



Research article

Intelligent synthesis of hyperspectral images from arbitrary web cameras in latent sparse space reconstruction

Yenming J. Chen¹, Jinn-Tsong Tsai^{2,3}, Kao-Shing Hwang^{2,4}, Chin-Lan Chen^{5,*} and Wen-Hsien Ho^{2,6,7,*}

¹ Department of Information Management, National Kaohsiung University of Science and Technology, Taiwan

² Department of Healthcare Administration and Medical Informatics, Kaohsiung Medical University, Taiwan

³ Department of Computer Science and Artificial Intelligence, National Pingtung University, Taiwan

⁴ Department of Electrical Engineering, National Sun Yat-Sen University, Taiwan

⁵ Department of Family Medicine, Yuan's General Hospital, Taiwan

⁶ Department of Medical Research, Kaohsiung Medical University Hospital, Taiwan

⁷ College of Professional Studies, National Pingtung University of Science and Technology, Taiwan

* **Correspondence:** Email: 900pchinlan@gmail.com; whho@kmu.edu.tw.

Abstract: Synthesizing hyperspectral images (HSI) from an ordinary camera has been accomplished recently. However, such computation models require detailed properties of the target camera, which can only be measured in a professional lab. This prerequisite prevents the synthesizing model from being installed on arbitrary cameras for end-users. This study offers a calibration-free method for transforming any camera into an HSI camera. Our solution requires no controllable light sources and spectrometers. Any consumer installing the program should produce high-quality HSI without the assistance of optical laboratories. Our approach facilitates a cycle-generative adversarial network (cycle-GAN) and sparse assimilation method to render the illumination-dependent spectral response function (SRF) of the underlying camera at the first part of the setup stage. The current illuminating function (CIF) must be identified for each image and decoupled from the underlying model. The HSI model is then integrated with the static SRF and dynamic CIF in the second part of the stage. The estimated SRFs and CIFs have been double-checked with the results by the standard laboratory method. The reconstructed HSIs have errors under 3% in the root mean square.

Keywords: spectral response function (SRF); current illuminating function (CIF); hyperspectral images (HSI); dichromatic reflection model (DRM); cycle-generative adversarial network (GAN); sparse ensemble assimilation (SEA)

1. Introduction

Hyperspectral images (HSI) can be derived from ordinary RGB cameras [1–6]. However, existing methods require detailed optical properties of the underlying camera, which can only be measured in a laboratory with expensive equipment, preventing ordinary users from using a plug-and-play webcam immediately.

This study proposes a calibration-free and illumination-independent method, as shown in Figure 1, to facilitate ordinary users' instant use of an arbitrary camera. Our algorithm can transform an ordinary webcam into an expensive HSI camera without the help of additional hardware. Mathematically, the forward transformation, mapping from high-dimensional HSI images to low-dimensional RGB images, is relatively easy, compared to the reverse one, mapping from low- to high-dimensional.

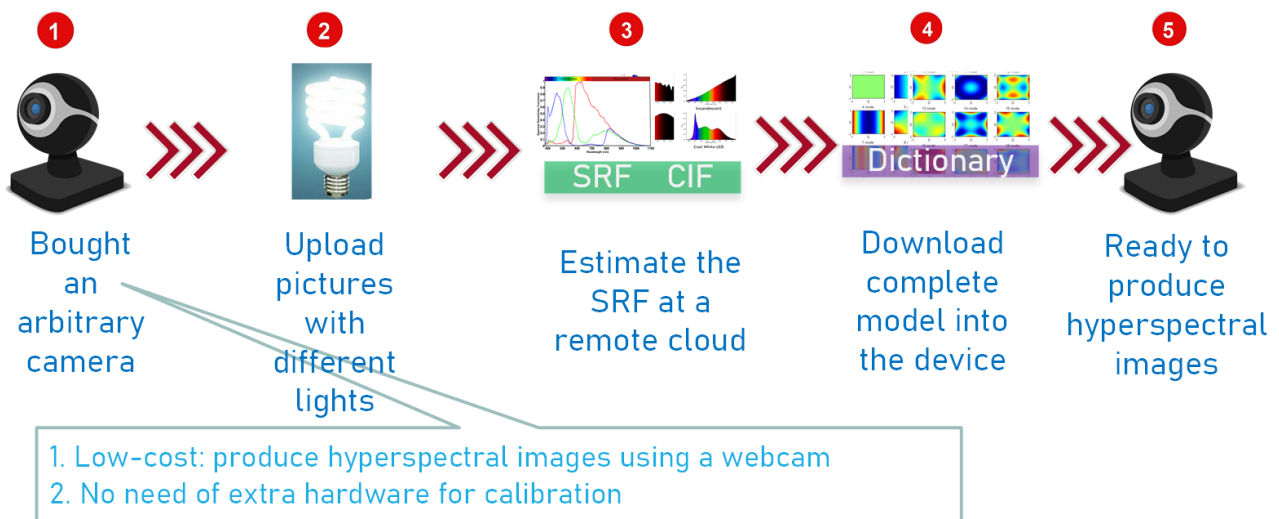


Figure 1. The setup and usage scenarios for any camera.

However, both the forward and reverse transformations involve a device-dependent response function for the underlying camera. In previous methods, the camera and light source must remain the same, and the spectral response function (SRF) of a camera and the current illuminating function (CIF) of the ambiance must be identified separately by specific equipment, such as standardized light sources, color-checkers, and spectrometers. The images used in the training database must match the same camera and light source. Such limitations prevent the algorithm from being plug-and-play.

Hyperspectral technology has been widely used in different fields [7]. Although HSI imaging possess tremendous advantages in wide fields of applications, the extremely high acquisition cost (US\$40 thousand or above) limits the suitable applications and usage scenarios. Existing computation methods converting RGB to HSI can significantly broaden the applicability of HSI.

In this study, a semi-finished model, independent of devices and illumination, is shipped with the

installation package, as shown in the Model part of Figure 1. Before use, a setup step is performed without additional hardware to extract the SRF of the underlying camera, as shown in step 1 in Figure 1. During each usage, the ambient CIF must be estimated for each taken image, as shown in step 2 in Figure 1. The final reconstruction with SRF and CIF then projects RGB images to the device and illumination-independent HSI.

To accumulate sparse projection kernels of hyperspectral signatures in terms of many hyperspectral priors, we render the camera SRF through a most probable algorithm. As shown in Figure 2, at the setup step, two parametric tasks are accomplished automatically without the involvement of users. SRF and CIF are estimated by uploading random images taken by the target camera to the cloud. There is no need to involve a calibration step for ordinary users. We develop a triangular generative adversarial assimilation comprising two different methods in supervising and unsupervised learning, each with advantages, pulling each other up to make a high-precision reconstruction.

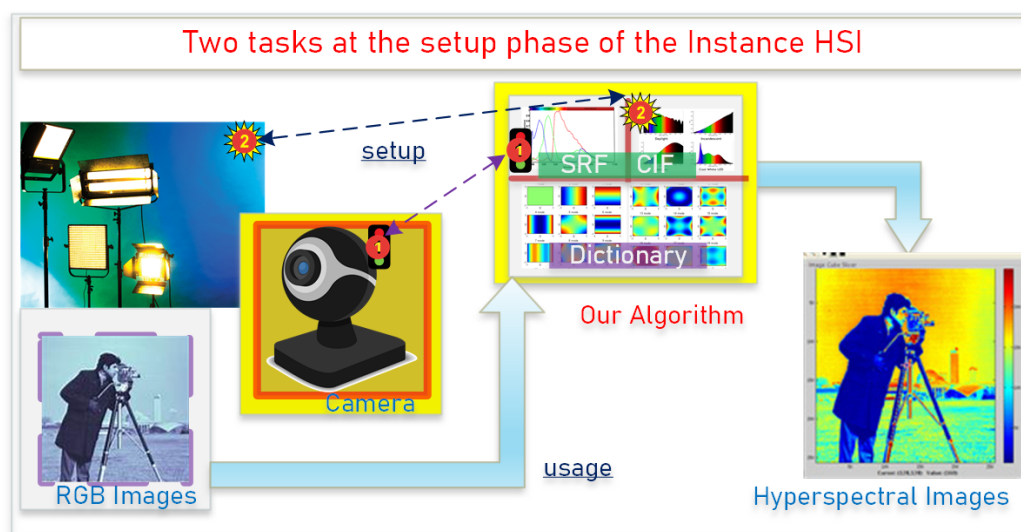


Figure 2. During setup, two parametric tasks are accomplished automatically without the involvement of users. SRF and CIF are estimated by uploading random images taken by the target camera to the cloud.

