Published in Image Processing On Line on 2013–12–04.
Submitted on 2012–12–22, accepted on 2013–06–18.
ISSN 2105–1232
This article is available online with supplementary materials, software, datasets and online demo at
https://doi.org/10.5201/ipol.2013.54

E-PLE : an Algorithm for Image Inpainting

Yi-Qing Wang[1]

[1] CMLA, ENS Cachan, France (yiqing.wang@cmla.ens-cachan.fr)

Communicated by Guillermo Sapiro *Demo edited by* Yi-Qing Wang

Abstract

Gaussian mixture is a powerful tool for modeling the patch prior. In this work, a probabilistic view of an existing algorithm piecewise linear estimation (PLE) for image inpainting is presented which leads to several theoretical and numerical improvements based on an effective use of Gaussian mixture.

Source Code

An ANSI C++ implementation of the algorithm has been peer reviewed and is accessible at the IPOL web page of this article[1].

Keywords: inpainting, expectation-maximization

1 Introduction

Inpainting is an interpolation technique developed for repairing a masked image by using information present in the visible parts of the same image.

Historically, one of the first acclaimed works in the field is a paper by Masnou et al. [12] in which the authors propose to connect level lines by minimizing a curvature functional due to their link made clear by the coarea formula. Later a total variation framework [13] is introduced along a similar line whose success can be explained by its insightful choice of functional space to avoid the blur that could be created by a more regular space such as H^1 under an otherwise identical optimization scheme. The subject has since gained some popularity and inspires a paper by Bertalmio et al. [2] where a high order PDE is used to propagate structural information to fill in relatively small gaps. To infer missing textural content, a similarity driven algorithm [9] is devised. The same idea of exploiting redundancy whenever possible has spawned an effective image processing paradigm [4, 6]. Building on these developments on structure and texture inpainting, some efforts [11, 3] have been made to unite these two by performing one preliminary step to separate two types of content before carrying out their respective dedicated procedure.

[1] https://doi.org/10.5201/ipol.2013.54

Another direction of research initiated by Aharon, Elad et al. [1, 10] targets an overcomplete dictionary for sparse representation of image patches. The orientation based K-LLD for image denoising [5] is another example. In a paper by Yu et al. [15] a similar algorithm, called PLE, was designed but intended to solve generic image related inverse problems. In a recent development [16], a Gaussian mixture modeling for patch prior is put forth with a new optimization scheme, which produces impressive results.

In this contribution, motivated in part by the works of Chatterjee et al. [5], Yu et al. [15], and Zoran et al. [16], we present E-PLE, or Enhanced PLE. Using a specialized Gaussian mixture initialized with real-world images, we adapt expectation maximization (EM) algorithm [7] to this particular setting and show its improved performance at inpainting.

Section 2 summarizes PLE. An account of E-PLE is provided in section 3. Section 4 presents the new algorithm outline, followed by several comparative empirical studies in section 5. The appendix is devoted to the EM algorithm.

2 PLE

In this section, PLE [15] is described to highlight its difference with E-PLE. PLE starts with building a number of directional models using synthetic samples and it retains all the eigenvectors from the estimated covariance matrices. Then one additional model is constructed using DCT as its basis to account for textural patches. Contrary to E-PLE, the model means and their covariance eigenvalues are arbitrarily fixed (see algorithm 1).

Algorithm 1 PLE initialization

Parameter: Number of Gaussian models K, patch dimension $\kappa \times \kappa$.

for $k = 0$ to $K-2$ **do**

Create and sample synthetic images

1. Create a binary image B of size 100×100 taking value in $\{0, 255\}$ with two sets $\{(r,u): B(r,u) = 0\}$ and $\{(r,u): B(r,u) = 255\}$ separated by a straight line inclined at $\frac{k}{K-1}\pi$ passing through the center of the image.
2. Blur B with Gaussian kernels of different standard deviations $(\sigma_b)_{1 \le b \le 4}$: $\sigma_b = 2b$ for all b.
3. Draw a large number of $\kappa \times \kappa$ patches from these blurred images to form the patch set $\mathcal{P}_k$.

Compute the statistics

1. Estimate the model mean and covariance:
$$\boldsymbol{\mu}_k = \frac{1}{|\mathcal{P}_k|} \sum_{P \in \mathcal{P}_k} P, \quad \boldsymbol{\Sigma}_k = \frac{1}{|\mathcal{P}_k|} \sum_{P \in \mathcal{P}_k} (P - \boldsymbol{\mu}_k)(P - \boldsymbol{\mu}_k)^T.$$
2. Define the k-th directional basis V_k using the spectral decomposition $\boldsymbol{\Sigma}_k = V_k \Lambda_k V_k^T$.
3. Set $\boldsymbol{\mu}_k = 0$. Replace the first leading eigenvector in V_k by a normalized DC component and apply Gram-Schmitt to orthogonalize the remaining vectors.[2]

end for

To this setup add a textural model whose basis is formed by DCT (with ascending component frequencies). Set its model mean to zero.

Take a sequence of κ^2 positive numbers of exponential decay (a working example: $m \in [0, \kappa^2 - 1] \cap \mathbb{Z} \mapsto 2^{20.5 - 0.5m}$) and make them the eigenvalues of all K Gaussian models just built.

Assume that there are K models in all. For each patch to restore, PLE produces K estimates under individual model assumption and keeps the one with the highest conditional probability to have both the observation and its estimate. This patch is assigned in the meantime to the same model.

Finally, all the models are updated with their assigned estimates. The last two steps, called estimation and maximization by the paper, are then repeated several times before the algorithm terminates (see algorithm 2).

Algorithm 2 PLE

Input: A masked gray image $\tilde{U}$, its mask M.
Parameter: Number of PLE iterations S
Run algorithm 1. Extract all $\kappa \times \kappa$ patches from $\tilde{U}$ and their associated masks from M, the collection of which is denoted by $\tilde{\mathcal{P}}$ and $\mathcal{M}$. With $|\tilde{\mathcal{P}}| = |\mathcal{M}|$, the i-th observed patch and its mask are $\tilde{P}_i$ and $\mathfrak{M}_i$.
for $t = 1$ to S **do**
Estimation:

1. Filter the patch under K model assumptions:

$$\begin{aligned}\forall(i,k),\quad \widehat{P}_i^{(k)} &= \operatorname*{argmax}_P p(P|\tilde{P}_i, \boldsymbol{\mu}_{k,t-1}, \boldsymbol{\Sigma}_{k,t-1}) \\ &= \operatorname*{argmax}_P p(P, \tilde{P}_i|\boldsymbol{\mu}_{k,t-1}, \boldsymbol{\Sigma}_{k,t-1}) \\ &= \operatorname*{argmin}_P \Big(\frac{\|\mathfrak{M}_i P - \tilde{P}_i\|^2}{\sigma^2} + (P-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_{k,t-1}^{-1}(P-\boldsymbol{\mu}_k)\Big).\end{aligned}$$

