly used in special

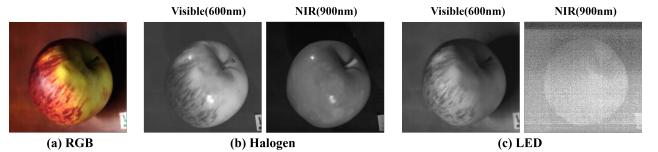


Figure 4: Effect of illumination on sample bands captured by a hyperspectral camera.

environments such as forensics labs and industrial plants. Therefore, capturing hyperspectral images in everyday indoor environments is difficult, even with actual hyperspectral cameras. Doing it on mobile phones with limited optical and processing resources is significantly more challenging.

To demonstrate the importance of illumination, we captured images with a hyperspectral camera when the scene was first illuminated with a halogen source and then with an LED source. We show sample bands for both cases in Figure 4. The figure shows that the NIR bands captured using the LED source are mostly damaged. In §3.4, we analyze the limitations of common illumination sources and present a solution to mitigate these limitations.

3 PROPOSED MOBISPECTRAL SYSTEM

3.1 Overview

Figure 5 shows an overview of the proposed MobiSpectral system. At the core of MobiSpectral is a *robust* hyperspectral reconstruction model (described in §3.3), which, unlike prior models in the literature, reconstructs bands in the visible and NIR ranges and handles realistic illumination conditions. As discussed before, information in the visible range alone is insufficient for spectral analysis of materials, as the most crucial information is typically found in the NIR range. Further, since our goal is to enable users to perform spectral analysis, e.g., to detect food fraud, using their phones in everyday environments such as homes and grocery stores, strict illumination cannot be expected.

The inputs to the reconstruction model are regular RGB and NIR images; both can be captured by recent phones as described in §3.2. The RGB image is scaled to the same resolution as the NIR image, and both images are then aligned. To handle the limitation of common illumination sources, which is the lack of emitted power in the NIR range, we utilize the infrared illumination source on recent phones, which is used in conjunction with the NIR camera for face identification, as detailed in §3.4. To increase the robustness of the model and make it usable with different phone cameras and diverse illumination conditions, we propose, in §3.5, an image normalization method that uses a deep learning model to abstract the processing pipelines of different cameras and transform the input images to a common representation.

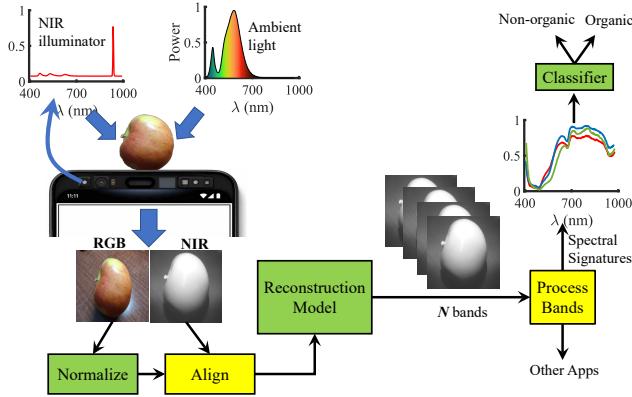


Figure 5: Overview of MobiSpectral.

The output of the reconstruction model is N bands, equally spaced across the entire 400–1000 nm spectrum. N is a configurable system parameter, where larger N values produce more accurate spectral signatures, but they require more processing and memory resources. The model infers the hyperspectral bands from the input RGB and NIR signals. That is, it uses the RGB signals in Figure 2.b to estimate the bands in Figure 3.b. As Figure 2.b shows, the RGB signals contain the needed spectral information, but in coarse grain and only in the 400–700 nm range. The NIR signal partially mitigates the lack of information in the 700–1000 nm range. Thus, the reconstruction model effectively upsamples the RGB+NIR signals into N spectral bands. This is similar in principle to super-resolution models in the computer vision community.

The reconstructed bands offer rich information in the spatial and spectral domains about every pixel in the captured scene. Thus, they can enable many new mobile applications and improve existing ones, including detecting food fraud, assessing the quality and ripeness of crops, analyzing the nutritional contents of food products, and identifying/classifying various materials in general. We provide a concrete case study on identifying organic foods in §4.

We note that the training of all models in MobiSpectral (reconstruction, image normalization, and classification) is done on a workstation with GPUs. Then, the trained models are uploaded to phones and used for inference in real time.

3.2 Obtaining Signals in the Invisible Range

Recent phones, e.g., Apple iPhone X, Google Pixel 4, and their sequels, contain multiple cameras. One of these cameras is typically used for face identification to unlock the phone. This camera operates in the NIR range, and we refer to it as the NIR camera. We propose using the NIR camera to obtain *some* information beyond the visible range. Recall that RGB cameras filter out all invisible signals beyond 700 nm, as

discussed in §2.1. The information from the NIR camera is then added to the RGB information captured by one of the regular cameras and used by the deep learning model in §3.3 to reconstruct bands across the 400–1000 nm range.

The NIR and RGB cameras are physically separated, which can introduce discrepancies in the captured pixels. In addition, NIR cameras typically have much lower resolutions than RGB cameras. Both of these issues pose a problem for reconstructing hyperspectral bands and lead to inaccurate spectral signatures. Recall that spectral signatures are computed per pixel across the frequency domain, which reflects the optical properties of the captured scene. To mitigate these issues, we rescale RGB images and align them with NIR images using the simple method in [12]. We note that we downsample the RGB images to match the resolution of NIR instead of upsampling the NIR, because NIR upsampling may introduce more errors than downsampling RGB images while not necessarily providing better reconstruction. In addition, actual hyperspectral cameras, including the one used in this paper, typically have low spatial resolutions (e.g., 512x512), which means that the reconstruction model would be trained on low-resolution bands in the first place.

3.3 Robust Hyperspectral Reconstruction

Limitations of Existing Models. Spectral reconstruction has been an active topic in the last few years in the computer vision community [4, 5]. Most prior works, however, reconstruct bands only in the visible range (400–700 nm) and assume ideal illumination conditions, mostly considering outdoor environments with sunlight that covers the entire 400–1000 nm spectral range. In addition, all previous models were trained and tested on datasets from the same data distribution. Specifically, all datasets used in previous models were captured by hyperspectral cameras, where a part of the datasets is used for training and another for testing. This, however, is not possible in our case because testing (reconstruction) must be done using images captured by phones, while training is done using images captured by hyperspectral cameras as they provide ground truth. And as explained in §2.1, hyperspectral cameras work in a completely different manner than phone cameras: they produce many narrow bands instead of three wide bands. This significant mismatch between the training and testing data distributions introduces substantial errors in the bands reconstructed by previous models, which would compromise the accuracy of any practical application designed to utilize these bands.

Further, previous spectral reconstruction works did not consider the differences in the sensitivity of the CMOS sensors (see Figure 2), which vary between manufacturers and camera models. The variation in the sensor sensitivity results in differences in pixel values, which affects the accuracy

of the reconstructed bands. Finally, each camera applies a sequence of steps, e.g., white balancing, demosaicing, and color transformation, on the captured raw data to produce the final RGB images. The implementation and tuning of these steps differ across camera manufacturers and models. Thus, the produced RGB images will have differences at the pixel level, despite looking visually similar. These differences, too, reduce the accuracy of the reconstructed bands.

Design Goals. We address all of the above limitations. Specifically, MobiSpectral is designed to: (i) produce accurate bands in the visible and NIR ranges from images captured by mobile phones, (ii) account for differences in the sensors and processing pipelines of various RGB cameras, and (iii) function under diverse illumination conditions in real environments such as homes and grocery stores.

Basic Reconstruction Model. To achieve the above goals, we build on the state-of-the-art MST++ model in [9]. MST++ is designed using transformers, which are recent deep-learning models that consider the long-range dependency among data points in the input sequences through a mechanism known as self-attention. MST++ introduces multiple ideas and constructs to consider the spatial and spectral characteristics of hyperspectral images, but it still suffers from the limitations mentioned above.

