

An Unsupervised Material Learning Method for Imaging Spectroscopy

Johannes Jordan¹ Elli Angelopoulou¹ Antonio Robles-Kelly^{2,3}

¹Pattern Recognition Lab, University of Erlangen-Nuremberg, 91058 Erlangen, Germany

²School of Engineering, Australian National University, Canberra ACT 0200, Australia

³ National ICT Australia (NICTA), Locked Bag 8001, Canberra ACT 2601, Australia

Abstract—In this paper we propose a method for learning the materials in a scene in an unsupervised manner making use of imaging spectroscopy data. Here, we view the input image spectra as a data point on a manifold which corresponds to a node in a graph whose vertices correspond to a set of parameters that should be inferred using the Expectation Maximisation (EM) algorithm. In this manner, we can pose the problem as a statistical unsupervised learning one where the aim of computation becomes the recovery of the set of parameters that allow for the image spectra to be projected onto a set of graph vertices defined a priori. Moreover, as a result of this treatment, the scene material prototypes can be recovered making use of a clustering algorithm applied to the parameter-set. This setting also allows, in a straightforward manner, for the visualisation of the spectra. We discuss the links between our method and self-organizing maps and illustrate the utility of the method as compared to other alternatives elsewhere in the literature.

I. INTRODUCTION

Imaging spectroscopy relies on associating each pixel in the image with a spectrum representing the intensity at each wavelength. As a result, imaging spectroscopy provides an information-rich representation of the scene which combines spatial and compositional information which can be used for a wide variety of applications spanning from remote sensing to food security and health [20].

For spectral image classification, each pixel spectra can be viewed as an input vector in a high dimensional space. This treatment opens up the possibility of representing a scene in terms of a number of spectral prototypes which correspond to naturally occurring materials such as wood, paint, etc. These materials have to be extracted from the scene and are, in general, unknown a priori. Moreover, these material prototypes should be consistent across pixels sharing similar spectra. We illustrate this in Figure 1 where we show, in the left-hand panel, a pseudocolour image taken from the CAVE dataset¹. The image depicts two faces, one is a printout and the other is an actual human subject. In the right-hand panel we show the corresponding material map where each of the corresponding prototypes corresponds to a different hue. Note that the two faces are well separated in hue, one being magenta and the other one red. In the figure, the backgrounds have also been separated, i.e. the printout background from the cloth in the scene background.



Fig. 1: Left-hand panel: Pseudocolour image depicting a print-out of a face and a human subject; Right-hand panel: Material map showing different hues for the spectral prototypes in the scene.

From the figure, we can also appreciate that automatic scene material extraction is often complicated due to the confounding factors introduced by illumination and the complex nature of real-world settings. Despite being a challenging task, the automatic recovery of these prototypes allows for avoiding cumbersome labelling of the scene spectral signatures, which, at the moment, is mainly effected through expert intervention [17]. The problem is somewhat related to the recognition of spectral signatures, where algorithms from statistical pattern recognition and machine learning have been adopted to perform pixel-level spectral feature extraction and classification [16]. These methods either directly use the complete spectra, or often make use of preprocessing and dimensionality reduction steps at input and attempt to recover statistically optimal solutions. Linear dimensionality reduction methods are based on the linear projection of the input data to a lower dimensional feature space. Typical methods include Principal Component Analysis (PCA) [12], Linear Discriminant Analysis (LDA) [6], and Projection Pursuit [11]. Almost all linear feature extraction methods can be kernelised, resulting in Kernel PCA [21], Kernel LDA [19] and Kernel Projection Pursuit [4]. These methods exploit nonlinear relations between different segments in the spectra by mapping the input data onto a high dimensional space through different kernel functions [21].

Note that the algorithms above aim at solving the classification problem at the pixel level rather than treating the problem as a blind source labelling task which can be viewed as an unsupervised learning problem. Such a problem can be tackled in a number of ways. Here we note that generative models have been proposed as probabilistic generalizations of ad-hoc

¹The data set can be accessed at <http://www.cs.columbia.edu/CAVE/databases/multispectral/>

learning methods mostly for unsupervised learning scenarios [2]. One particular example is the Generative Topographic Mapping [2], which can be seen as a generalization of Kohonen's self-organizing maps (SOMs) [15].

Moreover, neural networks and, more recently, deep belief networks (DBNs) have proved popular as methods for learning latent representations with the goal of tackling difficult problems in AI [8], [1]. Here, we explore the use of a generative model to produce a set of parameter vectors that can be used to describe a graph whose vertices correspond to a mapping of the spectra onto a feature space. We adopt a probabilistic treatment of the problem where these parameters can be inferred from the image data using the EM algorithm [3]. The use of these parameter vectors has two main advantages. Firstly, it allows for the visualisation of the image spectra in a straightforward manner. Secondly, the material prototypes can be easily recovered using clustering methods, such as mean shift [7]. Moreover, by using a graph theoretic setting, the inference process is a structured one which explicitly includes topological information, preserving similarity relations between input spectra.

II. PROBLEM SETTING

As mentioned above, here we cast the problem in hand into a graph-theoretic setting. Let $G_M = (Y, E_M)$ be a graph where the node set Y comprises of the samples in the input, i.e. the image spectra, $y_i \in \mathbb{R}^m$. E_M represents the connectivity between the spectral values. Determining E_M directly is a hard problem. Instead we employ a second graph $G_Q = (X, E_Q)$ as illustrated in Figure 2. While G_M is based on the image data, G_Q is constructed to help us derive the topology of the manifold on which the input spectral data lies. The vertex-sets for both graphs are linked as follows

$$x_j = \sum_{y_k \in Y} \mathbb{I}(x_j \sim y_k) \Gamma(y_k) \quad (1)$$

where Y is the vertex set of G_M , $\Gamma(\cdot)$ is a mapping function such that $\Gamma : \mathbb{R}^m \mapsto \mathbb{R}^q$ and $\mathbb{I}(x_j \sim y_k)$ is an indicator function which is unity if x_j is adjacent to y_k and zero otherwise. In the figure, we show these two graphs, each of which is realised in disjoint manifolds, i.e. $M \in \mathbb{R}^m$ and $Q \in \mathbb{R}^q$. This formalism is important since, as we will see later, we can use the metrics on these manifolds to define the affinities between vertices in each graph.