2. Select a model for each patch:

$$\begin{aligned}k_i &= \operatorname*{argmax}_{0\le k\le K-1} p(\widehat{P}_i^{(k)}, \tilde{P}_i|\boldsymbol{\mu}_{k,t-1}, \boldsymbol{\Sigma}_{k,t-1}) \qquad (1)\\ &= \operatorname*{argmin}_{0\le k\le K-1} \Big(\frac{\|\mathfrak{M}_i \widehat{P}_i^{(k)} - \tilde{P}_i\|^2}{\sigma^2} + (\widehat{P}_i^{(k)}-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_{k,t-1}^{-1}(\widehat{P}_i^{(k)}-\boldsymbol{\mu}_k) + \ln\det\boldsymbol{\Sigma}_{k,t-1}\Big)\end{aligned}$$

which leads to its estimate $\widehat{P}_i = \widehat{P}_i^{(k_i)}$ and assignment to the k_i-th model.[3]

Maximization: Denote $\mathcal{Q}_k$ the set of estimated patches attributed to the k-th model.
for $k = 0$ to $K-1$ **do**
Estimate the model mean and covariance:

$$\boldsymbol{\mu}_{k,t} = \frac{1}{|\mathcal{Q}_k|}\sum_{P\in\mathcal{Q}_k} P,\quad \boldsymbol{\Sigma}_{k,t} = \frac{1}{|\mathcal{Q}_k|}\sum_{P\in\mathcal{Q}_k}(P-\boldsymbol{\mu}_{k,t})(P-\boldsymbol{\mu}_{k,t})^T + \epsilon I$$

where ϵ is a small positive number to ensure the definiteness of $\boldsymbol{\Sigma}_{k,t}$.
end for
end for
Assign equal weights to all restored patches and recover the image.

[2]The implemented PLE leaves out both component substitution and basis orthogonalization because they can cause numerical instability as it is difficult to tell whether a set of vectors are collinear with the computer's limited precision. With DC components removed from the directional bases, PLE could discriminate better.

[3]It would be more natural to incorporate at this stage what we know from the observations. Experiments confirmed that the algorithm yielded better results if the estimated pixels were replaced with the visible ones wherever possible. Hence, we implemented PLE with this additional step.

The Gaussian model used by PLE lacks the mixing weights $\boldsymbol{w}$. for it to be a mixture

$$p(P) = \sum_{k=1}^{N} \boldsymbol{w}_k \mathcal{N}(P \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k). \tag{2}$$

And its synthetic image sampling cannot produce an estimate of that. Thus algorithm 2 is not an EM, a class of algorithms known to increase the likelihood of a mixture over iterations [7]. The absence of the mixing weights to knit the models also implies that the patch assignment step (1) is not statistically founded.

3 E-PLE

3.1 Masked Patch Classification with EM

To set up the Gaussian mixture for E-PLE, we follow the work by Wang et al. [14] and feed it with real-world data (see algorithm 3) so as to shorten the algorithm's learning phase, which is carried out by a version of EM developed for our partially observed data (see appendix). A patch $\tilde{P}$ is then classified using

$$\begin{aligned} k_* &= \operatorname*{argmax}_k p(\tilde{P} \text{ is generated by model } k \mid \tilde{P}) \\ &= \operatorname*{argmax}_k \boldsymbol{w}_k p(\tilde{P} \mid \boldsymbol{\mu}_k, \boldsymbol{F}_k). \end{aligned} \tag{3}$$

which can be shown to minimize Bayes risk [8]. The resulting patch-model association is called a *patch map*. Figure 1 illustrates such a classification example.

3.2 Adaptive Filtering

If a patch $\tilde{P}$ is found with (3) to be best described by the k-th model

$$\tilde{P} = \boldsymbol{F}_k c + \boldsymbol{\mu}_k + \boldsymbol{\sigma} N$$

where $\boldsymbol{F}_k$, c, $\boldsymbol{\mu}_k$, $\boldsymbol{\sigma}$ and N denote its factor loading matrix, random coefficient, model mean, noise standard deviation and a standard Gaussian random vector independent of c, Tikhonov regularization can be applied to construct an estimator. Assume without loss of generality that the column vectors $(\boldsymbol{F}_k^{(m)})_{1\le m\le l_k}$ of $\boldsymbol{F}_k$ are the orthogonal leading eigenvectors of the covariance matrix $\boldsymbol{F}_k\boldsymbol{F}_k^T$. Then the following Wiener filtering scheme with an adjustable parameter ξ controlling the degree of data fit

$$\begin{aligned} \widehat{P} &= \operatorname*{argmin}_{\exists \boldsymbol{\beta}, P=\boldsymbol{F}_k\boldsymbol{\beta}+\boldsymbol{\mu}_k} \sum_{m=1}^{l_k} \|\boldsymbol{F}_k^{(m)}\|^{-2} \left\langle P-\boldsymbol{\mu}_k, \|\boldsymbol{F}_k^{(m)}\|^{-1}\boldsymbol{F}_k^{(m)}\right\rangle^2 + \xi\|\mathfrak{M}P-\tilde{P}\|^2 \\ &= \operatorname*{argmin}_{\exists \boldsymbol{\beta}, P=\boldsymbol{F}_k\boldsymbol{\beta}+\boldsymbol{\mu}_k} \sum_{m=1}^{l_k} \left\langle P-\boldsymbol{\mu}_k, \|\boldsymbol{F}_k^{(m)}\|^{-2}\boldsymbol{F}_k^{(m)}\right\rangle^2 + \xi\|\mathfrak{M}(P-\boldsymbol{\mu}_k-\tilde{P}+\boldsymbol{\mu}_k)\|^2 \end{aligned}$$

defines a ξ-indexed mapping $\tilde{P} \in \mathbb{R}^{\kappa^2} \mapsto \widehat{P} \in \mathbb{R}^{\kappa^2}$. A much neater formulation of the same problem can be obtained with some additional auxiliary variables

$$\begin{aligned} \boldsymbol{\beta} &= [\beta_1, \cdots, \beta_{l_k}]^T, \quad \beta_m = \left\langle P-\boldsymbol{\mu}_k, \|\boldsymbol{F}_k^{(m)}\|^{-2}\boldsymbol{F}_k^{(m)}\right\rangle \\ \tilde{\boldsymbol{\beta}} &= [\tilde{\beta}_1, \cdots, \tilde{\beta}_{l_k}]^T, \quad \tilde{\beta}_m = \left\langle \tilde{P}-\boldsymbol{\mu}_k, \|\boldsymbol{F}_k^{(m)}\|^{-2}\boldsymbol{F}_k^{(m)}\right\rangle. \end{aligned}$$