2. Relation to prior research

Despite the extensive studies in reconstructing HSI from RGB images, numerous challenges remain unsolved. This study seeks to close three research gaps. Existing studies require a laboratory-calibrated camera, and the training is only bound to that specific camera. Our results contribute to the online retrieval of SRF and the SRF independent training offline.

Much research has contributed to reverse mapping from the low-dimensional RGB space to the high-dimensional hyperspectral space [1, 4, 6, 8, 9]. However, gaps still exist.

The essence of the forward problem is a spatial convolution between incident image spectral and camera optics [10]. The properties of camera spectral response functions have been studied extensively [11]. The reverse mapping in algebraic or iterative approximation methods is sufficient as long as the camera response function can be correctly acquired before applying the algorithms [4, 6]. The reverse

mapping in kernel representation is prevalent and stable [1, 8, 9]. Statistical inference methods have been reported effective in enhancing the accuracy of the reverse mapping [2, 3].

With the assistance of additional hardware, the recovered image can be more accurate [5, 12–14]. However, such assistance does not fit the goal of this study. Therefore, the last piece of the puzzle for the HSI reconstruction is still acquiring camera response functions without the help of expensive instruments.

The second research gap leads to a challenge in algorithm design. Because the central part of the forward problem involves the convolution in the camera SRF, the training algorithm may only retrieve the function mapping if the SRF is determined at the training phase. On the other hand, estimating the SRF is difficult if we do not have the pairs of images with input RGB and output HSI [15]. Our problem is to retrieve the HSI without training pairs and a system model. The original generative adversarial network (GAN) requires a pair of images (x, y) for discriminator D and generator G to evaluate $\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x|y)] + E_{z \sim p_z(z)} [\log 1 - D(G(z|y))]$. Thanks to modern progress of the maximum likelihood theorems and the sparse variation of the GAN, “double unknown retrieval” of those above has become possible [16–18]. One of the advantages of cycleGAN over standard GANs is that the training data need not be paired. This implies that, in our situation, only the images (RGB, HSI) are sufficient for estimating the output SRFs. The idea of conditional and cyclic adversarial networks with image-to-image translation has successfully solved the problem of double unknown retrieval [19].

When the training data are not completely paired, the latent space reconstruction becomes the key part of the double unknown retrieval [20–22]. Convolution particle filters have been proven effective in estimating unknown parameters in state-space [23].

The generative network synthesizes the SRFs based on a predictive distribution conditioned on the correctness of the recovery of the original RGB values [19, 24]. When the projection to the latent space is not linear, some studies apply advanced algorithms to address the nonlinearity issue [25].

The third gap in the challenges of HSI reconstruction is the interference of versatile illumination conditions. Many studies tried to decouple the influence of illuminations. Despite the simplicity, an accurate spectral reflectance reconstruction must come with a spectral estimation of illumination [13]. Some solve the issue at the tristimulus value space (RGB or CIE XYZ) by collecting images of multi-conditional illumination [26–28]. On the other hand, CIE XYZ starts from the spectral response of illumination; and, nevertheless, they primarily focus on the color appearance for our bare eyes, not the spectral reflectance of objects. Therefore, a research gap exists in decoupling the influence of unknown illumination for spectral response estimation algorithms.

The sparse assimilation algorithm can reconstruct the parameters from limited observations [29]. However, most iterative algorithms are computationally intensive. Thus, they are not suitable for applying online and a tractable algorithm is needed for computational feasibility. We use hierarchical Bayesian learning and Metropolis-Hastings algorithms in the algorithm [30, 31] to estimate joint probability densities. Therefore, this study exploits an assimilation method adapted to online computation.

3. Methods

3.1. Problem description

As shown in Figure 3, a camera with response function SRF $R_\phi(\lambda, i)$ takes a spectral reflectance input $h(j, \lambda)$ for wavelength $\lambda \in [1, N]$ and produces the tristimulus values $c(j, i)$ for $i \in [RGB]$ or $[1, 3]$, and the $j \in [1, n]$ is the j -th dataset pair in the database. The 24-patches' color-checker pixels have been arranged into a column vector for ease of dimension expression.

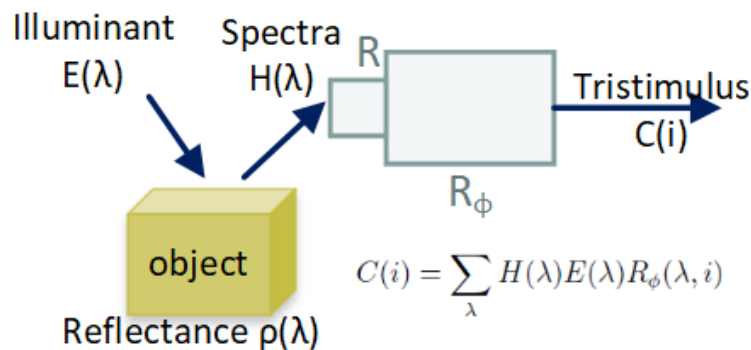


Figure 3. The forward projection, dichromatic reflection model (DRM), from object reflectance to the RGB values.

The forward problem is stated as a dichromatic reflection model (DRM)

$$\begin{aligned}
 h(\lambda) &= E(\lambda)\rho(\lambda), \forall \lambda, \\
 c(j, i) &= \sum_{\lambda} h(j, \lambda)R(\lambda, i), \forall j, i,
 \end{aligned} \tag{3.1}$$

or,

$$\mathbf{C}_{n \times 3} = \mathbf{H}_{n \times N} \mathbf{R}_{N \times 3}, \text{ for the } j\text{-th image in the database.} \tag{3.2}$$

where the matrix form is a collection of the indexes such that $\mathbf{C}_{n \times 3} = \{c(j, i)\}_{j \in [1, n], i \in [1, 3]}$.

By the representation theorem, we can find coefficients $\tilde{w}(k, k')_{k, k' \in [1, m]} = \tilde{\mathbf{W}}$ for taking m basis such that the RGB images \mathbf{C} can be approximated by

$$\mathbf{C}_{n \times 3} \approx \tilde{\mathbf{W}}_{n \times m} \mathbf{\Psi}_{m \times N} \mathbf{R}_{\phi N \times 3}. \tag{3.3}$$

The DRM (3.2) becomes

$$\tilde{\mathbf{W}} \mathbf{\Psi} \mathbf{R}_{\phi} = \mathbf{H} \mathbf{R}_{\phi} \tag{3.4}$$

Therefore,

$$\mathbf{H} = \tilde{\mathbf{W}} \mathbf{\Psi} \tag{3.5}$$

As discussed in the previous section, many methods exist to estimate the ill-posed back projection, but challenges still exist. Our problem is an inverse one. Given the tristimulus values \mathbf{C} , we want to reconstruct the \mathbf{H} . Because the dimension of \mathbf{H} is much higher than that of \mathbf{C} , the inverse project is ill-posed. Our method requires no pre-built calibration and prevents over-fitting after training.

3.2. Retrieving CIF from the white block

The CIF is estimated from the images of the white block of the color checker. The illumination light e is converted to electric signals through the camera response e_ϕ , and, therefore, $e(\lambda) = \sum_i e_\phi(\lambda, i)$. Utilizing the above-mentioned method, the CIF can be decomposed by a set of kernel functions. We have

$$e(\lambda) = \sum_i e_\phi(\lambda, i) = \sum_{i,k} \gamma_{i,k} b_k(\lambda, i), \forall \lambda \in [1, N], \quad (3.6)$$

where $b_k(\lambda)$ is the k -th basis for the CIF, and $\gamma_{i,k}$ is the coefficient.

3.3. Retrieving SRF by cycleGAN

Classical generative adversarial network applications use neural networks entirely for classification and regression applications. Such modeling suffers from the convergence problem. In our method, we use optical models as generators and neural networks as discriminators to maximize the knowledge of prior model information for the stochastic process of Lambert reflectance (Eq 3.2), which is implemented through a Monte Carlo radiative simulation. Through collected priors, the discriminator can distinguish between successful or unsuccessful generations. Therefore, the cyclic iteration can converge fast and accurately estimate the latent sparse space with sparse converging metrics.