We note that we are not claiming a new neural network design for hyperspectral reconstruction. Rather, our contribution is making the state-of-the-art model (MST++) robust to real-life illumination settings and extending it to produce bands in the NIR range, which makes it much more practical and useful for many applications.

Our first, simple, extension of the MST++ model is to add the NIR range, which is done by expanding the input layer to take the scaled RGB bands and the NIR image captured simultaneously by the phone. At a high level, the reconstruction model functions as follows. The input RGB and NIR signals, containing sparse spectral information in the visible and NIR range, are first upsampled to N spectral features. Then, each feature is embedded as a token in the spectral domain. This embedding maps similar spectral features to positions close to each other, which assists the self-attention mechanism in capturing the long-range inter-dependencies among features in the visible and NIR ranges. This helps in reconstructing bands in the NIR range, where the input signal is sparse but has similar characteristics to the nearby bands.

In the following subsections, we describe how we improve the robustness of this model.

3.4 Handling Limited Illumination

We analyze and address the limitations of common illumination sources. In hyperspectral imaging, the reflected light from the captured scene is divided into components using a

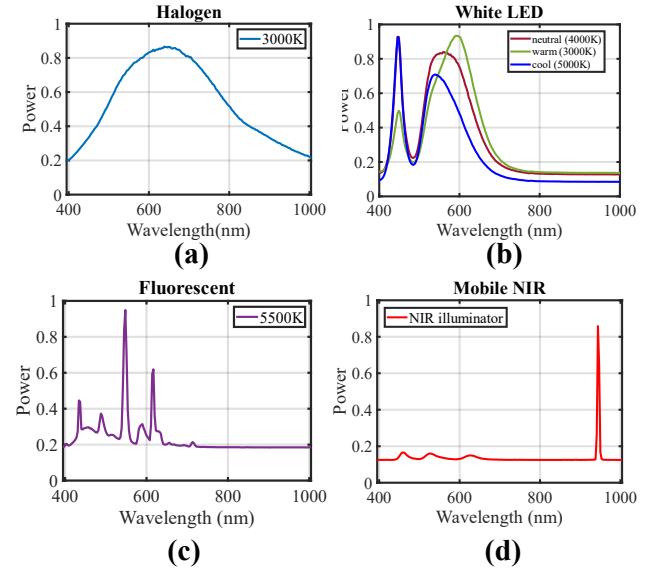


Figure 6: Spectral characteristics of light sources.

prism or grating [7, 26], as illustrated in Figure 3. We denote the image value in the hyperspectral band n at the spatial position (x, y) as $I(x, y, \lambda_n)$, which can be defined as [15]:

$$I(x, y, \lambda_n) = \int R(x, y, \lambda) \cdot L(x, y, \lambda) \cdot C_n(\lambda) \cdot d\lambda, \quad (1)$$

where $L(x, y, \lambda)$ is the illumination intensity at position (x, y) which is a function of the wavelength λ , $R(x, y, \lambda)$ is the surface reflectance at position (x, y) which depends on the optical properties of the materials in the scene, and $C_n(\lambda)$ is the sensitivity function of the camera for the n^{th} band which is specified by the camera manufacturer. In practice, the spectral information is discretized across the wavelength, and Equation (1) can be rewritten as:

$$I(x, y, \lambda_n) = \sum_{i=1}^N R(x, y, \lambda_i) \cdot L(x, y, \lambda_i) \cdot C_n(\lambda_i), \quad (2)$$

where $\lambda_n, n = 1, 2, \dots, N$, is the discrete representation of wavelength λ into N spectral bands. Equation (2) highlights the crucial role of illumination in hyperspectral imaging: the source $L(x, y, \lambda_n)$ should emit power across all wavelengths to be represented in the captured hyperspectral image.

We experimentally analyze the characteristics of different illumination sources and their impact on hyperspectral imaging. We consider three categories of common light sources: halogen, LED, and fluorescent. We plot in Figure 6.a the normalized power emitted across various bands when the halogen light source is used. We produced this figure by illuminating a *white* object, which reflects all wavelengths, with a halogen light source and capturing the reflected signals

using a hyperspectral camera. We repeated the experiment with multiple LED light sources that have different color temperatures (Figure 6.b) and a fluorescent light source (Figure 6.c). Figure 6 shows that the halogen source emits power across most of the 400–1000 nm range, while the LED and fluorescent sources emit power mostly in parts of the visible range and almost nothing in the NIR range. The curves in Figure 6 along with Equation (2) explain the reasons behind the damaged NIR bands in Figure 4.b, which were captured using an LED source.

In addition, Figure 6.b shows that the three LED sources have diverse spectral curves, and all are different from the fluorescent one in Figure 6.c. This variability leads to inconsistent pixel values of captured images, which creates another illumination challenge for hyperspectral applications.

In summary, practical illumination sources, e.g., LED and fluorescent, pose two main problems: (i) small power emitted in the NIR range and (ii) variability of the emitted power based on the color temperatures.

To partially mitigate the first problem, we propose using the infrared illumination source that is *already available* on recent phones. Specifically, as mentioned before, most recent phones have NIR cameras for face identification. For such NIR cameras to work in dark and low-light environments, manufacturers equip their phones with an infrared illumination source, which emits infrared light, typically around the 932 nm wavelength. We utilize this source to *complement* the ambient light sources in common everyday environments, which are usually LED and fluorescent sources.

In Figure 6.d, we analyze the spectral characteristics of the infrared light source on phones. Notice that, although the infrared source does not cover most of the NIR range, as in the case of the halogen source (Figure 6.a), it provides a much-needed power in that range. Our spectral reconstruction model capitalizes on this NIR power to improve its robustness under practical illumination conditions.

The variability of the emitted power from various light sources affects the captured images, and it interacts with the processing pipeline in the camera. We address this issue §3.5.

3.5 Handling Diversity in Illumination and Camera Processing Pipeline

Regular cameras perform multiple processing steps on the captured raw data to improve the visual quality of the final RGB image. These steps include white balancing, demosaicing, color transformation, and color rendering. All of these steps are performed by the integrated signal processor (ISP), which is part of the camera hardware. White balancing is the most crucial step in the processing pipeline since it depends on the sensitivity of the sensor and the scene's illumination, while other steps mostly apply standard functions.

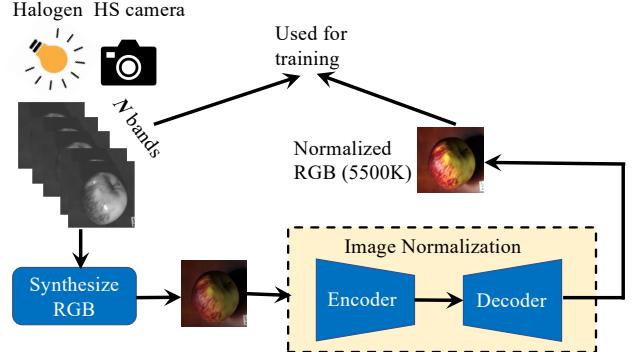


Figure 7: Image normalization in MobiSpectral.

White balancing strives to make all objects in the scene appear as if they were captured under an ideal “white light”. It consists of two steps. The first step estimates the camera sensor’s response to the illumination in the scene and produces the raw RGB channels. The second step applies linear scaling to each of the RGB channels to adjust the illumination. The camera sensor and the illumination estimation method vary across camera manufacturers and models. Further, manufacturers usually employ various proprietary enhancements to improve the image appearance. This indicates that the white balancing, and the entire processing pipeline in general, perform *non-linear* operations that vary across cameras, which may result in significant differences in the produced RGB images. This complicates the training of the hyperspectral reconstruction model and reduces its accuracy.