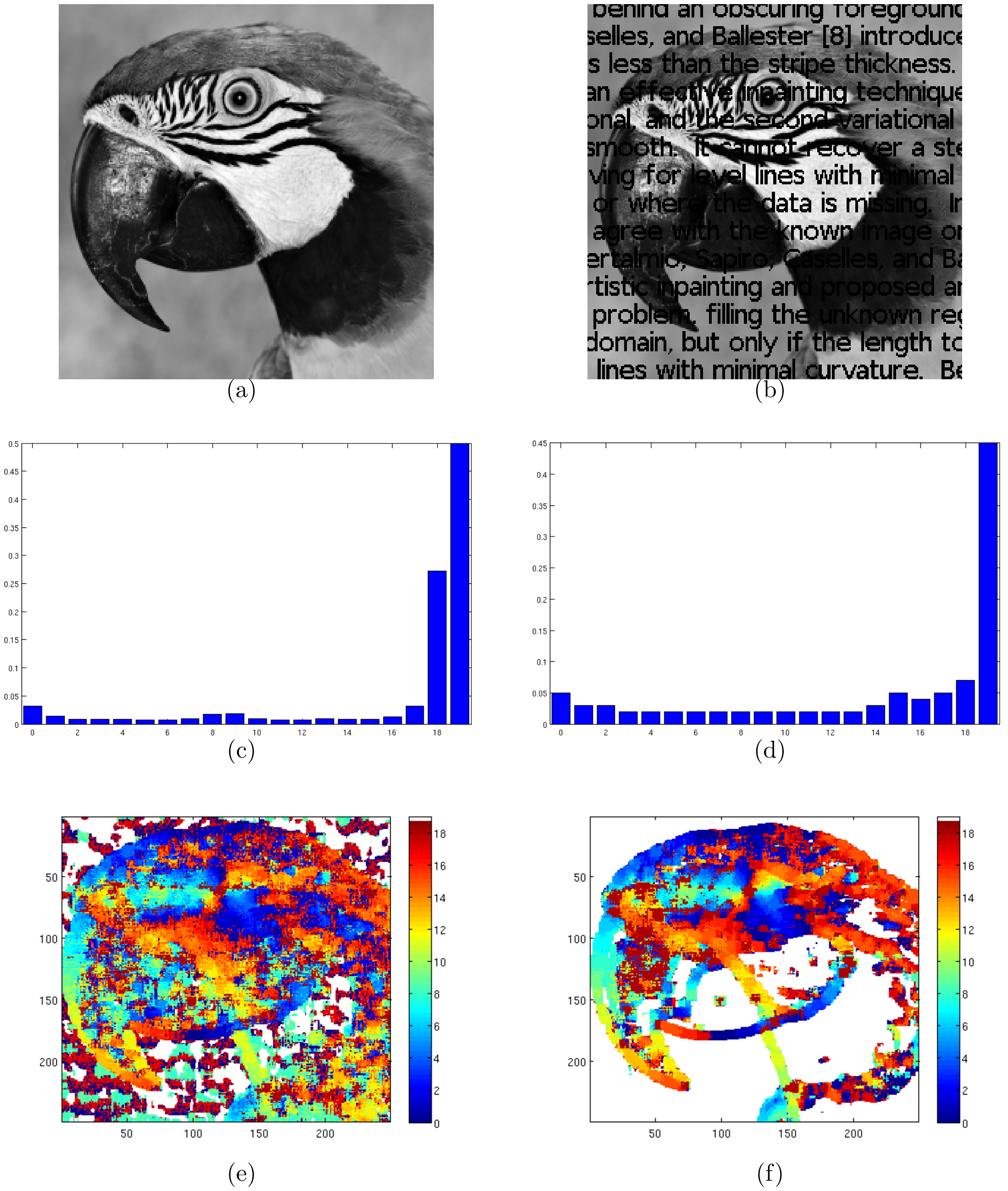


Figure 1: Masked patch classification with EM. (a) original image (b) masked image (c) initial mixing weights (d) mixing weights after three EM iterations (e) patch map formed after one EM iteration (f) patch map formed after three EM iterations. The flat patches are painted white.

Now it follows: $\widehat{P} = \boldsymbol{F}_k \boldsymbol{\beta}_* + \boldsymbol{\mu}_k, \qquad \boldsymbol{\beta}_* = \operatorname{argmin}_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|^2 + \xi \|\mathfrak{M}\boldsymbol{F}_k(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) - \mathfrak{M}R_{\tilde{P}}\|^2$
with the residual: $R_{\tilde{P}} = \tilde{P} - \boldsymbol{\mu}_k - \boldsymbol{F}_k \tilde{\boldsymbol{\beta}}$.
The solution to this quadratic minimization problem is straightforward

$$\boldsymbol{\beta}_* = \xi (I_{l_k} + \xi \boldsymbol{F}_k^T \mathfrak{M} \boldsymbol{F}_k)^{-1} \boldsymbol{F}_k^T \mathfrak{M} (\tilde{P} - \boldsymbol{\mu}_k).$$

Hence the linear estimator

$$\widehat{P} = \xi \boldsymbol{F}_k (I_{l_k} + \xi \boldsymbol{F}_k^T \mathfrak{M} \boldsymbol{F}_k)^{-1} \boldsymbol{F}_k^T \mathfrak{M} (\tilde{P} - \boldsymbol{\mu}_k) + \boldsymbol{\mu}_k. \tag{4}$$

Thanks to the presence of the identity I_{l_k}, the matrix inversion is well defined. In addition, in view of the linear filter's symmetric form, the factor orthogonalization in $\boldsymbol{F}_k$, otherwise required to meet the assumption of the analysis, can be effectively avoided.

4 Algorithm Outline

A recap of E-PLE (algorithms 3 and 4) working on a masked gray image. First, a Gaussian factor mixture is set up using natural images. Next, EM is called upon to infer its parameters from the image to inpaint. Finally, patch map (3) guided linear filters (4) are used to restore patches and hence the image.

For a color image, three color channels can be restored separately before forming the final result. One way to speed up the algorithm in this case however is to make EM only run on one channel and use the resulting patch map to guide the other two. It is the adopted approach in the current implementation.

Algorithm 3 E-PLE Gaussian mixture initialization

Input: Z noiseless natural gray images.
Parameter: Number of mixture components K, patch dimension $\kappa \times \kappa$.
For all $0 \leq k \leq K-1$, set N_k, the number of samples obtained for the k-th model, to 0.
Collect samples:
while $\min_{0\leq k\leq K-1} N_k < 5000$ **do**
  Randomly picks one among Z images and sample a $\kappa \times \kappa$ patch P from it.
  Calculate the eigenvalues (λ_b, λ_s) of $\sum_{(r,u)\in Dom(P)} \nabla P(r,u)(\nabla P(r,u))^T$ together with its eigenvector v associated with λ_b ($\lambda_b \geq \lambda_s$) where $\nabla P(r,u)$ represents the discrete gradient of P at (r,u).
  if $\lambda_b/\lambda_s < t_{\textbf{orient}}$ **then**
    if $\lambda_b < t_{\textbf{flat}}$ **then**
      Assign P to the flat model: $N_{K-1} \leftarrow N_{K-1} + 1$.
    else
      Assign P to the multi-oriented model: $N_{K-2} \leftarrow N_{K-2} + 1$.
    end if
  else
    Determine the orientation $\theta = \psi(\arctan \frac{y}{x})$ with $v = (x,y)^T$ and $\psi(a) = a1_{a\geq 0} + (\pi + a)1_{a<0}$.
    Assign P to the k-th mono-oriented model if $\theta \in [\frac{k}{K-2}\pi, \frac{k+1}{K-2}\pi)$: $N_k \leftarrow N_k + 1$.
  end if
end while
Compute the statistics:
for $k = 0$ to $K-1$ **do**
  Estimate the model prior: $\boldsymbol{w}_k = \frac{N_k}{\sum_{j=0}^{K-1} N_j}$.
  Estimate the model mean and covariance: denote $\mathcal{P}_k$ the set of patches attributed to the k-th model

$$\boldsymbol{\mu}_k = \frac{1}{|\mathcal{P}_k|}\sum_{P\in\mathcal{P}_k} P, \quad \boldsymbol{\Sigma}_k = \frac{1}{|\mathcal{P}_k|}\sum_{P\in\mathcal{P}_k} (P-\boldsymbol{\mu}_k)(P-\boldsymbol{\mu}_k)^T.$$