We hybridize a statistical generator and neural network discriminator to maximize the usage of prior model information for the stochastic process. Our generator can also take environment conditions as covariates in the random process and is robust to anti-symmetric station distribution.

A standard GAN comprises a generator G and a discriminator D [32]. The model G and D can be neural networks or any mathematical functions, as long as the optimization (3.7) with Lagrangian η has solutions.

$$G^* = \arg \min_G \max_D \mathcal{L}_{GAN}(G, D) + \eta \mathcal{L}_1(G). \quad (3.7)$$

The standard loss functions have the form

$$\mathcal{L}_{GAN}(G, D) = E_{Z,Z'} [\log D(Z, Z')] + E_{Z,Z1} [\log(1 - D(Z, G(Z, Z1)))], \quad (3.8)$$

$$\mathcal{L}_1(G) = E_{Z,Z',Z1} [\|Z' - G(Z, Z1)\|_1]. \quad (3.9)$$

The GAN used in this study is a type of cycleGAN, which contains 2 generators and 2 discriminators, as shown in Figure 4. Our cycleGAN takes two sets of images as input and obtains the output containing the corresponding SRF. The observation Z contains the pair $C_3 = (R, G, B)$ and H_{33} 33-band HSI images. The subscripts 3 and 33 represent the number of bands for the variables, and they can be ignored if the meaning is clear. The latent estimation represents the SRF $Z' = (R'_{33}, H'_{33}, C'_3)$. In the target domain, the Z' is compared to a small set of collected SRFs $Z1 = R1_{33}$, which are only served as shape templates. The re-estimated image pairs are $Z'' = (C''_3, H''_{33})$, where $C''_3 = H'_{33} R'_{33}$, H'_{33} is the coefficients of $G_{ZZ'}$, and H''_{33} is the direct output of $G_{Z'Z}$.

cycleGAN: $Z \rightarrow Z' \rightarrow Z''$

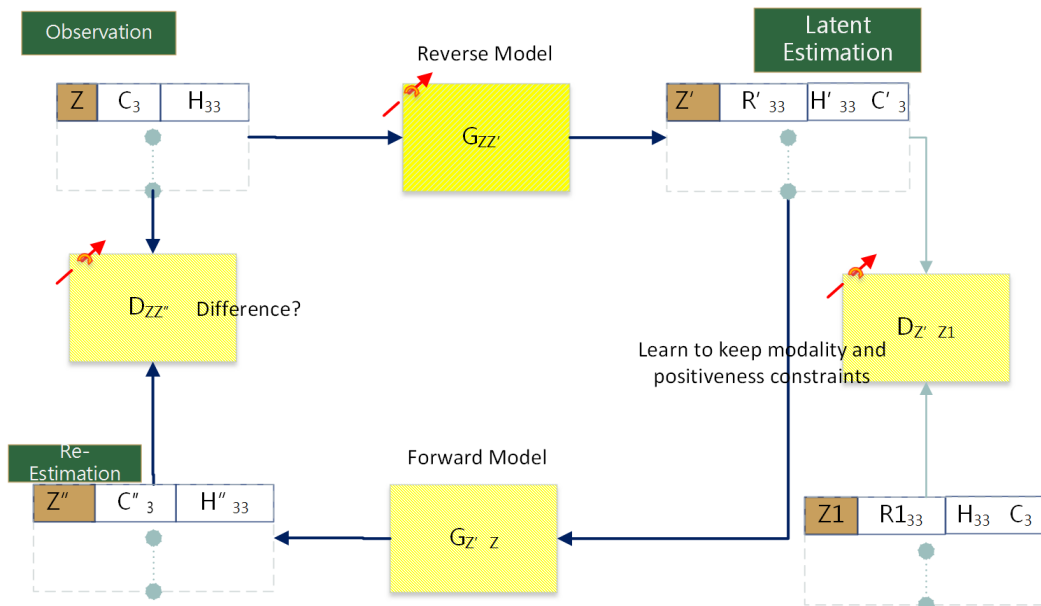


Figure 4. The generator and discriminator design in the cycleGAN. Systemically, we want to derive SRFs from the image pairs. Therefore, the observation Z contains the pair $C_3 = (R, G, B)$ and H_{33} 33-band HSI image. The latent estimation represents the SRF $Z' = (R'_{33}, H'_{33}, C'_3)$. In the target domain, the Z' are compared to a small set of collected SRFs $Z1 = (R1_{33}, H_{33}, C_3)$, where the $R1_{33}$ are only served as shape templates. The re-estimated image pairs are $Z'' = (C''_3, H''_{33})$, where $C''_3 = H'_{33}R'_{33}$, H'_{33} is the output of $G_{ZZ'}$, and H''_{33} is the output of $G_{Z'Z}$. The models in the cycleGAN are $G_{ZZ'}$, $D_{Z'Z1}$, $G_{Z'Z}$ and $D_{ZZ''}$, respectively. When the cycleGAN converges, the error between Z and Z'' will be minimized.

The models in the cycleGAN are $G_{ZZ'}$, $D_{Z'Z1}$, $G_{Z'Z}$, and $D_{ZZ''}$, respectively. The cycleGAN model takes C_3 and H_{33} as input dataset. We first prepare a set of (RGB,HSI) pairs for different color patches and a known camera (e.g., CIE 1964). Because we expected the same color patch to produce the same HSI image through different cameras, we prepared additional image pairs by taking the image of the target unknown camera as RGB and the HSI of the known camera as the HSI. We also need small samples of normal SRFs $Z1 = R1_{33}$ as a target template to prevent multiple solutions that are dissimilar to normal SRFs.

The generator $G_{ZZ'}$ is a reverse model, which takes $Z = (H, C)$ in the source domain and outputs $Z' = (R', H', C')$ in the target domain. During the generation, $G_{ZZ'}$ synthesizes SRFs $Z' = R'$, which is equivalent to applying a perturbation Δ such that $R' = R\Delta$ without violating the modality (Eq 3.10) and positiveness (Eq 3.11) constraints.

$$\begin{cases} R_i(\lambda_{k+1}) > R_i(\lambda_k), & k = 1, \dots, m_i - 1, \\ R_i(\lambda_{k+1}) \leq R_i(\lambda_k), & k = m_i, \dots, 33, \end{cases} \quad (3.10)$$

$$R_i(\lambda_k) \geq 0, \quad k = 1, \dots, 33, \quad (3.11)$$

where $i = \{1, 2, 3\}$ is the RGB channel, m_i is a predefined mode number (peak location) for channel i , and λ_k is the wavelength at each band. To integrate the constraints (3.10) to the loss function of the

first discriminator $D(Z', Z1)$, the constraints can be written in a form of violation score.

$$\zeta_d = \sum_{i=1,2,3} \sum_{k=1,\dots,33} \text{sgn}(k - m_i) \{R_i(\lambda_{k+1}) - R_i(\lambda_k)\} - R_i(\lambda_k), \quad (3.12)$$

where $\text{sgn}(s) = \{1, -1\}$ if $s > 0$ and $s \leq 0$, respectively.

The synthesized SRFs Z' will be rejected by $D_{Z'Z1}$ if they are dissimilar to normal SRFs $Z1 = R1_{33}$. The discriminator $D_{Z'Z1}$ is directly formed by a residual neural network (ResNet34). The network adds some jump connections that skip internal layers to avoid the vanishing gradient problem and the accuracy saturation problem. The construction of ResNet repeats a fixed pattern several times in which the strided convolution downsampler jumps, bypassing every 2 convolutions.

The $G_{Z'Z}$ is the forward model, which generates the re-estimated image pairs $Z'' = (C''_3, H''_{33})$. The C''_3 is directly applied to the estimated SRF R' from the high-dimensional H''_{33} . The re-estimated RGB images are computed as $C''_3 = H''_{33}R'_{33}$, where H''_{33} is the one of the output of $G_{ZZ'}$. The re-estimated HIS images H''_{33} are the output of $G_{Z'Z}$ taking input from C''_3 , another output of $G_{ZZ'}$.

Finally, a ResNet34 $D_{ZZ''}$, which takes the original images Z and the re-estimated images Z'' , rejects those having large errors. The discriminator $D_{Z'Z1}$ uses probability-based loss function, and $D_{ZZ''}$ uses a simple loss function in the mean squared error (mse) between the expected and predicted outputs.