We now view the graph G_Q as a Gibbs Field. This treatment is important since, by viewing the vertices of the graph as random variables, the joint probability in a Gibbs Field can be written as the product of clique potentials, i.e. any fully connected subset of the the graph. Consider the clique \mathcal{C}_{x_i} centred at the vertex x_i . We can now write the conditional probability of the vertex under consideration as a weighted product of the pairwise potentials as follows

$$P(x_i | \mathcal{C}_{x_i}) = \frac{1}{Z_Q} \prod_{x_j \in \mathcal{C}_{x_i}} f_Q(x_j, x_i) \quad (2)$$

where Z_Q is the partition function and $f_Q(x_j, x_i)$ is the potential function between the vertices x_j and x_i .

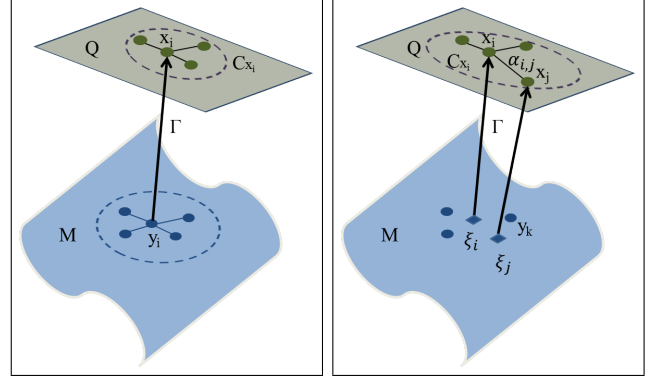


Fig. 2: Left-hand panel: Conceptual illustration of the relationship between the graphs and manifolds using in our problem setting; Right-hand panel: Introduction of the parameter vectors.

Note that the potential function $f_Q(x_j, x_i)$ can be effectively viewed as the edge-weight between the vertex pair x_j - x_i . This will provide, later on in the paper, a means to perform a maximum-likelihood estimation (MLE) on the vertex-set $Y \in G_M$ based upon the vertices $X \in G_Q$. Moreover, we can constraint this process using the metrics on the manifolds M and Q . Recall that, in Equation 1 we expressed the vertex x_j as a sum over the product $\mathbb{I}(x_j \sim y_k) \Gamma(y_k)$. If the actual correspondences $x_j \sim y_k$ are not in hand, we can employ the prior η_i so as to relax Equation 1 and write

$$x_j = \sum_{y_k \in Y} \eta_j \mathcal{K}(y_k, \xi_j) \quad (3)$$

where $\mathcal{K}(\cdot)$ is a kernel function with parameter set ξ_j . In the left-hand panel of Figure 2, we provide an intuitive interpretation to this relaxation process. Note that ξ_j is now, in practice, a kernel variable on the manifold M which has a correspondence relationship to the vertex x_j . This treatment leads to a setting such as that shown in the right-hand panel of Figure 2, where the vertices in G_Q correspond to each of the parameter vectors ξ_j , which, in turn, are supported by the vertex set Y .

By substituting Equation 3 into Equation 2, we get

$$P(x_i | \mathcal{C}_{x_i}, Y) = \frac{1}{Z_Q} \prod_{x_j \in \mathcal{C}_{x_i}} f_Q \left(\sum_{y_k \in Y} \mathcal{K}(y_k, \xi_j), x_i \right) \quad (4)$$

The equation above is telling since it implies that, by using the cliques in G_Q , the vertex sets X and Y , a maximum likelihood estimate of the parameter ξ_j can be obtained based upon our choice of kernel and potential functions.

III. PARAMETER LEARNING

Moreover, note that the potential function $f_Q(\cdot)$ are defined in the probability space corresponding to X , i.e. the graph G_Q , whereas the kernel function $\mathcal{K}(\cdot)$ and the parameter vector ξ_j are supported by the graph G_M . This opens-up the possibility to recover the parameter vector-set making use of maximum likelihood estimation. This is important since,

once these are in hand, the vertices of the graph G_M can be mapped onto the manifold Q based upon the topology of the graph G_M .

To take our analysis further, we note that, since the vertex-set for the graph G_Q is a Gibbs Field, it should follow a Gibbs distribution. Hence, we employ the potential function

$$f_Q(x_j, x_i) = \sum_{y_k \in Y} \alpha_{i,j} \exp \left\{ -\frac{1}{T} d_M(y_k, \xi_j)^2 \right\} \quad (5)$$

where $d_M(\cdot)^2$ is the squared geodesic distance on the Manifold M , $\alpha_{i,j}$ are pairwise mixture weights, and the temperature T controls the sharpness of the distribution. With the equation above, we can write the likelihood as follows

$$P(x_i | \mathcal{C}_{x_i}, Y) = \frac{1}{Z_Q} \prod_{x_j \in \mathcal{C}_{x_i}} \sum_{y_k \in Y} \alpha_{i,j} \exp \left\{ -\frac{1}{T} d_M(y_k, \xi_j)^2 \right\} \quad (6)$$

From inspection, it is clear that we can view the MLE task at hand as that corresponding to a mixture model which can be tackled using the EM algorithm [3]. The idea underpinning the EM algorithm is to recover maximum likelihood solutions to problems involving missing or hidden data. To do this, we view the parameter vectors ξ_j as a set of variables to be estimated and the mixture weights $\alpha_{i,j}$ as the posterior probabilities that the vertex x_i belongs to the j^{th} component of the mixture.