  Estimate the factor loading matrix: denote l_k the number of factors required by the k-th model. The spectral decomposition $\boldsymbol{\Sigma}_k = V\Lambda V^T$ with $V = [\phi_1, \cdots, \phi_{\kappa^2}]$ and $\Lambda = diag(\lambda_1, \cdots, \lambda_{\kappa^2})$ gives

$$\boldsymbol{F}_k = [\,(\lambda_1 - \sigma^2)^{1/2}\phi_1, \cdots, (\lambda_{l_k} - \sigma^2)^{1/2}\phi_{l_k}\,], \qquad \text{where} \quad \sigma^2 = \frac{1}{\kappa^2 - l_k}\sum_{m=l_k+1}^{\kappa^2} \lambda_m.$$

end for

Algorithm 4 E-PLE

Input: A masked gray image $\tilde{U}$, its mask M.
Parameter: Number of PLE iterations S.
Run algorithm 3. Extract all 8×8 patches from $\tilde{U}$ and their masks from M, the collection of which are denoted by $\tilde{\mathcal{P}}$ and $\mathcal{M}$. With $|\tilde{\mathcal{P}}| = |\mathcal{M}| = N$, observation i and its mask are denoted by $\tilde{P}_i$ and $\mathfrak{M}_i$.
for $t = 1$ to S **do**

Expectation:

1. Compute the mean and covariance of the coefficient posteriors: $\forall 1 \le i \le N,\ 0 \le s_i \le K-1$,

$$\boldsymbol{\Sigma}_{c_i|s_i} = \Big(\frac{\boldsymbol{F}_{s_i,t-1}^T \mathfrak{M}_i \boldsymbol{F}_{s_i,t-1}}{\boldsymbol{\sigma}_{t-1}^2} + I\Big)^{-1},\ \boldsymbol{\mu}_{c_i|s_i} = \boldsymbol{\Sigma}_{c_i|s_i} \frac{\boldsymbol{F}_{s_i,t-1}^T(\tilde{P}_i - \mathfrak{M}_i \boldsymbol{\mu}_{s_i,t-1})}{\boldsymbol{\sigma}_{t-1}^2}.$$

The parameter set Θ_{t-1} is made up of $(\boldsymbol{w}_{k,t-1}, \boldsymbol{F}_{k,t-1}, \boldsymbol{\mu}_{k,t-1})_{0 \le k \le K-1}$ and $\boldsymbol{\sigma}_{t-1}$ for $1 \le t \le S$. The couple $(\boldsymbol{\Sigma}_{c_i|s_i}, \boldsymbol{\mu}_{c_i|s_i})_{1 \le i \le N}$ evolves over time, but for notational convenience, their time index is omitted should no confusion arise.

2. Compute model responsibilities for all patches: $\forall 1 \le i \le N,\ 0 \le s_i \le K-1$,

$$\mathbb{P}_t(s_i|\tilde{P}_i) \propto \boldsymbol{w}_{s_i} \exp\Big(\frac{1}{2}\ln\det \boldsymbol{\Sigma}_{c_i|s_i} + \frac{\boldsymbol{\mu}_{c_i|s_i}^T \boldsymbol{\Sigma}_{c_i|s_i}^{-1} \boldsymbol{\mu}_{c_i|s_i}}{2} - \frac{\|\mathfrak{M}_i \boldsymbol{\mu}_{s_i} - \tilde{P}_i\|^2}{2\boldsymbol{\sigma}^2}\Big).$$

under the constraint $\sum_{k=0}^{K-1} \mathbb{P}_t(s_i = k|\tilde{P}_i) = 1$.

Maximization:

1. Update model priors: $\forall 0 \le k \le K-1,\ \ \boldsymbol{w}_{k,t} = \frac{1}{N}\sum_{i=1}^N \mathbb{P}_t(s_i = k|\tilde{P}_i)$.

2. Update noise variance:

$$\boldsymbol{\sigma}_t^2 = \frac{\sum_{i=1}^N \sum_{k=0}^{K-1} \mathbb{P}_t(s_i = k|\tilde{P}_i) \int dc_i p_i(c_i|s_i = k) \|\tilde{P}_i - \mathfrak{M}_i \tilde{\boldsymbol{F}}_{s_i,t-1} \tilde{c}_i\|^2}{\sum_{i=1}^N |\mathfrak{M}_i|}$$

where $|\mathfrak{M}_i|$ means the number of non-zero entries in $\mathfrak{M}_i$. See (5) for the integral.

3. Update model factors and means: solve the linear equation one row at a time

$$\sum_{i=1}^N \mathfrak{M}_i \tilde{\boldsymbol{F}}_{s_i,t} \mathbb{P}_t(s_i|\tilde{P}_i) \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{c_i|s_i}} + \boldsymbol{\mu}_{c_i|s_i}\boldsymbol{\mu}_{c_i|s_i}^T & \boldsymbol{\mu}_{c_i|s_i} \\ \boldsymbol{\mu}_{c_i|s_i}^T & 1 \end{pmatrix} = \sum_{i=1}^N \mathfrak{M}_i \mathbb{P}_t(s_i|\tilde{P}_i)\tilde{P}_i \begin{pmatrix} \boldsymbol{\mu}_{c_i|s_i}^T & 1 \end{pmatrix}$$

where $\tilde{\boldsymbol{F}}_{s_i,t} := [\boldsymbol{F}_{s_i,t}, \boldsymbol{\mu}_{s_i,t}]$.

end for
Create the patch map: with the parameter set Θ_S, define the patch to model mapping

$$f : \tilde{P}_i \in \tilde{\mathcal{P}} \mapsto \operatorname*{argmax}_{0 \le k \le K-1} \mathbb{P}_{\Theta_S}(s_i = k|\tilde{P}_i).$$

Filter: $\forall \tilde{P}_i \in \tilde{\mathcal{P}}$, take the model $k_i = f(\tilde{P}_i)$ and fill in $\tilde{P}_i$'s missing pixel values with the estimates from

$$\widehat{P}_i = \boldsymbol{F}_{k_i,S}\boldsymbol{\beta_i} + \boldsymbol{\mu}_{k_i,S} \text{ with } \boldsymbol{\beta}_i = \xi(I_{l_k} + \xi \boldsymbol{F}_{k_i,S}^T \mathfrak{M}_i \boldsymbol{F}_{k_i,S})^{-1} \boldsymbol{F}_{k_i,S}^T \mathfrak{M}_i(\tilde{P}_i - \boldsymbol{\mu}_{k_i,S}).$$

Assemble: assign equal weights to all restored patches and recover the image in the usual way.

If an image has a hole larger than the inpainting patch size, only the missing pixels bordering the hole can be estimated by the algorithm while the rest needs to be inferred differently: certainly the aforementioned techniques [12, 2, 9] can be applied here. Since this article is mainly concerned with inpainting images with sporadically missing pixels caused by masks as those shown in the examples, we assume that the sizes of masked parts, though occasionally bigger than 8×8, remain on a somehow manageable scale, which leads to a much simpler inpainting algorithm (algorithm 5).

Algorithm 5 E-PLE for inpainting arbitrary masked images

Input: A masked gray image $\tilde{U}$, its mask M.
Parameter: Number of PLE iterations S.