The loss functions for the two discriminators are expressed in scores

$$D(Z', Z1) = (p_d - 1)^2 + p_d^2 + \eta_1 \frac{(R' - R1)^2}{R'} + \eta_2 \zeta_d, \quad (3.13)$$

$$D(Z, Z'') = \frac{(C - C'')^2}{C} + \frac{(H - H'')^2}{H}, \quad (3.14)$$

where p_d is the probability output of the discriminator $(Z', Z1)$, and η_1 and η_2 are the scaling factors controlling the involvement of the similarity and constraints, respectively. When the cycleGAN converges, the error scores (3.13) and (3.14) between Z' and $Z1$, Z and Z'' , respectively, will be minimized (The convergence trend of the scores can be referred in Figure 9).

3.4. Reconstructing HSI by back projection

To increase the accuracy of back projection, we use a kernel method to decompose the HSI and tristimulus RGB images. The estimated projection over the kernels can further minimize the reconstruction errors [33]. We decompose the HSI in the DRM (3.2) into a set of kernels $\psi(\lambda, \lambda')_{\lambda=1,\dots,N,\lambda'=1,\dots,N}$,

$$h(x, j, \lambda) = \sum_{\lambda'} \tilde{H}(x, j, \lambda) \psi(\lambda, \lambda'), \forall x = [1, n], j = [1, m],$$

or

$$\mathbf{H}(j)_{n \times N} = \tilde{\mathbf{H}}(j)_{n \times N} \mathbf{\Psi}_{N \times N}. \quad (3.15)$$

At the training stage, images are indexed by $j \in [1, m]$. We aim to find a set of kernels $\tilde{\mathbf{H}}(j) = \{\tilde{h}(x, j, \lambda)\}_{x \in [1, n], \lambda \in [1, N]}$ that can span the tristimulus space.

$\tilde{\mathbf{H}}$ can be estimated by Yule-Walker equation [34] or a Nadaraya-Watson kernel estimator [35]. The kernels $\mathbf{\Psi}$ are chosen in the reproducing kernel Hilbert space, guaranteeing that the basis vectors exist [33].

By the representation theorem, the tristimulus values can be approximated by the transformed kernels $\sum_{\lambda} \tilde{h}(x, j, \lambda)R(\lambda, i)$ and the coefficients $\tilde{w}(x, x')_{x, x' \in [1, nm]} = \tilde{\mathbf{W}}$ such that

$$C(x, j, i) \approx \sum_{\lambda} \tilde{w}(x, x) \tilde{h}(x, j, \lambda) R(\lambda, i), \text{ for } x \in [1, n], j \in [1, m], i \in [1, 3]. \quad (3.16)$$

We decompose an RGB image by approximating

$$\mathbf{C}(j)_{n \times 3} \approx \tilde{\mathbf{W}}(j)_{n \times N} \mathbf{\Psi}_{N \times N} \mathbf{R}_{N \times 3}. \quad (3.17)$$

Together with the DRM (3.2), this implies that

$$\mathbf{H}(j)_{n \times N} \approx \tilde{\mathbf{W}}(j)_{n \times N} \mathbf{\Psi}_{N \times N}. \quad (3.18)$$

Because the back projection $\tilde{\mathbf{W}}$ may not align with the direction of the covariance of $\mathbf{\Psi}(j)\mathbf{R}$, we need to make additional assumptions in l_1 space. Assuming the tristimulus space is far smaller than one of the HSI space, the sparse partial least square regression partitions the space into two disjoint subspaces such that $\mathbf{H} = (\mathbf{H}_1, \mathbf{H}_2)$ spanned by relevant (\mathbf{H}_1) and irrelevant variables (\mathbf{H}_2) [36,37]. Such partition effectively isolates uncorrelated basis in the latent space.

We aim to find a set of coefficients that maximize the span of hyperspectral space and also try to avoid over-fitting. The goal is that

$$\max_{\mu} \min_{\tilde{\mathbf{W}}} \|\mathbf{C} - \tilde{\mathbf{W}}\mathbf{\Psi}\mathbf{R}_{\phi}\|^2 + \mu \left\{ \frac{1-\alpha}{2} \|\tilde{\mathbf{W}}\|_2^2 + \alpha \|\tilde{\mathbf{W}}\|_1 \right\}. \quad (3.19)$$

The challenge of the ill-posed problem is to invert a near-singular problem. The regularization methods in (3.19) with a positive perturbation in l_1 can suppress over-fitting effectively.

To avoid over-fitting, we further generalize the problem with an elastic net (3.19) in the Least Absolute Shrinkage and Selection Operator (Lasso) term [38] with Lasso penalty α with l_1 and l_2 norms, $\|\cdot\|_1, \|\cdot\|_2$.

Optimization (3.19) pushes the coefficients to zero if the covariates are insignificant due to the l_1 properties. The reconstruction efficiency increases when transformation manifests the sparsity properties [39]. The optimization (3.19) is regulated by the Lasso penalty ($\alpha = 1$) or the ridge penalty ($\alpha = 0$), and it takes advantage of the sparse l_1 norm in evaluating solutions for the ill-posed problem [40].

If $\alpha = 0$, the optimization in (3.19) reduces to an ordinary generalized matrix inverse, which serves as a comparison basis. The least-square estimation in the objective creates a significant variance when covariates exhibit multicollinearity.

Ridge regression performs optimization to compensate for the multicollinearity problem by finding a balance between variance and bias [41]. The ridge penalty effectively reduces the variance of the identified coefficients [42, 43]. The experiments demonstrate that the lasso and ridge estimation effectively reconstruct the HSI without over-fit.

Due to the properties of the l_1 space, an optimization (e.g., $\min_x \|x\|_1$, s.t. (3.19)) should possess minimal non-zero solutions, and yield strong reconstruction performance if several sparsity properties, such as restricted isometry and incoherence properties, are satisfied [39, 40, 44]. In our multivariate transformation matrix Ψ_i , the lasso penalty tends to reduce the coefficients of less important covariates to zero, thus generating more zero solutions and fitting our assumption about the hyperspectral space.

The ill-posed back projection from \mathbf{C} to \mathbf{H} still contains errors. We propose a machine learning model to further close the gap. In the training step, we aim to solve a model g such that the error

$$\|\mathbf{H}_j - g(\tilde{\mathbf{W}}_j)\Psi\|_2 \quad (3.20)$$

is minimized. This study employs an ensemble of regression learners, including random forest regression, and supports vector regression.

At the query stage, we are ready to retrieve the HSI from the decomposition matrix $\tilde{\mathbf{W}}$ of the RGB images \mathbf{C}_q .

$$\mathbf{H}_q = g(\tilde{\mathbf{W}}\Psi) \quad (3.21)$$

4. Analysis

Our method is robust to a wide range of SRFs. An experiment is designed to prove our algorithm. The reconstructed HSIs by our method are almost identical between the measured SRF and the estimated SRF. Despite the quality of the SRF, the reconstructed HSIs are always of high quality.

4.1. Retrieving algorithm in standard conditions

We first apply our algorithm in the standard CIE 1964 camera spectral response function (Figure 5 (a)). Our algorithm uses 33 bands from 400 nm to 720 nm. Based on the normal visible spectrum range, the 3-color images have been transformed into 33-color images. The reconstructed hyperspectral images match the original ground truth HSI closely (Figure 5 (b)).

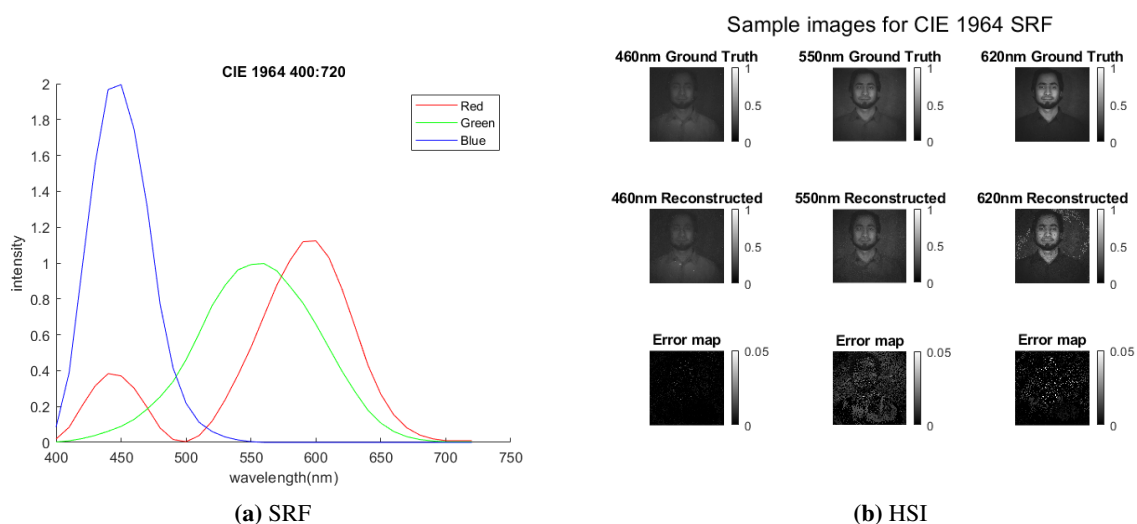


Figure 5. (a) The standard CIE 1964 camera spectral response function. (b) A sample of reconstructed HSIs in 460, 550 and 620 nm bands using the standard CIE 1964 SRF. The entire HSI contains 33 bands from 400 nm to 720 nm. The errors between the original and the reconstructed images are insignificant according to the visual representation. Few pixels have errors approximating to 0.05. Please refer to Table 1 for the summarized root mean square errors. (image source: [45])