Since the EM algorithm is an iterative one, we commence by indexing the expected log-likelihood to iteration number t . In the M-step, we aim at maximising the expected log-likelihood with respect to the parameter variables. This yields

$$\xi_j^{t+1} = \frac{\sum_{y_k \in Y} \alpha_{i,j}^t \exp \left\{ -\frac{1}{T} d_M(y_k, \xi_j^t)^2 \right\} y_k}{\sum_{y_k \in Y} \alpha_{i,j}^t \exp \left\{ -\frac{1}{T} d_M(y_k, \xi_j^t)^2 \right\}} \quad (7)$$

In the E-step, the posterior probabilities are estimated. For the likelihood in Equation 6, this yields

$$\alpha_{i,j}^{t+1} = \frac{\tau_i \sum_{y_k \in Y} \alpha_{i,j}^t \exp \left\{ -\frac{1}{T} d_M(y_k, \xi_j^t)^2 \right\}}{\sum_{x_l \in X} \tau_l \sum_{y_k \in Y} \alpha_{l,j}^t \exp \left\{ -\frac{1}{T} d_M(y_k, \xi_j^t)^2 \right\}} \quad (8)$$

where τ_i can be viewed as the posterior for each of the cliques in the graph. This is given by

$$\tau_i = \frac{\sum_{x_k \in \mathcal{C}_i} \alpha_{k,i}^t}{\sum_{x_k \in \mathcal{C}_i} \alpha_{i,k}^t} \quad (9)$$

IV. DISCUSSION

Note that, so far, we have arrived to Equations 7-9 making use of an unsupervised learning approach such that the topology of the graph G_Q constrains the inference process through its clique-set, whereas the parameter vectors ξ_i govern the kernel function $\mathcal{K}(\cdot)$.

A. Energy Functions

It is worth noting in passing that the use of the potential function in Equation 5 corresponds to a Gibbs measure. This accounts for a Boltzmann distribution [14] whose energy is the squared geodesic distance $d_M(\cdot)^2$. This can be easily seen by writing

$$f_Q(x_j, x_i) = \alpha_{i,j} \sum_{y_k \in Y} \exp \left\{ -\frac{1}{T} E_{\xi_j}(y_k) \right\} \quad (10)$$

where $E_{\xi_j}(y_k) = d_M(y_k, \xi_j)^2$ is the energy of y_k with respect to the parameter vector ξ_j and $\alpha_{i,j}$ is a function of the vertex-pair x_i - x_j . Note that, for instance, in Equation 5, its straightforward to set $\alpha_{i,j} = \eta_i \eta_j$ so as to write $f_Q(x_j, x_i) = \eta_i \eta_j \mathcal{K}(y_k, \xi_j)$ such that

$$\mathcal{K}(y_k, \xi_j) = \sum_{y_k \in Y} \exp \left\{ -\frac{1}{T} d_M(y_k, \xi_j)^2 \right\} \quad (11)$$

This treatment is consistent with that commonly given to priors in Bayesian inference on multivariate mixtures [18] and opens up the possibility of using other kernels such as box functions or robust estimators [9].

B. Relation to Self-organizing Maps

Note that other energy functions may be used without any loss of generality. Moreover, in practice, there is often the case whereby the alpha weights do not need to be estimated but rather can be computed from the vertex-set X . In such cases, the inference process presented earlier is reminiscent of a self-organizing map (SOM) [15]. Consider a sampling process in M and set

$$\alpha_{i,j} = \mathbb{I}(x_j \sim \rho) h \left(\frac{1}{T} d_Q(x_i, x_j)^2 \right) \quad (12)$$

where $h(\cdot)$ is a real-valued positive function, $d_Q(x_i, x_j)$ is the geodesic distance on Q between the vertices x_i and x_j , T is, as before, the temperature of the system and $\mathbb{I}(x_j \sim \rho)$ is an indicator function which is unity for the vertex x_j corresponding to the parameter variable ξ_j whose distance to the sample y_k is minimum, i.e. the best matching unit ρ , and zero otherwise.

Moreover, consider the case where the energy function $E_{\xi_j}(y_k)$ is constant. With these ingredients, the sampling process is then such that the update strategy in Equation 8 will draw the parameter vector ξ_j^{t+1} towards the vector ξ_ρ^t corresponding to the vertex ρ in G_Q which is closest to the input sample y_k , i.e. the neighbours of the best matching unit ρ . The resulting scheme is also consistent to the notion that, as the system ‘‘cools down’’, the influence of the geodesic distance upon the update decreases.

To illustrate this behaviour, in Figure 3 we show the 2-dimensional responses, visualised using a red-green colour space, when only the best matching unit is used for the inference process and the distance on Q in Equation 12 is given by $d_Q(x_i, x_j)^2 = (x_i - x_j)^2$ (SOM-Euclidean) or $d_Q(x_i, x_j)^2 \equiv a$, where a is a constant (SOM-Constant). This, from the structural point of view, can also be viewed