To handle the wide diversity in illuminations and phone cameras, we propose abstracting the processing pipelines of different cameras using a non-linear function. This function maps the camera’s raw input data to a *normalized* output format that is independent of the camera sensor sensitivity and the scene’s illumination. The normalized format will be used in both the training and inference of the reconstruction model, thus enabling our model to produce accurate bands for different cameras and illumination conditions.

Specifically, we design a deep-learning encoder-decoder model for image normalization, which is illustrated in Figure 7. Our design is based on the recent model in [2]. Unlike [2], however, we normalize all input images to a *common* illumination setting, regardless of the specific details of the camera(s) that captured these images and the likely different illumination(s) that were present during capturing, whereas the model in [2] tries to achieve the best white-balancing results for each camera. In our experiments, we normalize all images to the *daylight* illumination setting, which has a color temperature of 5500 Kelvin (K). While our model would work with any temperature, we chose daylight for better visualization.

Figure 7 summarizes the normalization process for the ground truth images captured by actual hyperspectral cameras and used in training the reconstruction model. Hyperspectral images must be captured using a halogen illumination source. In addition, hyperspectral cameras are not designed to produce regular RGB images; rather, they produce many narrow bands. We synthesize RGB images from the spectral bands. Then, the synthesized RGB images are normalized and used with the ground truth bands to train the reconstruction model.

The image normalization process is also used in the inference stage. All RGB images captured by various phone cameras and under diverse illuminations are normalized to the common setting. The normalized RGB images are then used with the NIR images to reconstruct their corresponding hyperspectral bands using the pre-trained reconstruction model. Note that we do not normalize the NIR images as they are not impacted by illumination in the visible range.

It is important to notice that without our normalization process, it would be difficult to train the reconstruction model. This is because the model needs to be trained using ground truth hyperspectral images captured in halogen illumination, which is not common in everyday environments. At the same time, the inference must be done on images captured by mobile phones under arbitrary illumination.

4 IDENTIFYING ORGANIC FOODS

4.1 Limitations of Current Approaches

Organic farming prohibits the use of most synthetic fertilizers, pesticides, and genetically modified organisms, which improves biodiversity and reduces the risk of exposure to carcinogenic substances for farm workers [16]. The global market of organic food has been steadily increasing over the past few decades, recently reaching \$120B USD annually [1]. Organic foods come at higher prices than non-organic ones, which may provide incentives for some producers or distributors to cheat. Currently, producers need to follow strict rules to obtain *organic certificates*, which allow them to label their products as organic. Most countries have organizations that issue organic certificates and periodically check the compliance of producers. Certification typically involves dispatching human agents to farms, sending food and soil samples to specialized labs for chemical and biological analyses, and inspecting used fertilizers and pesticides. This is a time-consuming and laborious process, which adds a significant cost to the production of organic foods. This increases the prices of organic foods, which, in turn, creates more incentives for cheating.

In addition to cost, the scale of the organic certification organizations is limited. Further, policing the truthful labeling

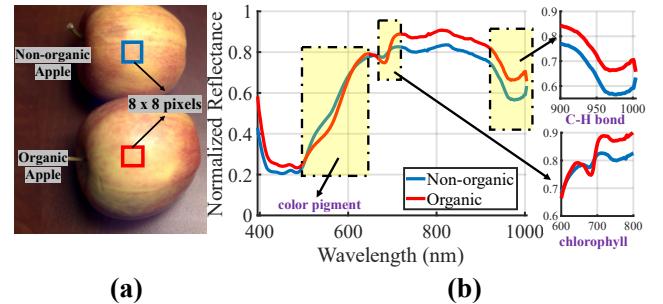


Figure 8: Spectral signatures of apples.

of organic foods is challenging, because of the multiple entities involved in the production, storage, and transportation of foods. For example, some small growers and businesses may make organic claims without realizing that they require certification. Illegal organic labeling can also occur outside the production farms by, for instance, third parties involved in the supply chains. Moreover, currently, there is no easy way for end consumers or even merchants to check the validity of organic claims other than trusting the organic labels, which may not provide enough deterrence to potential offenders.

Multiple works in the literature proposed using *spectrometers* operating in different spectral ranges to identify organic foods. For example, Song et al. [30] used the 900–1700 nm range to identify organic apples, Gupta et al. [24] used the 400–2100 nm range to identify organic apples, oranges, green onions, and bell peppers, and Xiao et al. [28] used the 800–2500 nm range to identify organic rice. Spectrometers are expensive (tens of thousands of dollars) and bulky devices, and they require careful calibration and strict illumination; they are mostly used in specialized laboratories. In addition, these spectrometers operate in spectral ranges not available in the sensors of mobile phone cameras.

4.2 Spectral Analysis of Organic Foods

We conduct a feasibility study to test the hypothesis that organic and non-organic foods can be distinguished utilizing spectral analysis in the 400–1000 nm range, which is the range available in phone cameras. This is unlike prior works, e.g., [24, 28, 30], that used specialized spectrometers operating in wider spectral ranges, as discussed above.

First, we note that multiple studies, e.g., [8, 16, 35], have analyzed the chemical and biological characteristics of organic foods and compared them to non-organic ones. For example, it has been shown that organic fruits and vegetables have lower concentrations of nitrate and higher concentrations of antioxidants, vitamins, minerals (e.g., iron, magnesium, phosphorous, and zinc), and bioactive compounds (e.g., carotenoids and tocopherols).

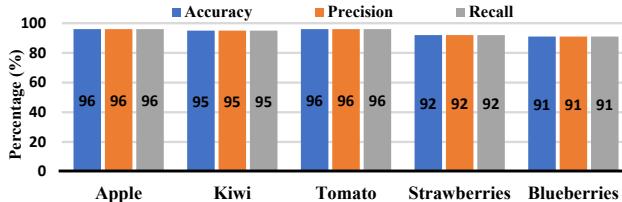


Figure 9: Accuracy of identifying organic fruits using spectral signatures in the 400–1000 nm range.

Second, light absorption, scattering, and reflection are affected by the chemical composition and biological structures of foods. For example, water, protein, and pigments absorb light at different wavelengths. Similarly, structural properties, such as density, particle size, and shape, affect the scattering of light. This means that various light wavelengths will undergo different scattering and absorption behavior as they reflect from biological materials. Since hyperspectral cameras provide detailed information across many wavelengths, they can potentially detect various *biomarkers* that identify organic foods in a non-destructive manner.

We start with a simple subjective analysis to explain the idea. We captured multiple hyperspectral images of organic and non-organic apples using a hyperspectral camera that operates in the 400–1000 nm range (the camera model is Specim IQ, and it captures 204 bands). We illuminated the scene using a halogen light source as recommended by the manufacturer of the camera. We plot the spectral signatures of two sample apples (one organic and one non-organic) in Figure 8 for illustration. Recall that a spectral signature is the amount of light energy reflected by an object across the different wavelengths in the spectral range for each pixel. Typically, the signature is averaged across multiple neighboring pixels for more robustness. In this example, the signatures are averaged across the rectangular areas shown in the left part of Figure 8, which are 8 by 8 pixels.

Two observations can be made on Figure 8. First, the signatures of organic and non-organic apples mostly overlap in the visible range, which is expected as the two apples have similar colors. Second, there are noticeable differences between the two signatures, especially beyond the visible range. Some critical areas are magnified in the two small sub-figure insets on the right. The 650–750 nm range is affected by the chlorophyll concentration, while the 900–1000 nm range captures the variations in the so-called C-H bond stretching, which relates to the sugar content in apples [34].