1. Iterate algorithm 4 without **Filter** and **Assemble**.
2. For those partially masked patches, replace the missing pixels with their estimates and update their associated masks so that those pixels are marked as visible.
3. Aggregate the newly estimated patches to form an inpainted image. Do the same to the masks so as to know if there are pixels left unfilled.
4. **if** some pixels remain masked **then**
   Reduce the newly inpainted image and its mask to patches and assume that all the partially masked patches belong to the textural model. Go back to Step 2.
5. **end if**

---

5 Numerical Results

Figure 2 displays the original images used in our experiments. and table 1 compares the results of different inpainting algorithms in terms of RMSE.

(a) (b) (c) (d)

Figure 2: The images for the empirical studies. (a) parrot (b) shapes (c) barbara (d) frog

Comments:

1. To ensure fairness in comparison, all the algorithms used the same masked images. A random mask has a certain fixed probability for each pixel to become invisible. Both PLE and E-PLE iterate six times. And EPLL refers to the algorithm developed by Zoran et al. [16].

Table 1: Algorithms Comparison in RMSE

text	barbara	frog	parrot	shapes
EPLL	5.7	4.8	6.4	6.5
PLE	**4.2**	**4.3**	**6.1**	5.9
E-PLE	5.0	4.7	6.6	**5.7**

rand 0.2	barbara	frog	parrot	shapes
EPLL	2.5	2.5	4.0	2.8
PLE	**1.7**	**2.0**	**3.7**	2.4
E-PLE	1.9	2.2	3.8	**2.3**

rand 0.4	barbara	frog	parrot	shapes
EPLL	4.7	4.5	**6.7**	5.6
PLE	**3.7**	**4.0**	6.9	5.4
E-PLE	**3.7**	4.1	**6.7**	**4.7**

rand 0.6	barbara	frog	parrot	shapes
EPLL	8.6	7.2	**9.5**	9.6
PLE	10.6	7.1	10.9	11.2
E-PLE	**7.9**	**6.8**	10.0	**8.8**

rand 0.8	barbara	frog	parrot	shapes
EPLL	**15.8**	11.2	15.0	17.6
PLE	20.1	11.0	16.0	19.4
E-PLE	16.9	**10.8**	**14.8**	**16.5**

2. The higher the masking ratio, the worse the recovery in all cases. A higher masking ratio also implies that an algorithm has to guess more so that a well constructed prior knowledge is the most needed. Lacking such a structure, PLE does not do as well as the other two.

3. For natural images, one single iteration of E-PLE usually suffices to achieve a good restoration (see figure 3). More iterations do guarantee an increase in likelihood [7], though not necessarily in RMSE. Yet for a highly degraded image, more iterations could allow better inpainting especially for those images rich in structure such as `barbara`.

4. On the contrary, in case of artificial images, it is desirable to have the algorithm update mixture components through learning in order to adapt itself to this unexpected reality. This explains why EPLL yields a consistently worse result with `shapes` (see figure 4).

5. E-PLE outperforms PLE in general and owing to a reduced set of factors, E-PLE runs faster as well. Moreover, due to a carefully calibrated prior, one iteration of E-PLE is usually sufficient in the sense that more iterations do not bring about significant gain in RMSE to justify the additional computational cost. The same cannot be said of PLE (see figure 5).

6 Appendix

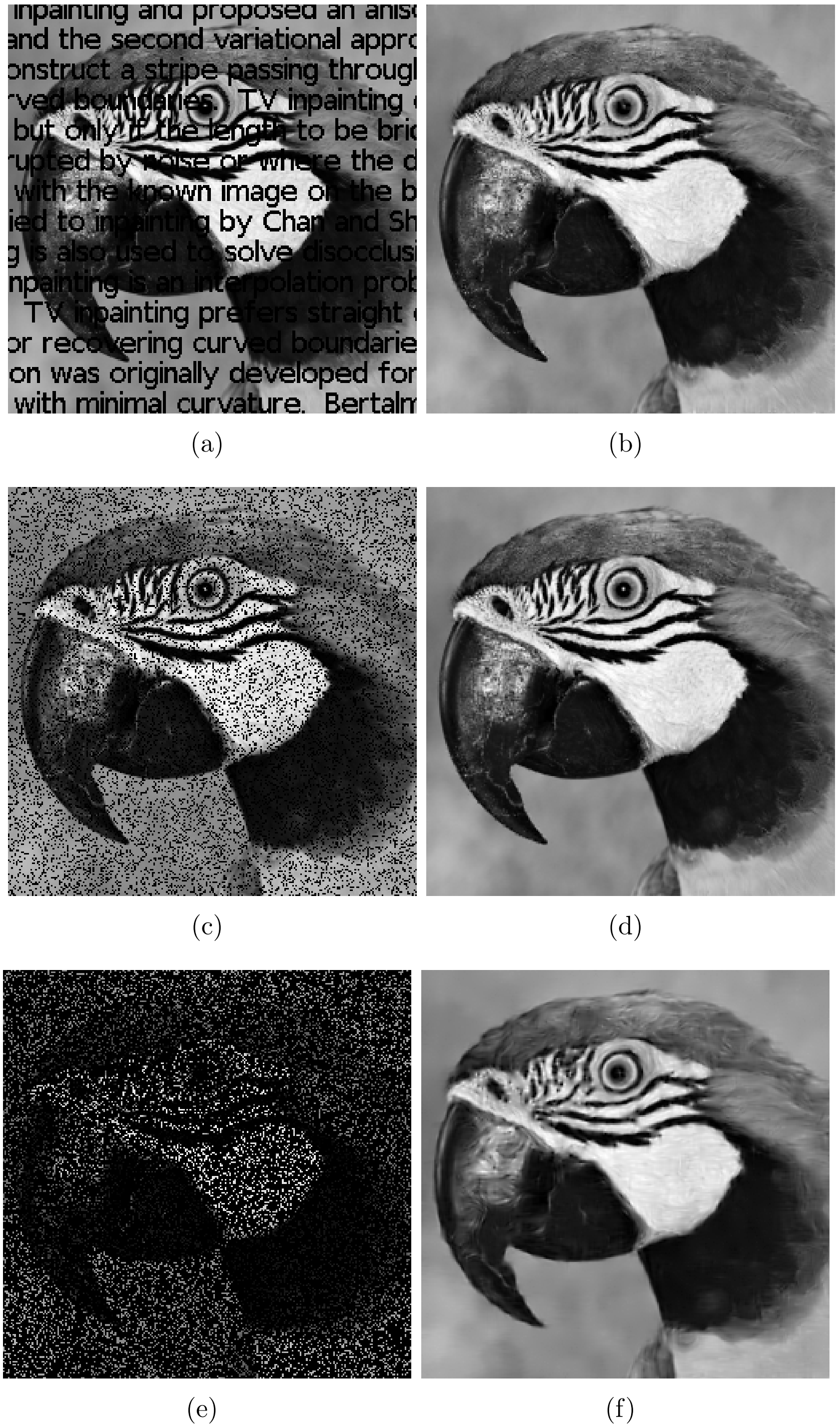
In this section, EM used in E-PLE is derived.
E-PLE's observation model is

$$\tilde{P} = \mathfrak{M}\big(\sum_{k=0}^{K-1} 1_{s=k} P + N\big) = \sum_{k=0}^{K-1} \mathfrak{M}(P+N) 1_{s=k}$$

whereby a patch undergoes noise N and linear distortion by a mask $\mathfrak{M}$. The patch model selector s is distributed according to the mixing weights $\boldsymbol{w}$. and independent of N.
Let Θ be the parameter set containing $\big(\boldsymbol{F}_k, \boldsymbol{\mu}_k, \boldsymbol{w}_k\big)_{0 \leq k \leq K-1}$, and the noise standard deviation $\boldsymbol{\sigma}$.