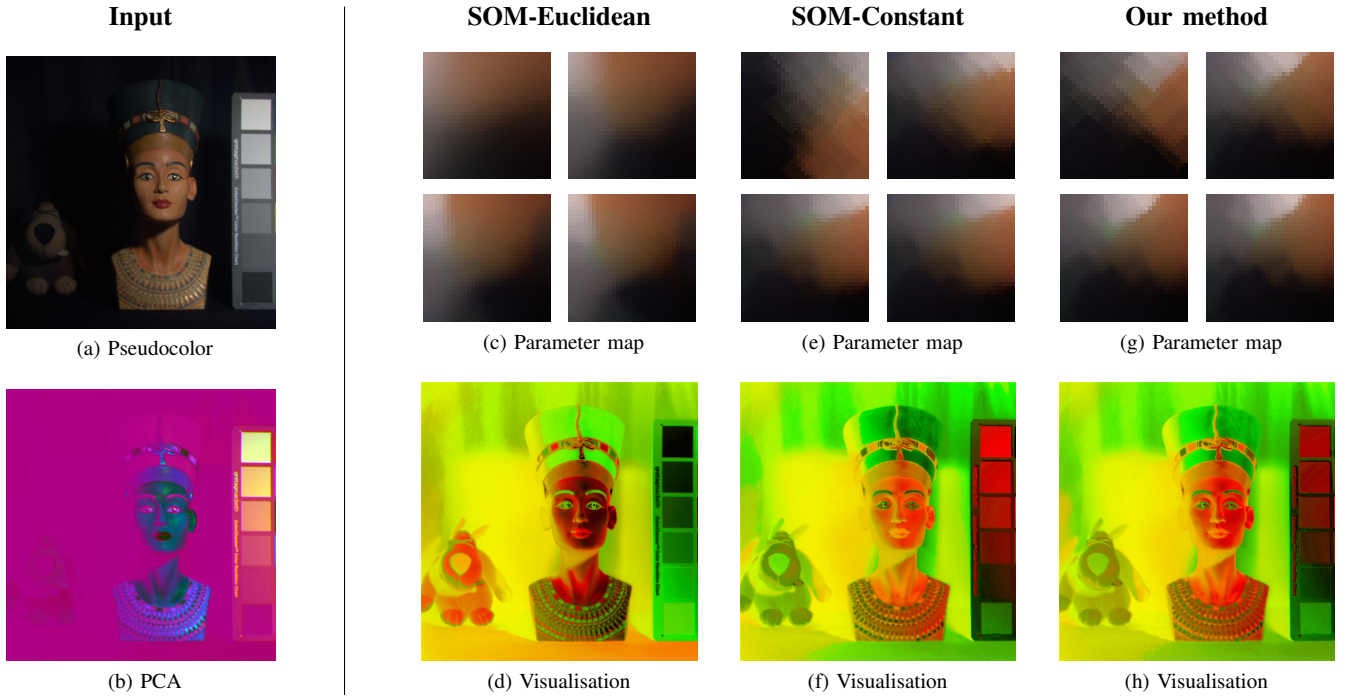


Fig. 3: Left-hand panels: **(a)** Input image in pseudocolour. **(b)** Input visualised using the first three principal components of the spectra. Right-hand panels: **(c), (e), (g)** Organisation of the ξ_j parameters after $1/4, 1/2, 3/4$ of iterations and the final result for three alternatives. **(d), (f), (h)** Visualisation of the image using the ξ_j variables as pixel values.

as the case where the edge-weights over each of the cliques in the graph G_Q are given by an exponential function or a constant, respectively. In the figure, we have used uniformly distributed vertices, i.e. units, and, for the clique size $|\mathcal{C}_{x_i}|$, we have linearly decreased the number of nearest neighbours on each dimension of the red-green lattice. We have done this for the sake of simplicity and so as to provide a direct analogy to a 2-dimensional SOM.

In Fig. 3, we also show the input in pseudo-colour as computed using the colour matching functions of Stiles and Burch [22] (Fig. 3a), the visualisation delivered by the three principal components (PCA, Fig. 3b) of the spectra when these are used as pixel colour values and the mapping delivered by our method (Fig. 3h) when $\alpha_{i,j} = \frac{1}{T} \exp \left\{ -\frac{1}{T} (x_i - x_j)^2 \right\}$, i.e. $d_Q(x_i, x_j)^2 = (x_i - x_j)^2$ in Equation 12.

C. Visualisation

For all the methods in Fig. 3 we have used a linear cooling schedule, i.e. the temperature T decreases at regular intervals throughout the update process. In the figure we also show the states of the graph on Q as it converges over 50 000 iterations (Fig. 3c,e,g), corresponding to 20% of the image data being sampled. In each of the panels, the top-left map corresponds to iteration $t = 12\,500$, top-right to $t = 25\,000$, bottom-left to $t = 37\,500$ and bottom-right to the final, i.e. 50 000th, update. In these panels, the graph nodes are distributed in a lattice and, for visualisation purposes, we have encoded the colour

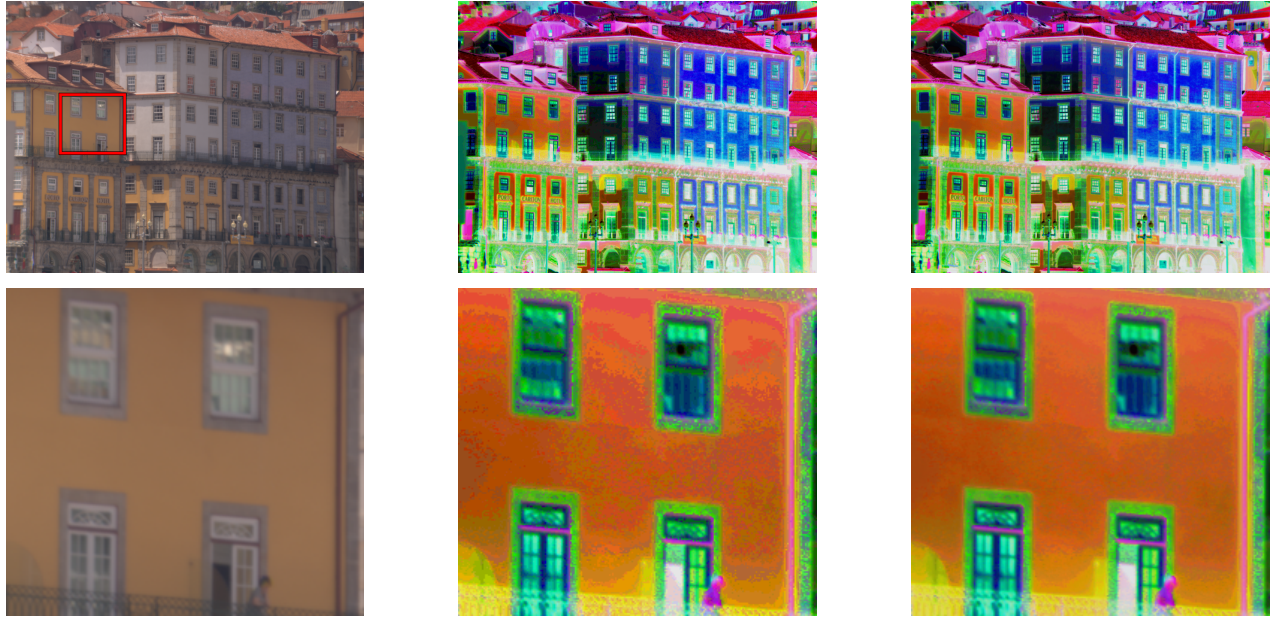
using the parameter vector ξ_j as follows

$$u = \sum_{\substack{y_k \in Y \\ x_j \in \mathcal{C}_\rho}} \mathbb{I}(x_j \sim y_k) \exp \left\{ -\frac{1}{T} d_Q(x_j, \rho)^2 \right\} \exp \left\{ -\frac{1}{T} (y_k - \xi_\rho)^2 \right\} \quad (13)$$

where we have written ξ_ρ and \mathcal{C}_ρ to imply that the parameter variable and clique under consideration correspond to the best matching unit ρ . Note that u is essentially the potential function $f_Q(\cdot)$ presented earlier when Equations 11 and 12 are combined and the corresponding distance and indicator functions are applied accordingly.