Next, we present a quantitative analysis of a large and diverse dataset. We captured 346 hyperspectral images of five fruits: apples, kiwis, tomatoes, strawberries, and blueberries. These fruits have different textures, colors, shapes, sizes, and

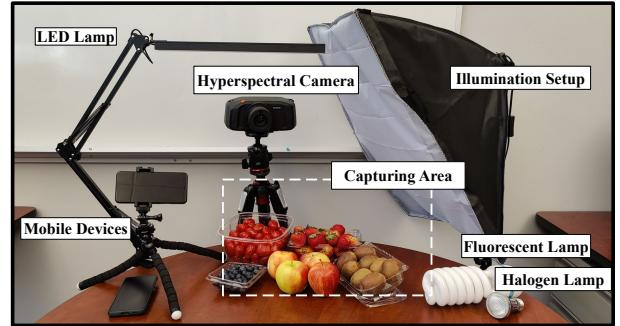


Figure 10: The setup used in our experiments.

chemical compositions. Half of the images were for organic fruits. All images were captured using our hyperspectral camera under ideal (halogen) illumination. Each image is a cube of 204 bands. More details are given in §5.1. We created spectral signatures, similar to the ones in Figure 8, for each hyperspectral image. We then designed and trained a simple neural network classifier to identify organic fruits using spectral signatures. The classifier is a Multilayer Perceptron (MLP) with three fully connected hidden layers that have 200, 150, and 100 neurons, respectively.

We plot the accuracy of classifying different organic fruits in Figure 9, which shows an accuracy between 91–96%. This accuracy is comparable to ones reported in prior works, e.g., [24, 28, 30], that used expensive spectrometers working in wider and more specialized spectral ranges. We note that the relatively lower accuracy for identifying organic strawberries and blueberries is mostly due to their small sizes.

In summary, the analysis in this section shows that hyperspectral imaging in the 400–1000 nm range can be used in identifying organic foods. However, the analysis was conducted with a high-end hyperspectral camera under perfect illumination conditions. MobiSpectral strives to achieve comparable performance using low-cost mobile phones working in common illumination conditions.

5 EVALUATION

We first show the accuracy of the reconstructed bands by MobiSpectral compared to the ground truth, which demonstrates the generality of MobiSpectral and the potential of using it in designing novel mobile applications. Then, we assess the accuracy of identifying organic foods using MobiSpectral. Finally, we analyze the performance impact of various components of MobiSpectral.

5.1 Experimental Setup and Datasets

Hardware Setup. The setup of our experiments is shown in Figure 10. We use a high-end hyperspectral camera (model: Specim IQ and costs about \$25K), which is a line-scanning



Figure 11: Samples from the mobile images dataset.

camera with a CMOS sensor operating in the 400–1000 nm range. It captures 204 bands with a spectral resolution of about 3nm. The camera is portable and has internal storage (32 GB) to save the captured data. It also has a mini display to help in focusing the camera lens before capturing. Captured images are later transferred to a workstation for processing. The camera comes with software to visualize the data.

In addition, we use two different mobile phones (Google Pixel 4XL and OnePlus 8 Pro) in the experiments as well as multiple diverse light sources, including halogen, LED, and fluorescent (CFL) sources.

Hyperspectral Images Dataset. We collected a total of 346 *hyperspectral images* of five different fruits using a halogen light source. These fruits have diverse characteristics in terms of textures, colors, shapes, sizes, and chemical compositions. Specifically, the dataset has images of: (i) 246 apples of various types (e.g., gala and ambrosia), (ii) 40 kiwis, (iii) 20 grape tomatoes, (iv) 20 strawberries, and (v) 20 blueberries. For each type of fruit, we captured 50% of the images for organic samples and the other 50% for non-organic ones. We captured each fruit individually and in groups of several pieces of the same type. In addition, we conducted our experiments over a long period of time (few months). Some fruits were kept for multiple weeks, and we took images on successive days. Further, the fruits were purchased from several stores and at different times throughout our study; that is, the fruits came from different sources/farms, and they had diverse freshness/ripeness levels. Thus, we believe our dataset is diverse and representative of realistic situations.

Each hyperspectral image has 204 bands, where each band is a grey-scale image with a resolution of 512 x 512 pixels. That is, this dataset has 70,584 bands of different fruits. This is a significant dataset in the hyperspectral imaging domain. Capturing hyperspectral images requires camera calibration and strict illumination and is time consuming; each image takes about 2 min to capture since the hyperspectral camera needs to scan the scene linearly (mechanically).

We synthesize the RGB image corresponding to each hyperspectral image using the software that comes with the camera. We then normalize the RGB images using the method in §3.5. The hyperspectral images dataset is used to train and

evaluate the accuracy of the reconstruction model, as it has *ground truth bands* for each associated RGB image.

Mobile Images Dataset. To evaluate the accuracy of identifying organic foods using MobiSpectral in real environments, we collected another dataset using the Google Pixel 4 XL phone. This dataset has pairs of RGB and NIR images. We note that we could not use other common phones, e.g., iPhones, as they currently do not allow public access to their NIR modules.

We consider multiple illumination settings while collecting this dataset: halogen, LED, fluorescent, and *arbitrary*. In the arbitrary setting, we do not control the light source at all. Rather we allow a mixture of sources, including the sunlight coming through the windows and the light bulbs installed in the ceiling of our lab. The arbitrary setting also contains images captured outdoors in direct sunlight as well as in the shade. In addition, we vary the capturing distance between 20 and 50 cm, as this is the operating range of the phone's NIR camera. Specifically, for each fruit sample, we choose an illumination setting and then capture it at multiple distances. Then, we repeat for a different illumination. Then, the whole experiment is repeated for another sample.

The mobile images dataset has a total of 900 *images* of: (i) 400 apples, (ii) 90 kiwis, (iii) 90 grape tomatoes, (iv) 126 strawberries, and (v) 124 blueberries. Half of the images are of organic fruits. Samples of this dataset are shown in Figure 11, demonstrating the diversity of objects, illumination settings, and capturing distances. The fruits in this dataset are *totally different* from the ones in the hyperspectral images dataset. The images in this dataset are normalized using the method in §3.5 to mitigate the effect of different illuminations. *This dataset is not used in training the reconstruction model.*

5.2 Accuracy of the Reconstructed Bands

We analyze the accuracy of the reconstructed bands by comparing them against the *ground truth* ones. For this comparison to be accurate/fair, the reconstructed and ground truth bands must be taken under the same conditions: camera sensor characteristics, lighting source, and capturing distance and angle. To achieve this, we use the hyperspectral camera, which produces actual hyperspectral bands as well as their corresponding RGB images. We apply our reconstruction model to those RGB images to reconstruct N bands. Then, we compare each of the N reconstructed bands with its corresponding ground truth using multiple metrics. For the experiments in this subsection and §5.3, we fix $N = 68$ bands, which means the distance between the reconstructed bands is only 9 nm, providing a fine-grained spectral resolution for most mobile applications. Previous spectral reconstruction works, e.g., [6], also used $N = 68$. In §5.4, we analyze the impact of varying N .

	MRAE	RMSE	SAM	SID	SSIM	PSNR
Apples	0.096	0.021	0.063	0.005	0.985	34.3
Kiwis	0.082	0.026	0.053	0.006	0.975	31.8
Blueberries	0.111	0.023	0.066	0.009	0.972	33.4
Average	0.096	0.023	0.060	0.006	0.977	33.2

Table 1: Average accuracy of the reconstructed bands relative to the ground truth.

We note that images captured by regular RGB cameras, e.g., ones on phones, do not have corresponding ground truth bands, and thus cannot be used in this comparison. Even if we were to capture the same scene with two cameras (phone and hyperspectral) simultaneously, the comparison would still be inaccurate. This is because the sensor sensitivity function, $C_n(\lambda_n)$, differs across cameras, and as shown by Equation (1), $C_n(\lambda)$ impacts the resulting pixel values. $C_n(\lambda)$ varies across cameras and manufacturers, and it is not publicly available.

We use the following performance metrics in the analysis, which are commonly used in hyperspectral imaging works. Due to space limitations and since they can easily be found in prior works, we remove the equations of these metrics.