4.2. With known SRF and CIF (requiring a spectrometer)

To evaluate the effectiveness of our method, we designed the experiments to compare the results of standard laboratory camera calibration and ambient light conditions. We bought a low-cost camera (under US\$30) from the Internet and set-up for laboratory calibration. The environment should mimic the one that consumer users have. A Macbeth color-checker with 24 patches, in Figure 6 (c), was used. The spectral response (λ) of light from these patches was measured by a spectrometer (PhotoResearchTM PR-670). The first two subgraphs (a) and (b) in Figure 6 show the light source and reflectance spectrum of patches and the blue patch of the color checker. As shown in the subgraph (d) in Figure 6, we fit the coefficients of the standard basis to estimate the actual SRF. With the measured SRF, we reconstruct a high-fidelity HSI in Figure 7. The error maps show small errors between the ground truth and reconstructed images. The quantitative errors are articulated in Table 1.

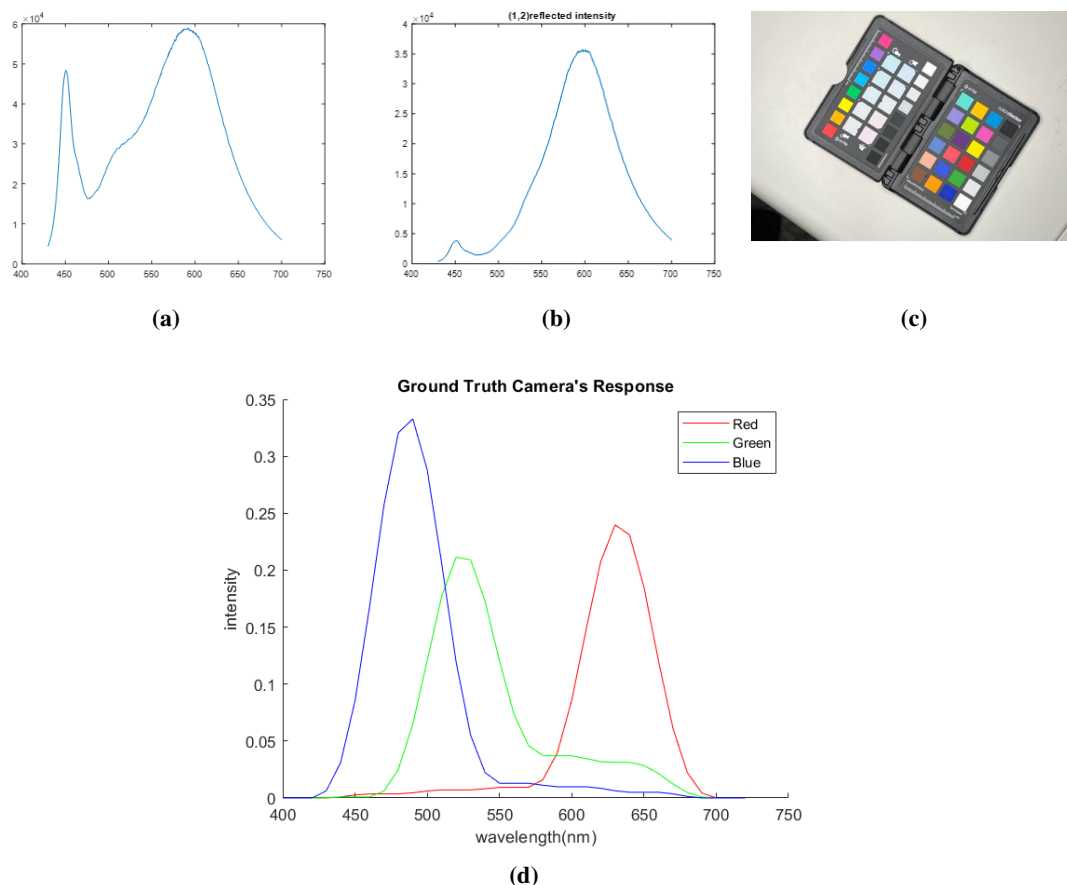


Figure 6. (a) The spectral response of the light source illuminates the experiment. (b) The color checker. (c) The spectral response of the blue patch at the (1,2) position of the color checker. (d) The fitted coefficients of the measured response.

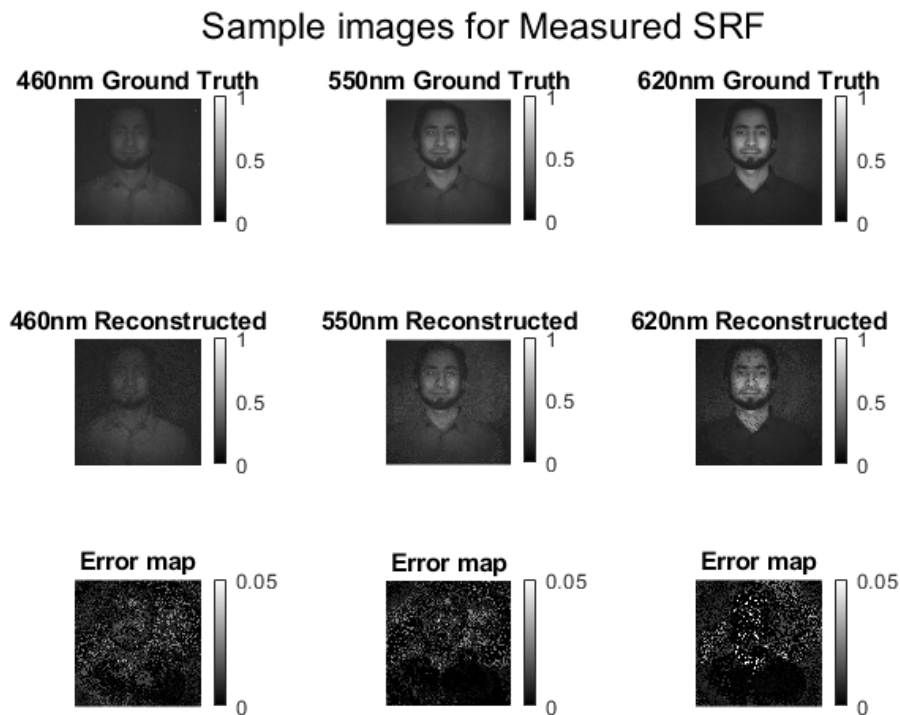


Figure 7. A sample of reconstructed HSI in 460, 550 and 620 nm bands using the measured SRF. The entire HSI contains 33 bands from 400 nm to 720 nm. The errors between the original and the reconstructed images are insignificant based on the visual representation. Few pixels have errors approximating to 0.05. Please refer to Table 1 for the summarized root mean square errors.

Table 1. Average prediction accuracy between the original and reconstructed HSI

method	SRF	rmse	rrmse
ours	CIE	0.029766	0.069987
	Measured	0.037874	0.098941
	Generated	0.038611	0.12115
[1]	CIE	0.052838	0.1111
	Measured	0.04472	0.15981
	Generated	0.08846	0.16791

rmse=root mean squared error; rrmse=relative root mean squared error.

4.3. With unknown SRF and CIF

To achieve the goal of being calibration-free, autonomous generating of SRF should be implemented. Without any help of additional hardware, we iteratively generated an SRF by the

cycleGAN (Figure 8). Through samples in the iteration in Figure 9, the cycleGAN converges promptly. Despite the slight non-smoothness in the spectral response in R-G-B bands, the reconstruction accuracy is still high. The error maps in Figure 10 still exhibit tiny errors.