Note that, for the methods in Figure 3, i.e. SOM-Euclidean, SOM-Constant and ours, the final parameter vectors ξ_j are organised in a similar manner. This is somewhat expected due to the similarities between the distance functions used in the three methods under consideration. This hints to the use of a SOM-like training process to recover the variables ξ_j and the use of the clique information for the evaluation of the potential functions for testing once the parameter vectors are in hand. This, in effect, delivers a means for a computationally efficient training stage while delivering the advantages of using clique information for the testing of the spectra.

This is more evident by visualising a natural scene from the imagery described in [5] using the ξ_j vectors as pixel values. In Figure 4, we have recovered the vectors ξ_j using the SOM-Constant alternative presented above. In the top row of the figure, from left-to-right, we show the input image and the visualisation induced by the use of the clique contributions by



(a) Input in pseudocolour

(b) Visualisation using only the best matching unit

(c) Visualisation recovered using clique information

Fig. 4: Left-hand column: Pseudocolour and detail, also in pseudocolour, of a natural scene from the imagery described in [5]; Right-hand columns: Visualisation of the input image and its detail using the ξ_j variables as pixel values for both, the case when only the best matching unit and clique information are used.

our method as compared with that of the best matching unit alone. In the bottom row, we show the detail, corresponding to the red square on the input image, for the visualisation in the top row. From the figure, it becomes evident that the use of the contribution of the vertices in the clique \mathcal{C}_ρ in Equation 13 in our method better preserves detail while still delivering a sharp differentiation between materials in the scene.

V. EXPERIMENTS

We now test our method’s capabilities on unsupervised material learning making use of publicly available data sets. Here, we use the images acquired by Foster et al. [5], which contain natural scenes. The data set is well suited for benchmarking as it contains natural scenes with a high spatial resolution. It was acquired using a multispectral sensor equipped with a tunable liquid-crystal filter and contain 33 bands in the range of 400 nm to 720 nm with an effective spatial resolution of approximately 1020×1340 pixels.

We have also used the CAVE dataset and the D.C. Mall remote sensing image². This provides a good mix of remote sensing, lab-acquired and real-world imagery for our evaluations with images of different spectral and spatial resolutions. This is as the CAVE imagery comprises 31 bands in the visible range at 10 nm intervals whereas the D.C. Mall image depicts 191 bands in the spectral range of $400\text{ nm} - 2475\text{ nm}$. In all our experiments, and as the sole preprocessing step, we normalise the spectral vectors (image pixels) to unit L_2 norm in accordance with [10]. Also, here, we compare our

method with that of Huynh and Robles-Kelly [10], hereby denoted as HRK. This method also employs deterministic annealing to recover material association probabilities and is generative in nature.

A. Implementation details

In our implementation, for the sake of efficiency, we train using a SOM-like scheme. Our graph G_Q is a three-dimensional lattice with an initial clique size of 10^3 , i.e. 10 neighbours per dimension. We randomly sample 100 000 image spectra and use a linear cooling schedule. Likewise, the clique size decreases linearly in each dimension as the iteration number increases. This is in accordance to regular SOM training, with a linear decay in both learning rate and neighborhood radius [15].

Based on the learned parameter vectors, we can then infer material prototypes as well as a global segmentation of the image according to material associations. To do this, we first perform clustering on the vectors ξ_j using a mode-seeking algorithm. Each image spectrum can then be associated to a cluster given the mapping described in Eq. 1. Furthermore, the modes obtained during clustering can be seen as the material prototypes of the scene. Here, we employ an efficient variant of mean shift, i.e. Fast Adaptive Mean Shift (FAMS) [7]. Our choice hinges in the fact that FAMS is well suited for a high dimensional feature space by using adaptive bandwidths and a fast nearest-neighbor approximation during its kernel density estimation. As our feature space only contains of $10^3 = 1000$ vectors, we can abstain from the nearest neighbour approximation. The bandwidth for each data point ξ_j is chosen so that $k = 0.6\sqrt{|X|}$ nearest neighbours

²The image can be accessed at <https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html>