- *Mean Relative Absolute Error (MRAE)*: measures the relative absolute error between the reconstructed and ground truth bands.
- *Root Mean Square Error (RMSE)*: measures the second-order error between the reconstructed and ground truth bands.
- *Spectral Angle Matching (SAM)*: measures the difference between two spectral signatures by computing the angle between them in the geometrical space [21].
- *Spectral Information Divergence (SID)*: measures the difference between the probability distributions of two spectral signatures [10].
- *Peak Signal to Noise Ratio (PSNR)*: measures the quality of the reconstructed bands relative to the ground truth.
- *Structural Similarity Index (SSIM)*: measures the similarity of the reconstructed bands to the ground truth.

Performance on Seen Fruits. We first consider training the reconstruction model only on three fruits: apples, kiwis, and blueberries. We divide the hyperspectral images of these three fruits into three *disjoint* partitions: 70% for training, 15% for validation, and 15% for testing.

We summarize the average results for all metrics in Table 1, which were computed on the testing partition of the dataset. The table shows that all error-related metrics (MARE, RMSE, SAM, and SID) are close to zero, which indicates the very close similarity of the reconstructed and ground truth bands. In hyperspectral imaging applications, SID and SAM are typically the most important metrics, because they measure

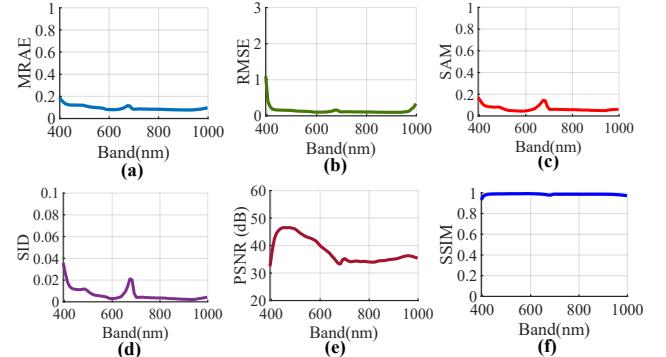


Figure 12: Accuracy of the reconstructed bands. Error metrics in (a)-(d) are close to zero for most bands. PSNR and SSIM metrics in (e) and (f) indicate good quality.

the similarity between the given two spectra. SID ranges between 0 and 1.0 and SAM between 0 to π . As the table shows, the average values of the SID and SAM metrics across all fruits are 0.006 and 0.06 radian (3.44 degrees), respectively. These are very small error values, which means that any hyperspectral application, e.g., classification or clustering, that uses our reconstructed bands will produce almost the same accuracy as if it were using bands captured by real hyperspectral cameras. Table 1 also shows that the SSIM metric is close to 1, which means that the reconstructed and ground truth bands have almost identical image structures. Further, the PSNR is consistently higher than 30 dB, indicating that the distortion level in the reconstructed bands is small.

We further analyze the accuracy of each band and present the results for all metrics in Figure 12. The figure confirms the high accuracy of all bands. We note that the PSNR values close to 400 nm (Figure 12.e) are relatively high because pixels in this range are quite dark, and both the reconstructed and ground truth bands do not show many details. We also note that bands around 400 nm and 700 nm have slightly higher error levels than other bands. This is because these bands are in the boundary regions in the spectrum: 400 nm is the transition from the ultraviolet (UV) range to the visible range, and 700 nm is the transition from the visible to the NIR range. And the reflected signals around these transition points are typically weak because of the sensor sensitivity.

Sample Visual Results. In Figure 13, we present samples of the reconstructed and ground truth bands and plot the difference (error) between corresponding bands as a heat map. The figure demonstrates the quality of the reconstructed bands and their similarity to the ground truth ones. The figure also shows that the error is fairly small, especially in the relevant areas that have objects (apples). Recall that the testing partition of the dataset may contain different numbers

	MRAE	RMSE	SAM	SID	SSIM	PSNR
Tomatoes	0.395	0.115	0.218	0.096	0.739	18.8
Strawberries	0.408	0.138	0.188	0.067	0.732	17.3
After Transfer Learning						
Tomatoes	0.063	0.015	0.043	0.004	0.983	36.9
Strawberries	0.124	0.026	0.073	0.011	0.959	31.8

Table 2: Accuracy of the reconstructed bands relative to the ground truth for two fruits not seen during the training, before and after transfer learning.

and arrangements of fruits than the training and validation partitions. Yet, the model was able to produce accurate bands.

Performance on Unseen Fruits. We test the reconstruction model on two fruits (tomatoes and strawberries) that were *not* seen at all during the training of the model. We present the results in the top part of Table 2, which shows that the error metrics are still small, although they are higher than the ones in Table 1. For example, the average values of SID and SAM are around 0.082 and 0.20 radian (or 11 degrees), respectively, which indicate that the produced spectral bands would still be useful for hyperspectral applications; our results on identifying organic fruits in §5.3 confirm this. The intuition behind why our reconstruction model worked on fruits that were never seen in training is that the *internal* structures of most fruits and vegetables are composed of similar chemical components, and thus they will likely have similar spectral characteristics.

We note that the PSNR and SSIM metrics are much lower in this case because tomatoes and strawberries have quite different shapes and colors than the fruits the model was trained on (apples, kiwis, and blueberries). PSNR and SSIM mostly model external features, which are less important in hyperspectral applications. The internal structures and compositions of fruits are captured by the variations across spectral bands, which are better modeled by SID and SAM. **Transfer Learning.** Next, we use transfer learning to fine-tune the reconstruction model on tomatoes and strawberries. We freeze all layers in the model except the last decoder layer and train only on a small number of images (20 tomatoes and 20 strawberries). The results, presented in the lower two rows of Table 2, show that all performance metrics have been significantly improved. That is, our model can easily be fine-tuned for new fruits to increase its accuracy further.

In summary, the results in this section show that our reconstruction model produces accurate hyperspectral bands, and it can easily be fine-tuned for new fruits.

5.3 Accuracy of Identifying Organic Foods

We developed a mobile application to identify organic foods and deployed it on two phones: Google Pixel 4 XL and OnePlus 8 Pro; a screenshot is shown in Figure 14.a. The NIR

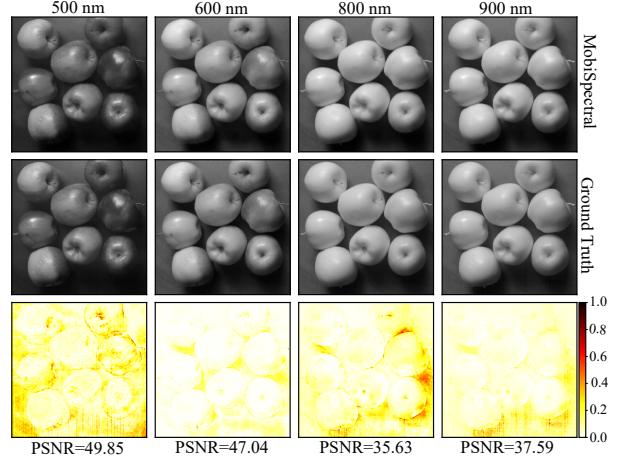


Figure 13: Similarity of the reconstructed and ground truth bands. The third row shows the differences between the corresponding images in the top two rows.

camera on Google Pixel is front facing, while on OnePlus, it is back facing. To facilitate capturing, especially with front-facing cameras, the application captures the RGB and NIR images after a short (3-sec) delay from clicking on the capture button. The application then shows a bounding box from which the spectral signatures will be created. Users can change the location of the box by tapping anywhere in the image. The area of the bounding box is divided into 16 equal size zones. We compute one spectral signature from each zone across all bands. The application processes all signatures and outputs Organic or Non-Organic. It also allows users to inspect bands and visualize spectral signatures.