Lemma 1 *Given the linear model*

$$\forall 1 \leq i \leq N, \;\; \tilde{P}_i = \sum_{k=0}^{K-1} \mathfrak{M}_i (\boldsymbol{F}_k c_i + \boldsymbol{\mu}_k + \boldsymbol{\sigma} n_i) 1_{s_i = k}$$



(a) (b) (c) (d) (e) (f)

Figure 3: E-PLE iterates once on the images on the left to produce those on the right. (a) text masked image (b) inpainted (RMSE = 7.9) (c) randomly masked image (ratio = 0.2) (d) inpainted (RMSE = 5.0) (e) randomly masked image (ratio = 0.8) (f) inpainted (RMSE = 14.9).

the posterior law of the coefficient c_i conditional on $(\tilde{P}_i, s_i)$ is Gaussian and its density $p_i(c_i|s_i)$ is characterized by the covariance matrix and mean

$$\boldsymbol{\Sigma}_{c_i|s_i} = \Big(\frac{\boldsymbol{F}_{s_i}^T \mathfrak{M}_i \boldsymbol{F}_{s_i}}{\boldsymbol{\sigma}^2} + I\Big)^{-1} \quad \text{and} \quad \boldsymbol{\mu}_{c_i|s_i} = \boldsymbol{\Sigma}_{c_i|s_i} \frac{\boldsymbol{F}_{s_i}^T (\tilde{P}_i - \mathfrak{M}_i \boldsymbol{\mu}_{s_i})}{\boldsymbol{\sigma}^2}.$$

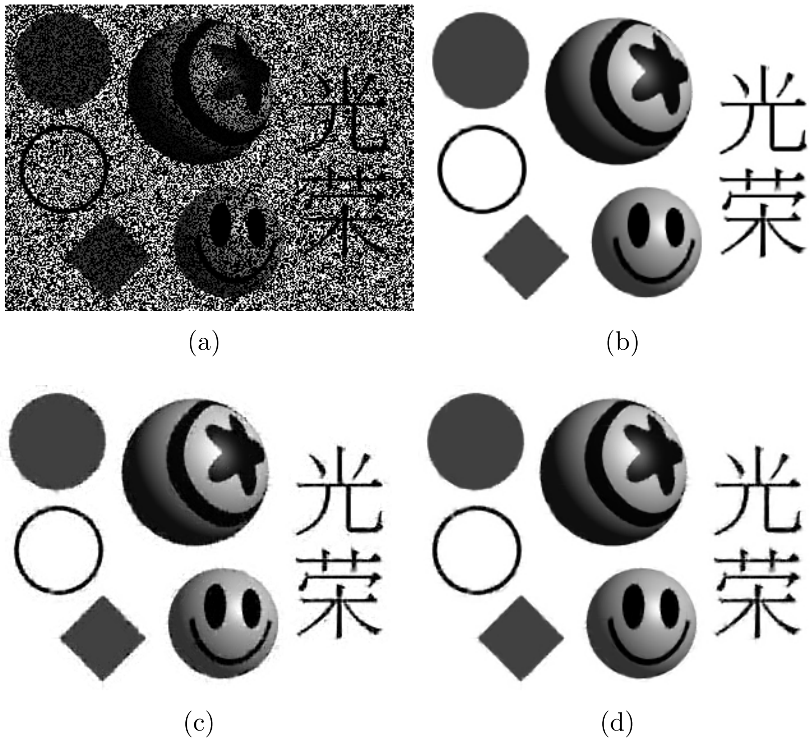


Figure 4: Inpainting an artificial image (a) masked `shapes` (40% pixels visible) (b) inpainted with EPLL (RMSE = 9.6) (c) inpainted with PLE (six iterations and RMSE = 11.2) (d) inpainted with E-PLE (six iterations and RMSE = 8.8).

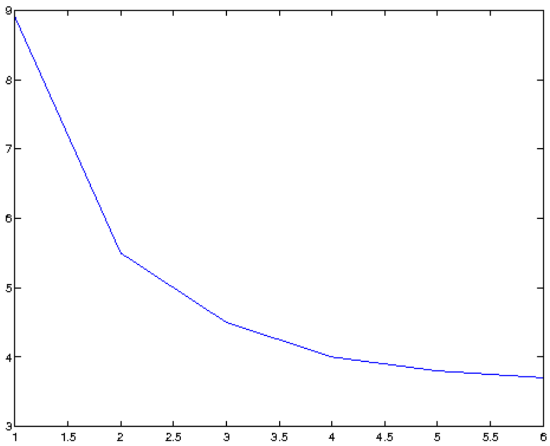


Figure 5: The importance of a good initialization of directional models: an example of how RMSE evolves over iteration with PLE.

Moreover, the density of $\tilde{P}_i$ given s_i is

$$p_\Theta(\tilde{P}_i|s_i) = C_{\mathfrak{M}_i,\boldsymbol{\sigma}^2} \exp\left(\frac{1}{2}\ln\det\boldsymbol{\Sigma}_{c_i|s_i} + \frac{\boldsymbol{\mu}_{c_i|s_i}^T\boldsymbol{\Sigma}_{c_i|s_i}^{-1}\boldsymbol{\mu}_{c_i|s_i}}{2} - \frac{\|\mathfrak{M}_i\boldsymbol{\mu}_{s_i} - \tilde{P}_i\|^2}{2\boldsymbol{\sigma}^2}\right)$$

for some positive constant $C_{\mathfrak{M}_i,\boldsymbol{\sigma}^2}$ only depending on $\mathfrak{M}_i$ and $\boldsymbol{\sigma}^2$.

Proof: an elementary application of Bayes formula implies

$$
\begin{aligned}
&p_\Theta(\tilde{P}_i|s_i) = \int dc_i p_\Theta(c_i|s_i) p_\Theta(\tilde{P}_i|c_i, s_i) = \\
=&\frac{C_{\mathfrak{M}_i,\boldsymbol{\sigma}^2}}{(2\pi)^{l_{s_i}/2}} \int dc_i \exp\Big(- \frac{\|c_i\|^2}{2} - \frac{\|\mathfrak{M}_i(\boldsymbol{F}_{s_i} c_i + \boldsymbol{\mu}_{s_i}) - \tilde{P}_i\|^2}{2\boldsymbol{\sigma}^2}\Big) = \\
=&p_\Theta(\tilde{P}_i|s_i) \int dc_i p_i(c_i|s_i) = \\
=&\frac{C_{\mathfrak{M}_i,\boldsymbol{\sigma}^2}}{(2\pi)^{l_{s_i}/2}} \int dc_i \exp\Big(- \frac{(c_i - \boldsymbol{\mu}_{c_i|s_i})^T \boldsymbol{\Sigma}_{c_i|s_i}^{-1} (c_i - \boldsymbol{\mu}_{c_i|s_i})}{2} + \frac{\boldsymbol{\mu}_{c_i|s_i}^T \boldsymbol{\Sigma}_{c_i|s_i}^{-1} \boldsymbol{\mu}_{c_i|s_i}}{2} - \frac{\|\mathfrak{M}_i \boldsymbol{\mu}_{s_i} - \tilde{P}_i\|^2}{2\boldsymbol{\sigma}^2}\Big) = \\
=&C_{\mathfrak{M}_i,\boldsymbol{\sigma}^2} \exp\Big(\frac{1}{2} \ln\det \boldsymbol{\Sigma}_{c_i|s_i} + \frac{\boldsymbol{\mu}_{c_i|s_i}^T \boldsymbol{\Sigma}_{c_i|s_i}^{-1} \boldsymbol{\mu}_{c_i|s_i}}{2} - \frac{\|\mathfrak{M}_i \boldsymbol{\mu}_{s_i} - \tilde{P}_i\|^2}{2\boldsymbol{\sigma}^2}\Big)
\end{aligned}
$$