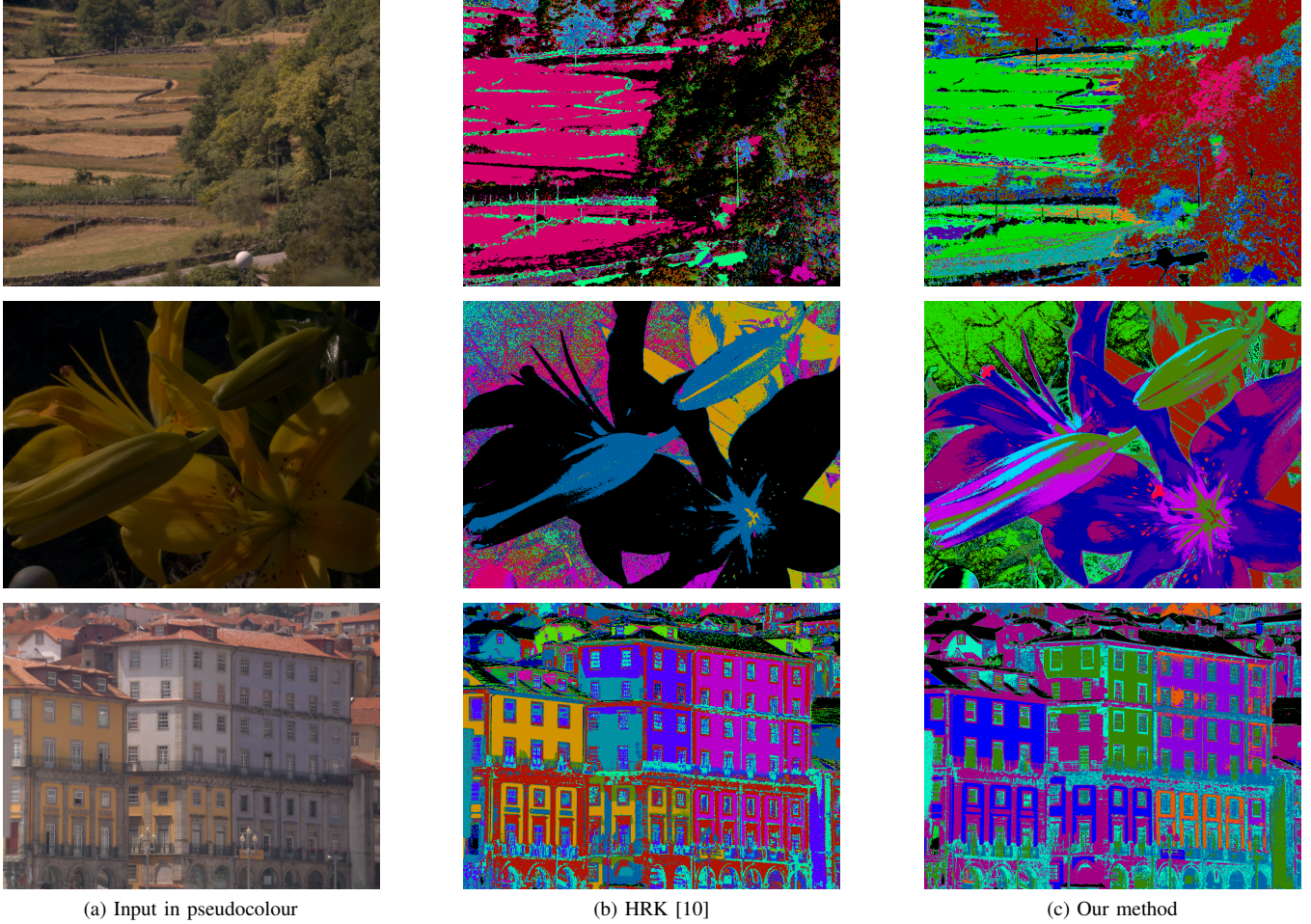


Fig. 5: Material segmentation results on the Cbrufefields, Cyflower, and Ribeira scenes in [5].

lie within its bounds, as suggested by [13]. Additionally, we weight the bandwidth of each ξ_j by the number of image spectra y_k that are associated with it. The rationale behind this is that, in this manner, an accurate approximation of the original spectral distribution can be easily obtained.

For the alternative, we have used the code provided by the authors and, due to memory constraints, performed a 1:2 subsampling of the images (a 4-fold reduction of image size without smoothing). Both methods were run with the same parameter settings on all images.

B. Material associations and Prototypes

We commence by illustrating the behaviour of our method for purposes of recovering material associations. Figure 6 depicts a segmentation of three images from Foster et al.’s dataset [5] based upon the abundance of each material scene at each pixel location, i.e. the material associations for each pixel. In the figure, the segment colors have been randomly assigned. Observe that both methods are capable of discerning most relevant materials in the scene. Nonetheless, in the middle row, i.e. Cyflower, HRK suffers from under-segmentation over several image regions. Further, several clusters are generated from the noisy background of the image. In the third image,

Ribeira (bottom row), HRK achieves a good material separation of the scene, but still is affected by noise on the roof of the buildings, which cause cluster fragmentation.

Figure 7 depicts the material prototypes found in the images as shown in Fig. 6. These prototypes are derived from the clustering results as follows. In the case of HRK, the mean of each cluster is considered as described in [10]. For our method, we use a mode-seeking algorithm [7] so as to take advantage of the parameter vectors directly. Fig. 6a shows the visualization of the normalised image spectra using the Parallel Coordinates method of the Gerbil hyperspectral visualization framework³. In the plot, all the spectra contained in the image are used to obtain a spectral power histogram. Fig. 6b and Fig. 6c show the extracted prototypes using our method and the alternative.

We can observe that both methods succeed in capturing the prominent materials in the scene as the extracted prototypes give a good hint at the respective image’s spectral power distribution. It is worth noting that HRK requires a user-determined number of clusters (20 in this experiment) while our mode-seeking approach is devoid of any parameters. In the