To assess the accuracy of identifying organic fruits using MobiSpectral, we use the reconstruction model trained on images of only three fruits (apples, kiwis, and blueberries) from the hyperspectral images dataset. We test the accuracy using the mobile images dataset that was captured under diverse and realistic illumination conditions and at different distances. The mobile images dataset was not used in the training of the model and has totally different fruit samples.

We use the neural network classifier mentioned in §4.2, and we train it for each fruit. We divide the mobile images dataset into two partitions: 80% for training and 20% for testing. Then, we use 4-fold cross-validation to divide the training data into train-validation splits, and we run the classifier four times and compute the average precision, recall, and accuracy on the test data.

We report the accuracy in Figure 14.b for five different fruits. The results show that MobiSpectral is fairly accurate in all cases. Specifically, MobiSpectral can identify organic apples with an accuracy of 93% and kiwis with an accuracy of 94%, from images captured by regular phones. Although

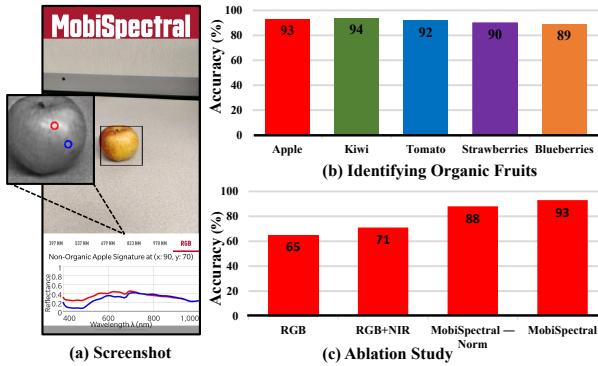


Figure 14: Analysis and performance of MobiSpectral.

blueberries are small and have irregular shapes, MobiSpectral was able to identify whether they are organic with an accuracy of 89%. In addition, MobiSpectral was not trained on organic tomatoes and strawberries. Yet, it was able to identify them with accuracies of 92% and 90%, respectively. As explained in §5.2, while the reconstructed bands have some visual distortions (shown by PSNR and SSIM) for the case of unseen fruits that have quite different *external* shapes and structures, MobiSpectral was able to create accurate spectral signatures (quantified by SID and SAM). And since the performance of applications mostly relies on spectral signatures, the accuracy of identifying organic tomatoes and strawberries was not significantly affected.

5.4 Analysis of MobiSpectral

Processing Time and Memory Usage. The average processing time (measured across more than 100 images) of MobiSpectral is 2.5 sec on Google Pixel and 2.2 sec on OnePlus Pro. Most of the time is spent in the spectral reconstruction step; all other steps took only a few msec. MobiSpectral used less than 300MB of memory on both phones.

Ablation Study. We analyze the performance contributions of the main components of MobiSpectral: spectral reconstruction and image normalization. The reconstruction model is trained to predict hyperspectral bands from the input RGB and NIR images, whereas image normalization transforms all images to a common reference to mitigate the effects of different capturing environments.

First, we do not use the reconstruction model at all. We only use the RGB and NIR images captured by the phone to identify organic apples. Then, we add the reconstruction model but do not use the image normalization step. Finally, we use all components of MobiSpectral. We also consider identifying organic fruits using only RGB images as a base reference. Our results, shown in Figure 14.c, confirm the importance of both the spectral reconstruction and image normalization components. The first provides the fine-grained

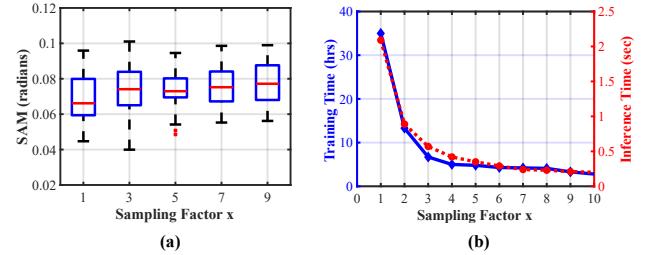


Figure 15: Generality of MobiSpectral in reconstructing different number of spectral bands.

bands from which we create accurate spectral signatures, which help in the identification process. The second adds robustness to the model by handling various illuminations.

Impact of capturing distance. The NIR camera has a limited operating range, typically between 10 and 50 cm. We analyze the impact of the capturing distance on the accuracy of the produced spectral signatures by MobiSpectral. We capture an apple at five distances: 10, 20, 30, 40, and 50 cm, and we create the spectral signature in each case. From our experiments, we found out that the ideal capturing distance for the NIR camera is around 30 cm. Thus, we use it as a reference distance. Then, we measure the deviations (using the SID metric) of signatures created at other capturing distances relative to the reference distance. Our results (the figure is omitted due to space limitations) show that the SID value is very small (between 0.02 and 0.06) for capturing distances in the range of 20 to 40 cm. The signatures are distorted when capturing at 10- and 50-cm distances, where the SID values are 6.75 and 4.18, respectively. This is because the strength of the NIR signal starts to decay rapidly beyond 50 cm, and when the phone is too close to the object (10 cm or less), it does not capture all reflected NIR signals.

Generality of MobiSpectral. MobiSpectral is designed to be flexible to support various applications. We demonstrate that MobiSpectral can accurately reconstruct different numbers of bands N , and we analyze the associated costs in each case. As mentioned before, our hyperspectral camera captures 204 bands, which represent the ground truth. We sample these bands with a factor x , where $x \in [1, 2, 3, \dots, 10]$. $x = 1$ means that we consider each band, that is, $N = 204$. When $x = 2$, we sample every other band, i.e., $N = 102$, and so on. Then, we train ten different reconstruction models, one for each value of N . We reconstruct the bands in each case and compare them against their corresponding ground truth ones using the SAM and SID metrics.

The results for the SAM metric are shown in Figure 15.a; SID results are similar. For each value of x , we display a boxplot summarizing the distribution of the SAM metric computed from all bands. We show only a few samples to

avoid cluttering the figure. The results demonstrate that MobiSpectral consistently produces accurate bands in all cases. For example, the maximum SAM is less than 0.1 radian (or 5.7 degrees) for all values of x .

Reconstruction models that produce a larger number of bands contain more trainable parameters, and thus they take longer to train. In Figure 15.b, we report the training and inference times for different values of x , which, as expected, decrease as the model reconstructs fewer bands.

Deployment Options and Limitations. MobiSpectral is designed for end users to assess the quality of their foods. Currently, it only supports verifying organic claims. But it can be extended for other applications, such as determining fat contents in meats. We also believe that MobiSpectral can be of interest to grocery stores, restaurants, and producers of organic foods. For example, grocery stores can perform fast analysis of samples from food shipments, whereas expensive methods can be used in suspicious cases. Restaurants that use organic products can quickly check the claims of new suppliers. Some organic food suppliers may promote the use of MobiSpectral to protect their products and fight cheating.

MobiSpectral makes use of NIR cameras, which exist in many phones, but not all of them. In addition, some manufacturers, e.g., Apple and Samsung, currently do not allow application developers to access their NIR modules. Furthermore, the reconstruction model in MobiSpectral requires hyperspectral images for training. This can partially be mitigated by using our pre-trained models and possibly fine-tuning them via transfer learning with a small number of images.

6 RELATED WORK

Mobile Applications. Several novel applications have been proposed for mobile devices capitalizing on their increasing processing and sensing capabilities, e.g., [22, 27, 38]. Mao et al. [22] present an acoustic imaging approach for mobile devices to help imaging in dark environments and in presence of obstructions. Prakash et al. [27] enable real-time augmented reality on mobile devices by estimating the illumination of the scene. Zhang et al. [38] use phones as part of a home-based rehabilitation system, utilizing various sensors for real-time analysis of human activities. The ideas proposed in this paper bring features of hyperspectral imaging to mobile devices, which can enable many more innovative applications on such devices.