Hence the lemma's claims. □
As a by-product, we find the posterior probability

$$
\begin{aligned}
\mathbb{P}_\Theta(s_i|\tilde{P}_i) &\propto p_\Theta(\tilde{P}_i|s_i)\mathbb{P}_\Theta(s_i) \\
&\propto \boldsymbol{w}_{s_i} \exp\Big(\frac{1}{2} \ln\det \boldsymbol{\Sigma}_{c_i|s_i} + \frac{\boldsymbol{\mu}_{c_i|s_i}^T \boldsymbol{\Sigma}_{c_i|s_i}^{-1} \boldsymbol{\mu}_{c_i|s_i}}{2} - \frac{\|\mathfrak{M}_i \boldsymbol{\mu}_{s_i} - \tilde{P}_i\|^2}{2\boldsymbol{\sigma}^2}\Big).
\end{aligned}
$$

Hence $\tilde{P}_i$ is best associated to

$$
k_i = \underset{0 \le s_i \le K-1}{\operatorname{argmax}} \Big(\ln \boldsymbol{w}_{s_i} + \frac{1}{2} \ln\det \boldsymbol{\Sigma}_{c_i|s_i} + \frac{\boldsymbol{\mu}_{c_i|s_i}^T \boldsymbol{\Sigma}_{c_i|s_i}^{-1} \boldsymbol{\mu}_{c_i|s_i}}{2} - \frac{\|\mathfrak{M}_i \boldsymbol{\mu}_{s_i} - \tilde{P}_i\|^2}{2\boldsymbol{\sigma}^2}\Big).
$$

With the parameter set Θ_t known at time t, EM first calculates the conditional expectation of the log-likelihood completed with latent variables $(s_i, c_i)_{1\le i\le N}$ (by abuse of notation, the probabilities $\mathbb{P}$ and densities p are mixed up if the context is clear)

$$
\begin{aligned}
&\sum_{i=1}^N \mathbb{E}_{\Theta_t}\big[\ln \mathbb{P}_\Theta(\tilde{P}_i, s_i, c_i)|\tilde{P}_i\big] = \\
=&\sum_{i=1}^N \sum_{k=0}^{K-1} \mathbb{E}_{\Theta_t}\big[\ln \mathbb{P}_\Theta(\tilde{P}_i, s_i, c_i)|\tilde{P}_i, s_i = k\big]\mathbb{P}_{\Theta_t}(s_i = k|\tilde{P}_i) = \\
=&\sum_{i=1}^N \sum_{k=0}^{K-1} \Big(\mathbb{E}_{\Theta_t}\big[\ln \mathbb{P}_\Theta(\tilde{P}_i, c_i|s_i = k)|\tilde{P}_i, s_i = k\big] + \ln \boldsymbol{w}_k\Big)\mathbb{P}_{\Theta_t}(s_i = k|\tilde{P}_i) = \\
=&\sum_{i=1}^N \sum_{k=0}^{K-1} \Big(\mathbb{E}_{\Theta_t}\big[- \frac{\|c_i\|^2}{2} - \frac{\|\tilde{P}_i - \mathfrak{M}_i(\boldsymbol{F}_k c_i + \boldsymbol{\mu}_k)\|^2}{2\boldsymbol{\sigma}^2} - \frac{|\mathfrak{M}_i|}{2} \ln \boldsymbol{\sigma}^2 + C_{k,\mathfrak{M}_i}|\tilde{P}_i, s_i = k\big] + \ln \boldsymbol{w}_k\Big)\mathbb{P}_{\Theta_t}(s_i = k|\tilde{P}_i)
\end{aligned}
$$

where $|\mathfrak{M}_i|$ is the number of non-zero elements in $\mathfrak{M}_i$ and $C_{k,\mathfrak{M}_i}$ is a constant that depends only on the couple $(k, \mathfrak{M}_i)$. The only variables that remain random in the conditional expectation are $(c_i)_{1\le i\le N}$ and this allows us to put the previous lemma to good use. Since only the second order moments are involved, the computation is straightforward:

$$
\begin{aligned}
&\mathbb{E}_{\Theta_t}\big[\|c_i\|^2|\tilde{P}_i, s_i\big] = tr(\boldsymbol{\Sigma}_{\boldsymbol{c_i|s_i}} + \boldsymbol{\mu}_{c_i|s_i}\boldsymbol{\mu}_{c_i|s_i}^T) := tr(\mathcal{C}_i) \\
&\mathbb{E}_{\Theta_t}\big[\|\tilde{P}_i - \mathfrak{M}_i(\boldsymbol{F}_k c_i + \boldsymbol{\mu}_k)\|^2|\tilde{P}_i, s_i\big] = \|\tilde{P}_i - \mathfrak{M}_i\boldsymbol{\mu}_k\|^2 - 2\big\langle \tilde{P}_i - \boldsymbol{\mu}_k, \mathfrak{M}_i \boldsymbol{F}_k \boldsymbol{\mu}_{c_i|s_i}\big\rangle + tr(\mathcal{C}_i \boldsymbol{F}_k^T \mathfrak{M}_i \boldsymbol{F}_k). \qquad (5)
\end{aligned}
$$

Next, EM maximizes the expectation just obtained w.r.t. the model parameters. For a more compact expression, let us combine the factor loading matrix $\boldsymbol{F}_{s_i}$ with the mean $\boldsymbol{\mu}_{s_i}$ to form $\tilde{\boldsymbol{F}}_{s_i}$ (thus the coefficient c_i is extended by one additional constant equal to 1). Now derive the expectation w.r.t. $\tilde{\boldsymbol{F}}_k$ and set it to zero

$$\frac{\partial}{\partial \tilde{\boldsymbol{F}}_k} \sum_{i=1}^{N} \mathbb{P}_{\Theta_t}(s_i = k|\tilde{P}_i) \int d\tilde{c}_i p_i(\tilde{c}_i|s_i = k) \|\tilde{P}_i - \mathfrak{M}_i \tilde{\boldsymbol{F}}_k \tilde{c}_i\|^2 = 0$$

which leads to

$$\sum_{i=1}^{N} \mathfrak{M}_i \tilde{\boldsymbol{F}}_k \mathbb{P}_{\Theta_t}(s_i = k|\tilde{P}_i) \int d\tilde{c}_i p_i(\tilde{c}_i|s_i = k) \tilde{c}_i \tilde{c}_i^T = \sum_{i=1}^{N} \mathfrak{M}_i \tilde{P}_i \mathbb{P}_{\Theta_t}(s_i = k|\tilde{P}_i) \int d\tilde{c}_i p_i(\tilde{c}_i|s_i = k) \tilde{c}_i^T.$$