³Downloadable at <http://www.gerbilvis.org/>

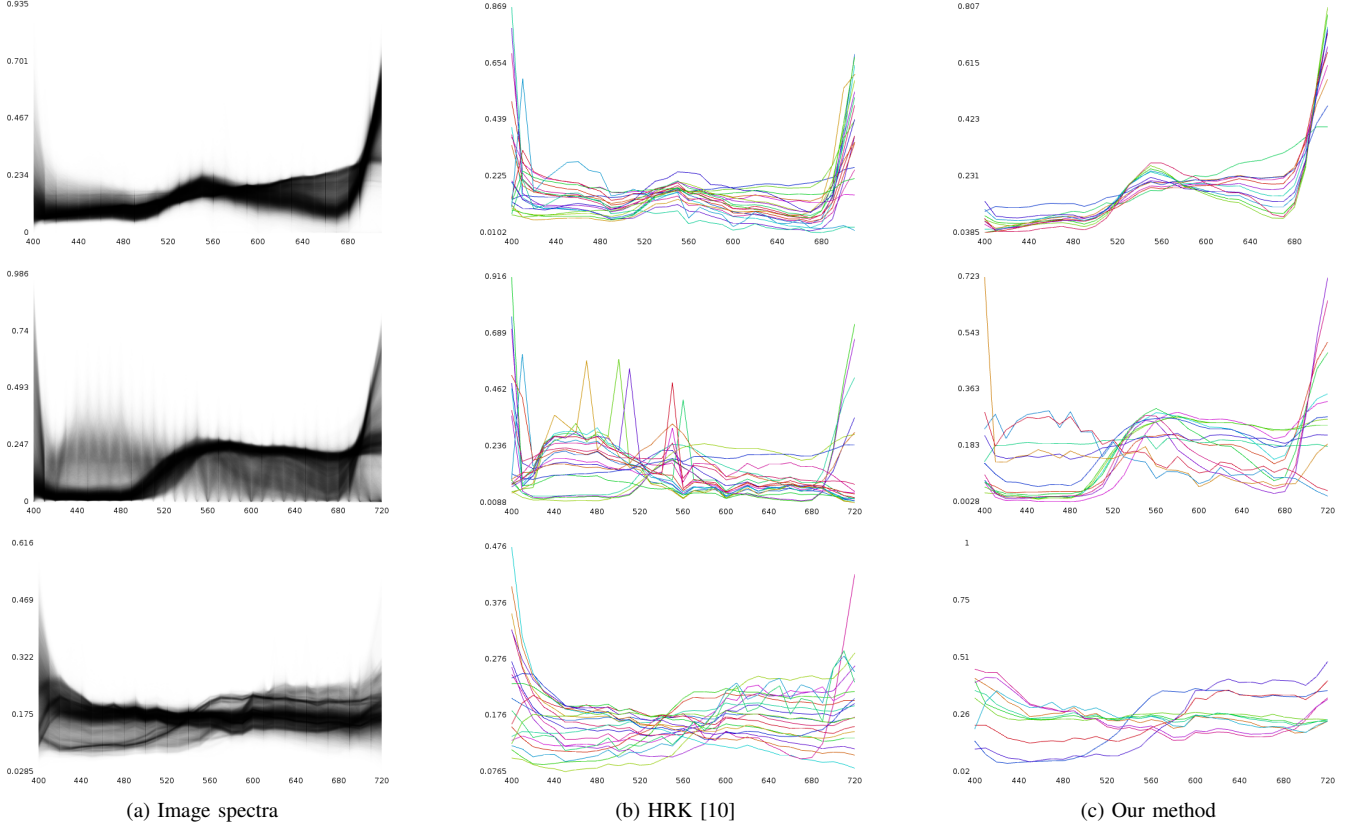


Fig. 6: Comparison of extracted prototypes for the Cbrufefields, Cyflower, and Ribeira scenes. **(a)** Spectral power histogram; **(b)**, **(c)** Material prototypes yielded by our method and the alternative.

case of HRK, some of the clusters found are more descriptive than others. A good example is the *Cyflower* scene, where several prototypes delivered by HRK correspond to noise in the data. In this particular case, some materials prototypes are affected by noise. This produces “spiky” spectra that do not follow the intuition of smooth radiance in the spectral domain across the image.

C. Distance preservation

We now turn our attention to a qualitative analysis of the mapping $\Gamma : \mathbb{R}^m \mapsto \mathbb{R}^q$. Recall that, as described in Eq. 1, our method provides a mapping from a manifold with dimension m to one of dimensionality q . In the case of our experiments, m accounts for the number of wavelength indexed bands in the imagery and $q = 3$. By assuming local linearity in the mapping between the manifolds M and Q (this is reasonable since we have used the equivalent of Euclidean distance functions $d_M(\cdot)$ and $d_Q(\cdot)$), we can assess the relationship between distances on both manifolds employing the notion that those spectral radiance vectors y_j that are close in M should be linked to vectors x_j on our mapping that are, likewise, close to one another in Q .

To evaluate this property, for each pixel y_j in an image, we randomly select 40 other pixels and, for each of those selected y_i , we compute $\|y_i - y_j\|_2$ as well as $\|u - v\|_2$, where u and v correspond to the visualisation vectors for y_i

and y_j , respectively, computed using Eq. 13. In this manner, the resulting distance pairs can be computed so as to assess the linearity of the distance relationships induced by the mapping Γ . In Figure 7, we show the scatter plot histograms for these pairwise relationships. In the figure, we show plots for three images taken from each of the datasets under consideration. We have done this so as to explore the effects of spectral and spatial resolution on the linearity of Γ .

From the figure, we can appreciate that small distances exhibit a high correlation for the three images. Larger distances diverge slightly from a strictly linear relationship. This is somewhat expected since underlying manifolds may not be well approximated by a Euclidean metric as the pairwise distances increase.

D. Running times

Note that a straightforward approach to obtain a similar clustering as ours would be to run the mean shift algorithm directly on the normalized image spectra. We would like to mention in passing that the mean shift has a computational complexity of $O(n^2)$ in the number of pixels. Our profiling throughout the experiments revealed that over 95% of the computation time for the mean shift is spent on the vector distance calculation. This makes the high dimensionality of multispectral data an additional burden to the method. This contrasts with our method, where we effectively find a sparse

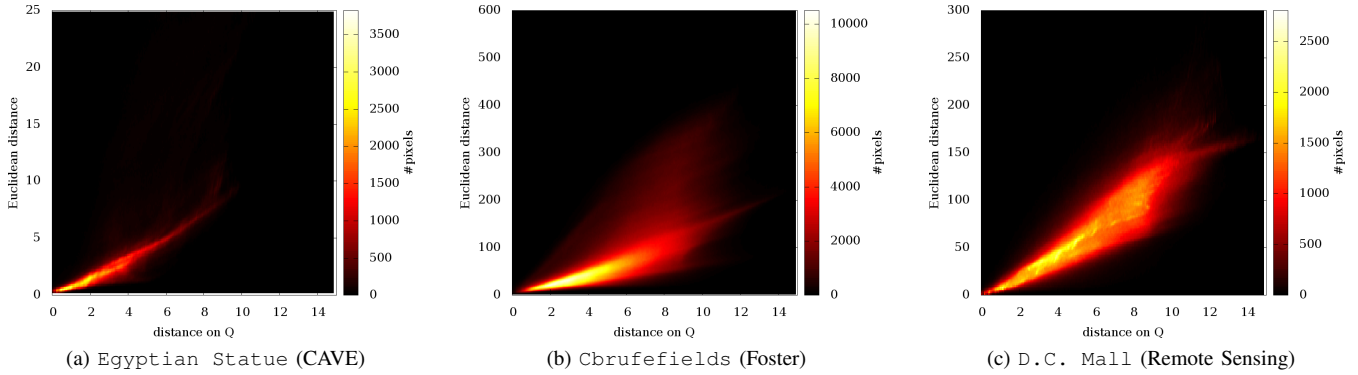


Fig. 7: Scatter plot histograms for pairwise distances induced by the mapping Γ . Colour bars indicate the frequency-hue scale.

representation of the spectral distribution first and then run mean shift on a reduced data set.