Recently, multiple works proposed mobile applications to detect food fraud, e.g., [19, 32, 36]. LiquidHash [32] detects adulteration in bottled liquids by tracking the shape and movement of air bubbles that form inside the bottles. Vi-Liquid [19] identifies various liquids by measuring their viscosity coefficients using the built-in accelerometer in phones. CapCam [36] estimates the surface tension of liquids and

uses it to identify water contamination and alcohol concentration. These works rely on inspecting various physical properties of *liquids*, and they do not apply to solid foods.

Hyperspectral Imaging. Multiple prior works have considered various aspects of capturing and processing hyperspectral images. For example, Goel et al. present HyperCam [17], which captures and adaptively prioritizes frames based on the application’s needs. Arab et al. [3] capture multiple hyperspectral bands and use them with RGB images to improve the accuracy of face-based authentication systems. These works require special setups that are not possible for mobile applications. For example, HyperCam requires a special illumination source that has 17 LEDs in different wavelengths.

Recent works, e.g., [7, 18, 20, 25, 31], have considered obtaining some aspects of hyperspectral imaging without expensive hyperspectral cameras. However, unlike MobiSpectral, all prior works require additional hardware, change the camera system, and/or require extensive calibration. For example, Baek et al. [7] propose adding a prism in front of the lens of a regular RGB camera. Prisms are expensive and highly engineered components. Kim et al. [20] use color filters to capture hyperspectral images, which requires changing the camera hardware. He and Wang [18] reconstruct 16 bands from RGB images using the Wiener estimation method. But their method involves complex calibration using an expensive snapshot hyperspectral camera and reference color charts, and it assumes a fixed illumination setting.

7 CONCLUSIONS AND FUTURE WORK

Hyperspectral imaging systems offer rich information and enable many important applications. They, however, are expensive, complex, and require strict illumination conditions; thus, they are not widely used for end-user applications. We proposed MobiSpectral, which enables hyperspectral applications on phones without requiring any hardware changes. MobiSpectral takes regular RGB and near-infrared (NIR) images captured by phones and converts them to many hyperspectral bands covering the entire visible and NIR spectral range (400–1000 nm). We presented methods to make MobiSpectral robust and to function in real environments with limited and diverse illuminations. We demonstrated the accuracy of the produced bands by comparing them against ground-truth ones captured by a high-end hyperspectral camera for a large dataset. In addition, we developed a proof-of-concept prototype of MobiSpectral on Android and showed that it can identify organic foods with high accuracy.

MobiSpectral can potentially be used for many applications, including determining: fat contents in meat (e.g., for steak grading), ripeness degree of fruits (e.g., for selecting avocados), sugar and fiber contents of vegetables (e.g., for sorting potatoes), which we will consider in our future work.

REFERENCES

- [1] 2020. *Global organic market*. Retrieved December 2022 from <https://www.fibl.org/en/info-centre/news/global-organic-market-unprecedented-growth-in-2020>
- [2] Mahmoud Afifi and Michael S. Brown. 2020. Deep White-Balance Editing. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 1394–1403. <https://doi.org/10.1109/CVPR42600.2020.00147>
- [3] Mohammad Arab, Puria Moghadam, Mohamed Hussein, Wael Abdelmageed, and Mohamed Hefeeda. 2020. Revealing True Identity: Detecting Makeup Attacks in Face-based Biometric Systems. In *Proc. of ACM Conference on Multimedia (MM'20)*. Seattle, WA, 3568–3576.
- [4] Boaz Arad, Ohad Ben-Shahar, Radu Timofte, Luc Van Gool, Lei Zhang, and Ming-Hsuan Yang. 2018. NTIRE 2018 Challenge on Spectral Reconstruction from RGB Images. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'18) Workshops*.
- [5] Boaz Arad, Radu Timofte, Ohad Ben-Shahar, Yi-Tun Lin, and Graham Finlayson. 2020. The 2020 NTIRE Challenge on Spectral Reconstruction from RGB Images. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'20) Workshops*. Seattle, WA.
- [6] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Luc Van Gool, Shuai Liu, Yongqiang Li, Chaoyu Feng, Lei Lei, Jiaoqiao Li, Songcheng Du, Chaoxiong Wu, Yihong Leng, Rui Song, Mingwei Zhang, Chongxing Song, Shuyi Zhao, Zhiqiang Lang, Wei Wei, Lei Zhang, Renwei Dian, Tiansi Shan, Anjing Guo, Chengguo Feng, Jinyang Liu, Mirko Agarla, Simone Bianco, Marco Buzzelli, Luigi Celona, Raimondo Schettini, Jiang He, Yi Xiao, Jiajun Xiao, Qiangqiang Yuan, Jie Li, Liangpei Zhang, Taesung Kwon, Do-hoon Ryu, Hyokyoung Bae, Hao-Hsiang Yang, Hua-En Chang, Zhi-Kai Huang, Wei-Ting Chen, Sy-Yen Kuo, Junyu Chen, Haiwei Li, Song Liu, Sabarinathan, K Uma, B Satyha Bama, and S. Mohamed Mansoor Roomi. 2022. NTIRE 2022 Spectral Recovery Challenge and Data Set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 863–881.
- [7] Seung-Hwan Baek, Incheol Kim, Diego Gutierrez, and Min H. Kim. 2017. Compact Single-Shot Hyperspectral Imaging Using a Prism. *ACM Transaction on Graphics* 36, 6, Article 217 (Nov. 2017).
- [8] Marcin Barański, Dominika Średnicka Tober, Nikolaos Volakakis, Chris Seal, Roy Sanderson, Gavin B. Stewart, Charles Benbrook, Bruno Biavati, Emilia Markellou, Charilaos Giotis, and et al. 2014. Higher antioxidant and lower cadmium concentrations and lower incidence of pesticide residues in organically grown crops: a systematic literature review and meta-analyses. *British Journal of Nutrition* 112, 5 (2014), 794–811. <https://doi.org/10.1017/S0007114514001366>
- [9] Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Radu Timofte, and Luc Van Gool. 2022. MST++: Multi-stage Spectral-wise Transformer for Efficient Spectral Reconstruction. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 744–754. <https://doi.org/10.1109/CVPRW56347.2022.00090>
- [10] Chein-I Chang. 2000. An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis. *IEEE Transactions on Information Theory* 46, 5 (2000), 1927–1932.
- [11] Gamal ElMasry and Da-Wen Sun. 2010. Chapter 1 - Principles of Hyperspectral Imaging Technology. In *Hyperspectral Imaging for Food Quality Analysis and Control*, Da-Wen Sun (Ed.). Academic Press, San Diego, 3–43. <https://doi.org/10.1016/B978-0-12-374753-2.210001-2>
- [12] Georgios D. Evangelidis and Emmanouil Z. Psarakis. 2008. Parametric Image Alignment Using Enhanced Correlation Coefficient Maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 10 (2008), 1858–1865. <https://doi.org/10.1109/TPAMI.2008.113>
- [13] Chen Feng, Shaojie Zhuo, Xiaopeng Zhang, Liang Shen, and Sabine Süsstrunk. 2013. Near-infrared guided color image dehazing. In *Proc. of IEEE International Conference on Image Processing (ICIP'13)*. Melbourne, Australia, 2363–2367.
- [14] Qingtao Fu, Cheolkon Jung, and Chen Su. 2019. Color Recovery from Multi-Spectral NIR Images Using Gray Information. In *Proc. of ACM Multimedia Asia (MMAsia'19)*. Beijing, China, Article 16, 6 pages.
- [15] Ying Fu, Yongrong Zheng, Lin Zhang, and Hua Huang. 2018. Spectral reflectance recovery from a single rgb image. *IEEE Transactions on Computational Imaging* 4, 3 (2018), 382–394.
- [16] Francesca Giampieri, Luca Mazzoni, Danila Cianciosi, José M. Alvarez-Suarez, Lucia Regolo, Cristina Sánchez-González, Franco Capocasa, Jianbo Xiao, Bruno Mezzetti, and Maurizio Battino. 2022. Organic vs conventional plant-based foods: A review. *Food Chemistry* 383 (2022), 132352. <https://doi.org/10.1016/j.foodchem.2022.132352>
- [17] Mayank Goel, Eric Whitmire, Alex Mariakakis, T. Scott Saponas, Neel Joshi, Dan Morris, Brian Guenter, Marcel Gavriliu, Gaetano Borriello, and Shwetak N. Patel. 2015. HyperCam: Hyperspectral Imaging for Ubiquitous Computing Applications. In *Proc. of ACM Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'15)*. Osaka, Japan, 145–156.
- [18] Qinghua He and Ruikang Wang. 2020. Hyperspectral imaging enabled by an unmodified smartphone for analyzing skin morphological features and monitoring hemodynamics. *Biomedical Optics Express* 11 (2 2020), 895. Issue 2. <https://doi.org/10.1364/boe.378470>
- [19] Yongzhi Huang, Kaixin Chen, Yandao Huang, Lu Wang, and Kaishun Wu. 2021. Vi-Liquid: Unknown Liquid Identification with Your Smartphone Vibration. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking (New Orleans, Louisiana) (MobiCom '21)*. Association for Computing Machinery, New York, NY, USA, 174–187. <https://doi.org/10.1145/3447993.3448621>
- [20] Seoong Kim, Dongrae Cho, Jihun Kim, Manjae Kim, Sangyeon Youn, Jae Eun Jang, Minkyu Je, Dong Hun Lee, Boreom Lee, Daniel L. Farkas, and Jae Youn Hwang. 2016. Smartphone-based multispectral imaging: system development and potential for mobile skin diagnosis. *Biomedical Optics Express* 7, 12 (December 2016), 5294–5307.
- [21] F.A. Kruse, A.B. Lefkoff, J.W. Boardman, K.B. Heidebrecht, A.T. Shapiro, P.J. Barloon, and A.F.H. Goetz. 1993. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment* 44, 2 (1993), 145–163.
- [22] Wenguang Mao, Mei Wang, and Lili Qiu. 2018. AIM: Acoustic imaging on a mobile. In *Proc. of ACM International Conference on Mobile Systems, Applications, and Services (MobiSys'18)*. Munich, Germany, 468–481.
- [23] Puria Azadi Moghadam, Neha Sharma, and Mohamed Hefeeda. 2021. Enabling Hyperspectral Imaging in Diverse Illumination Conditions for Indoor Applications. (2021), 23–35. <https://doi.org/10.1145/3458305.3459594>
- [24] Gupta O, Das AJ, Hellerstein J, and Raskar R. 2018. Machine learning approaches for large scale classification of produce. 8, 1, Article 5 (March 2018). <https://doi:10.1038/s41598-018-23394-3>
- [25] Seoung Wug Oh, Michael S. Brown, Marc Pollefeys, and Seon Joo Kim. 2016. Do It Yourself Hyperspectral Imaging with Everyday Digital Cameras. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'16)*. Las Vegas, NV, 2461–2469.
- [26] Marco A. C. Potenza, Daniele Nazzari, Llorenç Cremonesi, Ilaria Denti, and Paolo Milani. 2017. Hyperspectral imaging with deformable gratings fabricated with metal-elastomer nanocomposites. *Review of Scientific Instruments* 88, 11 (2017).
- [27] Siddhant Prakash, Alireza Bahremand, Linda D. Nguyen, and Robert LiKamWa. 2019. GLEAM: An Illumination Estimation Framework for Real-Time Photorealistic Augmented Reality on Mobile Devices. In *Proc. of ACM Conference on Mobile Systems, Applications, and Services*