Updating $(\tilde{\boldsymbol{F}}_k)_{0 \le k \le K-1}$ amounts to solving a linear equation: denoting by $(M)_q$ the q-th row of a matrix M, we have

$\forall 1 \le q \le \kappa^2$,

$$(\tilde{\boldsymbol{F}}_k)_q \sum_{i=1}^{n} \delta_{\mathfrak{M}_{i(q,q)},q} \mathbb{P}_{\Theta_t}(s_i|\tilde{P}_i) \int d\tilde{c}_i p_i(\tilde{c}_i|s_i) \tilde{c}_i \tilde{c}_i^T = \Big(\sum_{i=1}^{N} \mathfrak{M}_i \tilde{P}_i \mathbb{P}_{\Theta_t}(s_i|\tilde{P}_i) \int d\tilde{c}_i p_i(\tilde{c}_i|s_i) \tilde{c}_i^T \Big)_q$$

where $\delta_{\cdot,\cdot}$ is the Kronecker delta and

$$\int d\tilde{c}_i p_i(\tilde{c}_i|s_i) \tilde{c}_i \tilde{c}_i^T = \begin{pmatrix} \mathcal{C}_i & \boldsymbol{\mu}_{c_i|s_i} \\ \boldsymbol{\mu}_{c_i|s_i}^T & 1 \end{pmatrix}, \quad \int d\tilde{c}_i p_i(\tilde{c}_i|s_i) \tilde{c}_i^T = \begin{pmatrix} \boldsymbol{\mu}_{c_i|s_i}^T & 1 \end{pmatrix}.$$

Hence, if none of the observed patches has a visible pixel at row q, we will not be able to estimate the factors' or means' coordinate at that position. However, it rarely happens if we have a large enough dataset and that the mask behaves sufficiently randomly.
Similarly, the new model prior can be found via the optimization problem

$$\underset{\boldsymbol{w}_1, \cdots \boldsymbol{w}_{K-1}}{\operatorname{argmax}} \sum_{k=0}^{K-1} \ln \boldsymbol{w}_k \sum_{i=1}^{N} \mathbb{P}_{\Theta_t}(s_i = k|\tilde{P}_i) \text{ s.t. } \min_{0 \le k \le K-1} \boldsymbol{w}_k \ge 0 \text{ and } \sum_{k=0}^{K-1} \boldsymbol{w}_k = 1$$

whose solution is

$$\forall 0 \le k \le K-1, \quad \boldsymbol{w}_k = \frac{1}{N} \sum_{i=1}^{N} \mathbb{P}_{\Theta_t}(s_i = k|\tilde{P}_i).$$

Finally, the noise level can be estimated by

$$\frac{\partial}{\partial \boldsymbol{\sigma}^2} \sum_{i=1}^{N} \sum_{k=0}^{K-1} \mathbb{P}_{\Theta_t}(s_i = k|\tilde{P}_i) \int dc_i p_i(c_i|s_i = k) \Big(- \frac{|\mathfrak{M}_i|}{2} \ln \boldsymbol{\sigma}^2 - \frac{\|\tilde{P}_i - \mathfrak{M}_i \tilde{\boldsymbol{F}}_{s_i} \tilde{c}_i\|^2}{2\boldsymbol{\sigma}^2} \Big) = 0$$

whose solution is quite intuitive:

$$\boldsymbol{\sigma}^2 = \frac{\sum_{i=1}^{N} \sum_{k=0}^{K-1} \mathbb{P}_{\Theta_t}(s_i = k|\tilde{P}_i) \int dc_i p_i(c_i|s_i = k) \|\tilde{P}_i - \mathfrak{M}_i \tilde{\boldsymbol{F}}_{s_i} \tilde{c}_i\|^2}{\sum_{i=1}^{N} |\mathfrak{M}_i|}.$$

where the integral is the same as (5).

Image Credits

 Pascal Getreuer, CC-BY

 John D. Willson, CC-BY

  standard test images, copyright unknown

References

[1] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: Design of dictionaries for sparse representation. *Proceedings of Signal Processing with Adaptive Sparse Structured Representations*, 5:9–12, 2005. http://dx.doi.org/10.1109/TSP.2006.881199.

[2] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Proceedings of the 27th annual conference on computer graphics and interactive techniques*, pages 417–424. ACM Press/Addison-Wesley Publishing Co., 2000. http://dx.doi.org/10.1145/344779.344972.

[3] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher. Simultaneous structure and texture image inpainting. *IEEE Transactions on Image Processing*, 12(8):882–889, 2003. http://dx.doi.org/10.1109/TIP.2003.815261.

[4] A. Buades, B. Coll, and J.M. Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4(2):490–530, 2005. http://dx.doi.org/10.1145/344779.344972.

[5] P. Chatterjee and P. Milanfar. Clustering-based denoising with locally learned dictionaries. *IEEE Transactions on Image Processing*, 18(7):1438–1451, 2009. http://dx.doi.org/10.1109/TIP.2009.2018575.

[6] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image restoration by sparse 3D transform-domain collaborative filtering. In *Electronic Imaging 2008*. International Society for Optics and Photonics, 2008. http://dx.doi.org/10.1117/12.766355.

[7] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977.

[8] L. Devroye, L. Györfi, and G. Lugosi. *A probabilistic theory of pattern recognition*, volume 31. Springer, 1996. ISBN 978-0-387-94618-4.

[9] A.A. Efros and T.K. Leung. Texture synthesis by non-parametric sampling. In *International Conference on Computer Vision*, volume 2, pages 1033–1038. IEEE, 1999. http://dx.doi.org/10.1109/ICCV.1999.790383.

[10] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006. http://dx.doi.org/10.1109/TIP.2006.881969.

[11] M. Elad, J.L. Starck, P. Querre, and D.L. Donoho. Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA). *Applied and Computational Harmonic Analysis*, 19(3):340–358, 2005. http://dx.doi.org/10.1016/j.acha.2005.03.005.

[12] S. Masnou and J.M. Morel. Level lines based disocclusion. In *International Conference on Image Processing*, pages 259–263. IEEE, 1998. http://dx.doi.org/10.1109/ICIP.1998.999016.

[13] J. Shen and T.F. Chan. Mathematical models for local nontexture inpaintings. *SIAM Journal on Applied Mathematics*, 62(3):1019–1043, 2002. http://dx.doi.org/10.1137/S0036139900368844.

[14] Y.Q. Wang and J.M. Morel. SURE guided gaussian mixture image denoising. *SIAM Journal on Imaging Sciences*, 6(2):999–1034, 2013. http://dx.doi.org/10.1137/120901131.

[15] G. Yu, G. Sapiro, and S. Mallat. Solving inverse problems with piecewise linear estimators: from Gaussian mixture models to structured sparsity. *IEEE Transactions on Image Processing*, 21(5):2481–2499, 2012. http://dx.doi.org/10.1109/TIP.2011.2176743.

[16] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. In *International Conference on Computer Vision*, pages 479–486. IEEE, 2011. http://dx.doi.org/10.1109/ICCV.2011.6126278.