As shown in Table 1, the generated SRFs by cycleGAN are accurate. The *rmse* (root mean squared error) and *rrmse* (relative root mean squared error) were 0.038 and 0.12, respectively. The convergence is relatively straightforward because the dimension in the latent space is the same as the HSI bands. The estimated SRFs are effective according to the low *rmse*, both by our kernel projection method or the dictionary learning method in [1].

To further visualize the performance of our reconstruction and generation algorithms, we show the original and the synthesized picture from the recovered HSI side-by-side in Figure 11. It is evident that the two pictures are almost identical, which implies that the reconstructed HSI is sufficiently accurate.

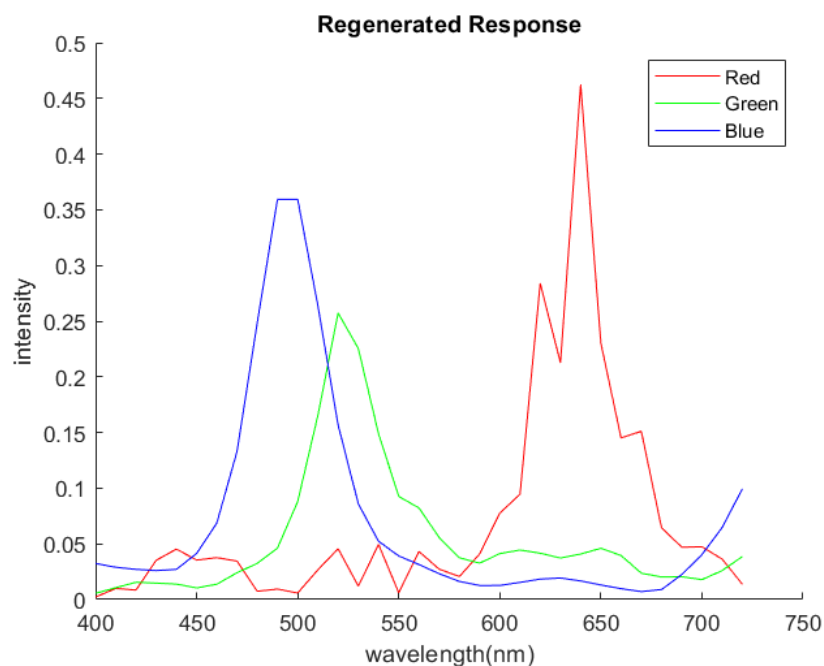
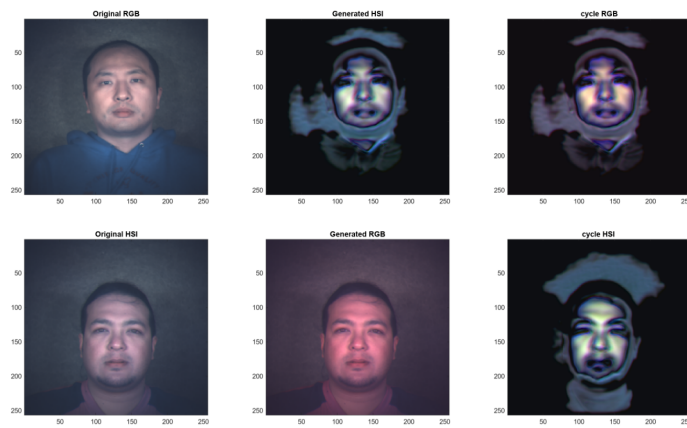
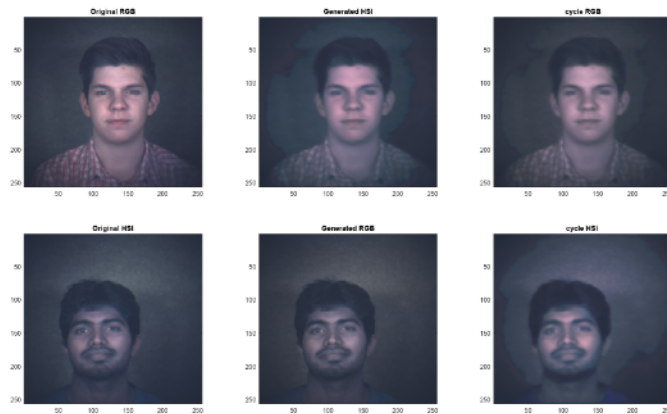


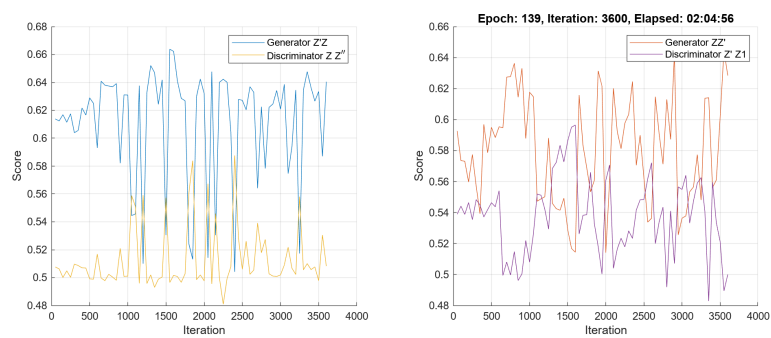
Figure 8. The generated SRF by our algorithm.



(a) iteration=50



(b) iteration=4000



(c) trends

Figure 9. (a) A sample of large error from the early iteration. (b) A sample of small error from the final iteration. (c) Score changes of (3.13) and (3.14) in the process of converging.

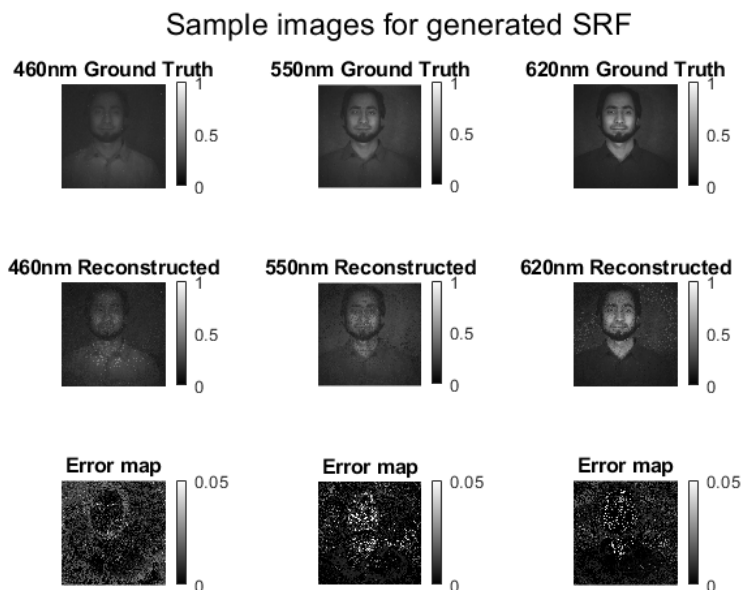


Figure 10. A sample of reconstructed HSI in 460, 550 and 620 nm bands using the generated SRF in Figure 8. The entire HSI contains 33 bands from 400 nm to 720 nm. The errors between the original and the reconstructed images are insignificant based on the visual representation. Not many pixels have errors approximating to 0.05. Please refer to Table 1 for the summarized root mean square errors.