To compare performance with our method, we run the FAMS algorithm using an Intel Core i7-2600 CPU and 16 GB RAM on the Foster images. In this case, we employ the approximate nearest neighbour speedup scheme [7]. With these settings, the clustering computation on the raw image radiance takes 10 to 11 hours per image. For our method, the running time consistently lies under 120 seconds per image being in average about 1 min. Note that the training time is low due to the updates being constrained by the clique size and the temperature of the system. Moreover, the timing above for our method comprises the full training, clustering and back-propagation steps of the algorithm. This makes our method also considerably faster than that of Huynh and Robles-Kelly [10], which took, in average 6 mins per image and is typically performed on subsampled versions of the input imagery.

VI. CONCLUSIONS

In this paper we propose an unsupervised method for learning the materials in a scene from imaging spectroscopy data. We have presented a structural unsupervised statistical learning method where the inference process is effected using the EM algorithm. Furthermore, we have discussed the links between our method and the SOM and the relationship between our method and energy functions used elsewhere in the literature. We have illustrated the ability of our method to recover material prototypes, perform scene material segmentation, visualization and dimensionality reduction on several datasets comprising imagery acquired in lab environments, real-world settings and remote sensing platforms. Moreover, our method provides a margin of improvement as compared to a state-of-the-art method for material association recovery, being more computationally efficient in terms of computational power and memory use.

VII. ACKNOWLEDGEMENTS

The authors would like to thank Cong Phuoc Huynh for the experimental results of [10] and Stefan Ploner for the code to generate the plots in Fig. 7. This work was partially funded by the German Research Foundation (GRK 1773).

REFERENCES

- [1] Y. Bengio and Y. LeCun. Scaling learning algorithms towards AI. In *Large-Scale Kernel Machines*. MIT Press, 2007.
- [2] C. M. Bishop, M. Svensén, and C. K. I. Williams. GTM: The generative topographic mapping. *Neural Computation*, 10(1):215–234, 1998.
- [3] A. Dempster, N. Laird, and D. Rubin. Maximum-likelihood from incomplete data via the EM algorithm. *J. Royal Statistical Soc. Ser. B methodological*, 39:1–38, 1977.
- [4] M. Dunder and D. Landgrebe. Toward an optimal supervised classifier for the analysis of hyperspectral data. *IEEE Transaction on Geoscience and Remote Sensing*, 42(1):271–277, 2004.
- [5] D. H. Foster, K. Amano, S. M. C. Nascimento, and M. J. Foster. Frequency of metamerism in natural scenes. *Journal of the Optical Society of America A*, 23(10):2359–2372, 2006.
- [6] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, second edition, 1990.
- [7] B. Georgescu, I. Shimshoni, and P. Meer. Mean shift based clustering in high dimensions: A texture classification example. In *Computer Vision, 2003 Ninth IEEE International Conference on*, volume 1, pages 456–463, oct. 2003.
- [8] G. Hinton and R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504 – 507, 2006.
- [9] P. Huber. *Robust Statistics*. Wiley, 1981.
- [10] C. P. Huynh and A. Robles-Kelly. A probabilistic approach to spectral unmixing. In *Structural, Syntactic, and Statistical Pattern Recognition*, pages 344–353. Springer, 2010.
- [11] L. Jimenez and D. Landgrebe. Hyperspectral data analysis and feature reduction via projection pursuit. *IEEE Transaction on Geoscience and Remote Sensing*, 37(6):2653–2667, 1999.
- [12] I. Jolliffe. *Principal Component Analysis*. Springer, 2002.
- [13] J. Jordan and E. Angelopoulou. Mean-shift Clustering for Interactive Multispectral Image Analysis. In IEEE, editor, *20th IEEE International Conference on Image Processing (ICIP)*, pages 3790–3794, 2013.
- [14] R. Kindermann and J. L. Snell. *Markov Random Fields and Their Applications*. American Mathematical Society, 1980.
- [15] T. Kohonen. *Self-organizing maps*, volume 30 of *Springer series in information sciences*. Springer, 3rd edition, 2001.
- [16] D. Landgrebe. Hyperspectral image data analysis. *IEEE Signal Process. Mag.*, 19:17–28, 2002.
- [17] M. Lennon, G. Mercier, M. C. Mouchot, and L. Hubert-Moy. Spectral unmixing of hyperspectral images with the independent component analysis and wavelet packets. In *Proc. of the International Geoscience and Remote Sensing Symposium*, 2001.
- [18] G. McLachlan and T. Krishnan. *The EM Algorithm and Extensions (Wiley Series in Probability and Statistics)*. Wiley-Interscience, 2008.
- [19] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. Muller. Fisher discriminant analysis with kernels. In *IEEE Neural Networks for Signal Processing Workshop*, pages 41–48, 1999.
- [20] A. Robles-Kelly and C. P. Huynh. *Imaging Spectroscopy for Scene Analysis*. Springer, 2013.
- [21] B. Scholkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2001.
- [22] W. S. Stiles and J. M. Burch. Interim report to the Commission Internationale de l’Éclairage Zurich, 1955, on the National Physical Laboratory’s investigation of colour-matching. *Optica Acta*, 2:168–181, 1955.