- (*MobiSys'19*). Seoul, Republic of Korea, 142–154.
- [28] D. Zhang Y. Ma R. Xiao, L. Liu and M. O. Ngadi. 2019. Discrimination of organic and conventional rice by chemometric analysis of NIR spectra: a pilot study. *Journal of Food Measurement and Characterization* 13, 1 (2019), 238–249.
- [29] Viktor Slavkovikj, Steven Verstockt, Wesley De Neve, Sofie Van Hoecke, and Rik Van de Walle. 2015. Hyperspectral Image Classification with Convolutional Neural Networks. In *Proc. of ACM International Conference on Multimedia(MM'15)*. Brisbane, Australia, 1159–1162.
- [30] Weiran Song, Hui Wang, Paul Maguire, and Omar Nibouche. 2016. Differentiation of organic and non-organic apples using near infrared reflectance spectroscopy – A pattern recognition approach. In *2016 IEEE SENSORS*. 1–3. <https://doi.org/10.1109/ICSENS.2016.7808530>
- [31] Mary Stuart, Andrew McGonigle, Matthew Davies, M. Hobbs, Nicholas Boone, Leigh Stanger, Chengxi Zhu, T.D. Pering, and Jon Willmott. 2021. Low-Cost Hyperspectral Imaging with A Smartphone. *Journal of Imaging* 7 (08 2021), 136. <https://doi.org/10.3390/jimaging7080136>
- [32] Bangjie Sun, Sean Rui Xiang Tan, Zhiwei Ren, Mun Choon Chan, and Jun Han. 2022. Detecting Counterfeit Liquid Food Products in a Sealed Bottle Using a Smartphone Camera. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services* (Portland, Oregon) (*MobiSys '22*). Association for Computing Machinery, New York, NY, USA, 42–55. <https://doi.org/10.1145/3498361.3539776>
- [33] Sabine Süsstrunk, Clément Fredembach, and Daniel Tamburrino. 2010. Automatic Skin Enhancement with Visible and Near-Infrared Image Fusion. In *Proc. of ACM International Conference on Multimedia(MM'10)*.
- Firenze, Italy, 1693–1696.
- [34] Tsuyoshi Temma, Kenkoh Hanamatsu, and Fujitoshi Shinoki. 2002. Measuring the Sugar Content of Apples and Apple Juice by Near Infrared Spectroscopy. *Optical Review* 9 (01 2002), 40–44. <https://doi.org/10.1007/s10043-002-0040-1>
- [35] Vanessa Vigar, Stephen Myers, Christopher Oliver, Jacinta Arellano, Shelley Robinson, and Carlo Leifert. 2020. A Systematic Review of Organic Versus Conventional Food Consumption: Is There a Measurable Benefit on Human Health? *Nutrients* 12, 1 (2020). <https://doi.org/10.3390/nu12010007>
- [36] Robert Xiao, Scott Hudson, and Chris Harrison. 2016. CapCam: Enabling Rapid, Ad-Hoc, Position-Tracked Interactions Between Devices. In *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces* (Niagara Falls, Ontario, Canada) (*ISS '16*). Association for Computing Machinery, New York, NY, USA, 169–178. <https://doi.org/10.1145/2992154.2992182>
- [37] Jonghee Yoon, Alexandru Grigoriu, and Sarah E. Bohndiek. 2020. A background correction method to compensate illumination variation in hyperspectral imaging. *PLOS ONE* 15, 3 (03 2020), 1–21. <https://doi.org/10.1371/journal.pone.0229502>
- [38] Hanbin Zhang, Gabriel Guo, Emery Comstock, Baicheng Chen, Xingyu Chen, Chen Song, Jerry Ajay, Jeanne Langan, Sutanuka Bhattacharjya, Lora A Cavuoto, and Wenyao Xu. 2020. RehabPhone: A Software-Defined Tool Using 3D Printing and Smartphones for Personalized Home-Based Rehabilitation. In *Proc. of ACM Conference on Mobile Systems, Applications, and Services (MobiSys'20)*. Toronto, Canada, 434–447.