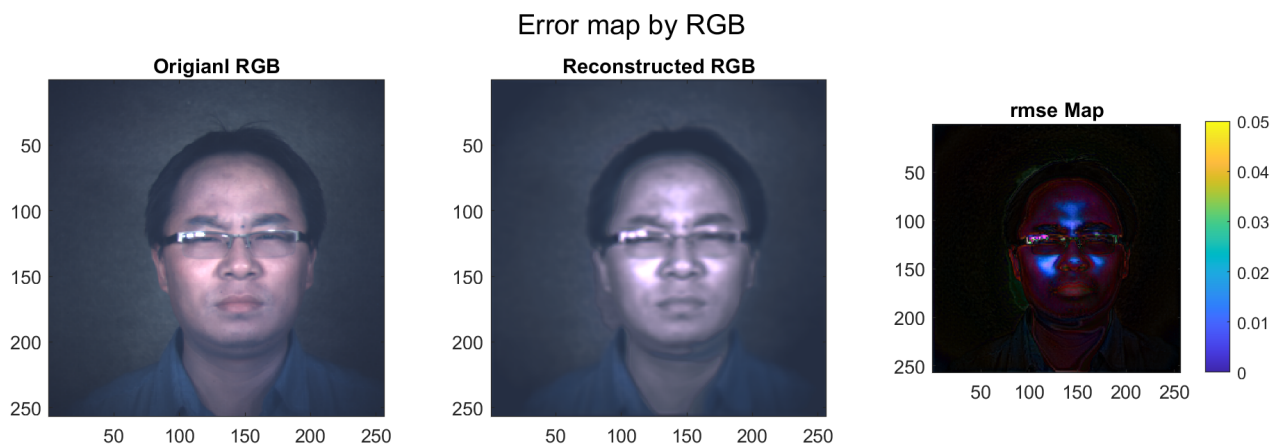
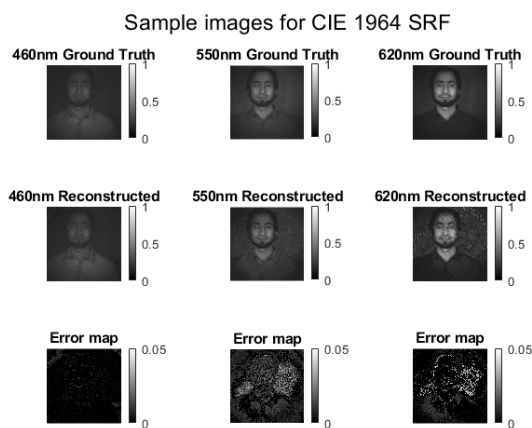
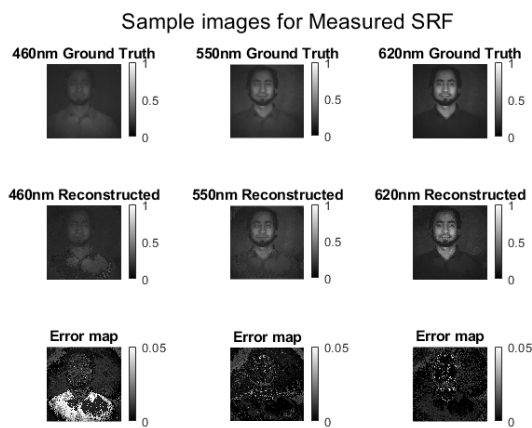


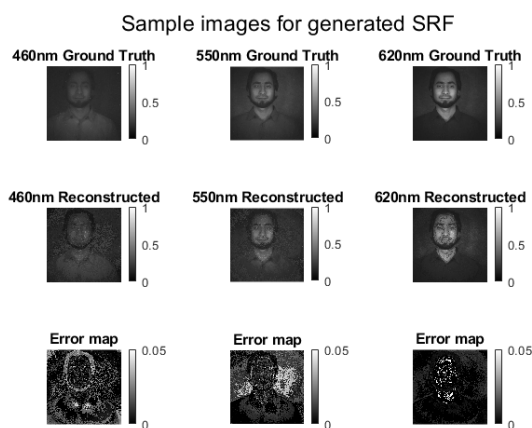
Figure 11. Using the generated SRF in Figure 8, the HSI was reconstructed in Figure 10. For better illustration, the RGB images were compared side-by-side between the original picture (left) and the synthesized picture from the recovered HSI (right). The detailed error maps in HSI were shown in Figure 10.



(a)



(b)



(c)

Figure 12. The error maps by other projection method [1]. (a) The HSI by other methods with CIE SRF. (b) The HSI by other methods with measured SRF. (c) The HSI by other methods with our generated SRF.

4.4. Discussion

The proposed method is superior to existing methods because it does not require laboratory measurement of SRF for a new camera. The experiments demonstrate that our automatic estimated SRFs are almost identical to the laboratory measurement.

Our generated SRFs are accurate and effective. The reconstructed HSIs with generated SRFs have low *rmse*, both by our kernel projection method and the existing method. We compared our result to an existing method [1] (Figure 12 and the second row of Table 1).

Our learning process, however, has errors. The reconstructed HSIs from a 30-dollar webcam will not be as accurate as a 40-thousand-dollar camera. In Figure 10, the comparison of RGB shows a large 3% mse error. The maximal errors happened when the images appeared to be reflected by highlights. The highlighted parts are sensitive to unknown artifacts and cause the model confusion in the reconstruction process.

Fortunately, the target application scenarios mainly need cheap solutions and do not care much about accuracy. Moreover, noises, such as modeling errors and light source disturbance, influence the accuracy of SRF identification and HSI reconstruction.

5. Conclusions

This research contributes to deep reverse analytics by integrating an automatic calibration procedure. This paper offers two contributions targeting HSI reconstruction. The estimated SRFs and CIFs match the results measured by the standard laboratory method, and the estimated HSIs achieve less than 3% errors in *rmse*. Therefore, our method possesses apparent advantages compared to other methods. Experimental results in real examples demonstrated the effectiveness of our method.

Limitations and future research

Our proposed algorithms work under several assumptions. We assume noise has no significant effect on the reconstruction process. In the future, we can use noise representation in the latent space. We will consider the spatial interaction effects resulting from the optical system in the future. In this study, we assume the exposure for each photo-transistor is independent. Therefore, we convert a picture to a 1D array in the model learning process to save computation time. In the future, we will use 2D array modeling.

Use of AI tools declaration

The authors declare they have not used artificial intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported in part by the National Science and Technology Council, Taiwan, under Grant Numbers MOST 110-2622-E-992-026, MOST 111-2221-E-037-007, NSTC 112-2218-E-992-004, and NSTC 112-2221-E-153-004. The authors also thank Mr. Huang, Yuan's General Hospital (ST110006), and NSYSU-KMU Joint Research Project (#NSYSU-KMU-112-P10).

Conflict of interest

The authors have no competing interests to declare.

References

1. B. Arad, O. Ben-Shahar, Sparse recovery of hyperspectral signal from natural RGB images, in *European Conference on Computer Vision*, Springer, 19–34. (2016), https://doi.org/10.1007/978-3-319-46478-7_2
2. I. Choi, D. S. Jeon, G. Nam, D. Gutierrez, M. H. Kim, High-quality hyperspectral reconstruction using a spectral prior, *ACM T. Graphic.*, **36** (2017), 1–13. <http://dx.doi.org/10.1145/3130800.3130810>
3. W. Jakob, J. Hanika, A low-dimensional function space for efficient spectral upsampling, *Comput. Graph. Forum*, **38** (2019), 147–155. <https://doi.org/10.1111/cgf.13626>
4. Y. Jia, Y. Zheng, L. Gu, A. Subpa-Asa, A. Lam, Y. Sato, et al., From RGB to spectrum for natural scenes via manifold-based mapping, in *Proceedings of the IEEE international conference on computer vision*, 4705–4713, (2017). <https://doi.org/10.1109/ICCV.2017.504>
5. H. Kwon, Y. W. Tai, RGB-guided hyperspectral image upsampling, in *Proceedings of the IEEE International Conference on Computer Vision*, 307–315, (2015). <https://doi.org/10.1109/ICCV.2015.43>
6. S. W. Oh, M. S. Brown, M. Pollefeys, S. J. Kim, Do it yourself hyperspectral imaging with everyday digital cameras, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2461–2469, (2016). <https://doi.org/10.1109/CVPR.2016.270>
7. Q. Li, X. He, Y. Wang, H. Liu, D. Xu, F. Guo, Review of spectral imaging technology in biomedical engineering: achievements and challenges, *J. Biomed. Opt.*, **18** (2013), 100901–100901. <https://doi.org/10.1117/1.JBO.18.10.100901>
8. M. Aharon, M. Elad, A. Bruckstein, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, *IEEE T. Signal Proces.*, **54** (2006), 4311–4322. <https://doi.org/10.1109/TSP.2006.881199>
9. Z. Xing, M. Zhou, A. Castrodad, G. Sapiro, L. Carin, Dictionary learning for noisy and incomplete hyperspectral images, *SIAM J. Imag. Sci.*, **5** (2012), 33–56. <https://doi.org/10.1137/110837486>
10. O. Burggraaff, N. Schmidt, J. Zamorano, K. Pauly, S. Pascual, C. Tapia, et al., Standardized spectral and radiometric calibration of consumer cameras, *Optics Express*, **27** (2019), 19075–19101. <https://doi.org/10.1364/OE.27.019075>
11. J. Jiang, D. Liu, J. Gu, S. Süsstrunk, What is the space of spectral sensitivity functions for digital color cameras?, in *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, IEEE, 168–179, (2013). <https://doi.org/10.1109/WACV.2013.6475015>
12. S. Han, Y. Matsushita, I. Sato, T. Okabe, Y. Sato, Camera spectral sensitivity estimation from a single image under unknown illumination by using fluorescence, in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 805–812, (2012). <https://doi.org/10.1109/CVPR.2012.6247752>

13. G. Wu, L. Qian, G. Hu, X. Li, Spectral reflectance recovery from tristimulus values under multi-illuminants, *J. Spectrosc.*, 2019. <https://doi.org/10.1155/2019/3538265>
14. L. Yan, X. Wang, M. Zhao, M. Kaloorazi, J. Chen, S. Rahardja, Reconstruction of hyperspectral data from RGB images with prior category information, *IEEE T. Comput. Imag.*, **6** (2020), 1070–1081. <https://doi.org/10.1109/TCI.2020.3000320>
15. M. D. Grossberg, S. K. Nayar, Determining the camera response from images: What is knowable?, *IEEE T. Pattern Anal.*, **25** (2003), 1455–1467. <https://doi.org/10.1109/TPAMI.2003.1240119>
16. Y. Choi, M. Choi, M. Kim, J. W. Ha, S. Kim, J. Choo, StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 8789–8797, (2018). <https://doi.org/10.48550/arXiv.1711.09020>
17. J. Schneider, Domain transformer: Predicting samples of unseen, future domains, in *2022 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 1–8, (2022). <https://doi.org/10.1109/IJCNN55064.2022.9892250> .
18. L. Yan, J. Feng, T. Hang, Y. Zhu, Flow interval prediction based on deep residual network and lower and upper boundary estimation method, *Appl. Soft Comput.*, **104** (2021), 107228. <https://doi.org/10.1016/j.asoc.2021.107228>
19. P. Isola, J. Y. Zhu, T. Zhou, A. A. Efros, Image-to-image translation with conditional adversarial networks, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125–1134, (2017). <https://doi.org/10.1109/CVPR.2017.632>
20. Y. J. Chen, L. C. Lin, S. T. Yang, K. S. Hwang, C. T. Liao, W. H. Ho, High-reliability non-contact photoplethysmography imaging for newborn care by a generative artificial intelligence, *IEEE Access*, **11** (2022), 90801–90810. <https://doi.org/10.1109/ACCESS.2023.3307637>
21. K. Yin, Z. Chen, H. Huang, D. Cohen-Or, H. Zhang, Logan: Unpaired shape transform in latent overcomplete space, *ACM T. Graphic.*, **38** (2019), 1–13. <https://doi.org/10.1145/3355089.3356494>
22. H. You, Y. Cheng, T. Cheng, C. Li, P. Zhou, Bayesian cycle-consistent generative adversarial networks via marginalizing latent sampling, *IEEE T. Neur. Net. Learn. Syst.*, **32** (2020), 4389–4403. <https://doi.org/10.1109/TNNLS.2020.3017669>
23. F. Campillo, V. Rossi, Convolution particle filter for parameter estimation in general state-space models, *IEEE T. Aero. Elec. Syst.*, **45**. <https://doi.org/10.1109/TAES.2009.5259183>
24. K. Vo, E. K. Naeini, A. Naderi, D. Jilani, A. M. Rahmani, N. Dutt, et al., P2e-wgan: Ecg waveform synthesis from PPG with conditional wasserstein generative adversarial networks, in *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, 1030–1036, (2021). <https://doi.org/10.1145/3412841.3441979>
25. G. Tsialiamanis, M. Champneys, N. Dervilis, D. J. Wagg, K. Worden, On the application of generative adversarial networks for nonlinear modal analysis, *Mech. Syst. Signal Pr.*, **166** (2022), 108473. <https://doi.org/10.1016/j.ymsp.2021.108473>
26. S. A. Burns, Chromatic adaptation transform by spectral reconstruction, *Color Res. Appl.*, **44** (2019), 682–693. <https://doi.org/10.1002/col.22384>

27. M. Störning, H. J. Andersen, E. Granum, Physics-based modelling of human skin colour under mixed illuminants, *Robot. Auton. Syst.*, **35** (2001), 131–142. [https://doi.org/10.1016/S0921-8890\(01\)00122-1](https://doi.org/10.1016/S0921-8890(01)00122-1)
28. X. Zhang, Q. Wang, J. Li, X. Zhou, Y. Yang, H. Xu, Estimating spectral reflectance from camera responses based on cie xyz tristimulus values under multi-illuminants, *Color Res. Appl.*, **42** (2017), 68–77. <https://doi.org/10.1002/col.22037>
29. J. F. Galantowicz, D. Entekhabi, E. G. Njoku, Tests of sequential data assimilation for retrieving profile soil moisture and temperature from observed L-band radiobrightness, *IEEE T. Geosci. Remote*, **37** (1999), 1860–1870. <https://doi.org/10.1109/36.774699>
30. J. S. Liu, F. Liang, W. H. Wong, The multiple-try method and local optimization in metropolis sampling, *J. Am. Stat. Assoc.*, **95** (2000), 121–134. <https://doi.org/10.1080/01621459.2000.10473908>
31. L. Martino, J. Read, D. Luengo, Independent doubly adaptive rejection metropolis sampling within gibbs sampling., *IEEE T. Signal Proces.*, **63** (2015), 3123–3138. <https://doi.org/10.1109/TSP.2015.2420537>
32. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., Generative adversarial nets, in *Advances in neural information processing systems*, 2672–2680, (2014). <https://doi.org/10.48550/arXiv.1406.2661>
33. P. Vincent, Y. Bengio, Kernel matching pursuit, *Mach. Learn.*, **48** (2002), 165–187. <https://doi.org/10.1023/A:1013955821559>
34. G. Aneiros-Pérez, R. Cao, J. M. Vilar-Fernández, Functional methods for time series prediction: A nonparametric approach, *J. Forecasting*, **30** (2011), 377–392. <https://doi.org/10.1002/for.1169>
35. E. Masry, Nonparametric regression estimation for dependent functional data: asymptotic normality, *Stoch. Proc. Appl.*, **115** (2005), 155–177. <https://doi.org/10.1016/j.spa.2004.07.006>
36. H. Chun, S. Keleş, Sparse partial least squares regression for simultaneous dimension reduction and variable selection, *J. Roy. Stat. Soc. B*, **72** (2010), 3–25. <https://doi.org/10.1111/j.1467-9868.2009.00723.x>
37. G. Zhu, Z. Su, Envelope-based sparse partial least squares, *Ann. Stat.*, **48** (2020), 161–182. <https://doi.org/10.1214/18-AOS1796>
38. H. Zou, T. Hastie, Regularization and variable selection via the elastic net, *J. Roy. Stat. Soc. B*, **67** (2005), 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>
39. E. J. Candes, T. Tao, Decoding by linear programming, *IEEE T. Inf. Theory*, **51** (2005), 4203–4215. <https://doi.org/10.1109/TIT.2005.858979>
40. E. J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *IEEE T. Inf. Theory*, **52** (2006), 489–509. <https://doi.org/10.1109/TIT.2005.862083>
41. D. W. Marquardt, R. D. Snee, Ridge regression in practice, *Am. Stat.*, **29** (1975), 3–20. <https://doi.org/10.1080/00031305.1975.10479105>

42. P. Exterkate, P. J. Groenen, C. Heij, D. van Dijk, Nonlinear forecasting with many predictors using kernel ridge regression, *Int. J. Forecasting*, **32** (2016), 736–753. <https://doi.org/10.1016/j.ijforecast.2015.11.017>
43. C. García, J. García, M. López Martín, R. Salmerón, Collinearity: Revisiting the variance inflation factor in ridge regression, *J. Appl. Stat.*, **42** (2015), 648–661. <https://doi.org/10.1080/02664763.2014.980789>
44. E. J. Candes, The restricted isometry property and its implications for compressed sensing, *CR Math.*, **346** (2008), 589–592. <https://doi.org/10.1016/j.crma.2008.03.014>
45. M. Uzair, Z. Khan, A. Mahmood, F. Shafait, A. Mian, Uwa hyperspectral face database, 2023, <https://dx.doi.org/10.21227/8714-kx37>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)