Image Fusion Using Gaussian Mixture Models

Katie Heaps[1]
heaps006@umn.edu

Josh Koslosky[1]
kosl0052@umn.edu

Glenn Sidle[2]
sidleg@duq.edu

Stacey Levine[2]
sel@mathcs.duq.edu

[1] Department of Mathematics
University of Minnesota
Minneapolis, MN, USA

[2] Department of Mathematics and Computer Science
Duquesne University
Pittsburgh, PA, USA

Recently, a number of works have show that patch-based image features can lead to more robust models than their pixel based counterparts. Critical to their success is that natural image patches can be expressed sparsely in appropriately defined dictionaries which can be tuned to a variety of applications. Yu, Sapiro and Mallat [1] demonstrated that estimating image patches from multivariate Gaussians is equivalent to finding sparse representations in a structured overcomplete PCA-based dictionary and can be solved using a straightforward piecewise linear estimator (PLE). In this work we show how a similar PLE can be formulated to fuse images with various linear degradations and different levels of additive noise.

We consider the degradation model $y = Uf + w$, in which a given image f has undergone some linear degradation U and is corrupted by additive noise $w \sim \mathcal{N}(0,\sigma^2)$. Decomposing f into overlapping $\sqrt{n}\times\sqrt{n}$ *vectorized* patches $f_i \in \mathbb{R}^n$, $i = 1..,I$ and noting that each patch may have a unique linear degradation U_i and unique additive noise w_i, we can express the degraded image patches as $y_i = U_i f_i + w_i$ for $i = 1,..,I$. Recovering f from y then becomes the problem of recovering f_i from y_i and rebuilding f from $\{f_i\}_{i=1}^I$.

In this work we propose a model for recovering f from J images, $y^j = U^j f + w^j$, $j = 1,...,,J$, where the linear degradations $U^j$and noise levels $w^j \sim \mathcal{N}(0,\sigma_j^2)$ may vary amongst images. Following the single image recovery model proposed in [1], we show that given a fixed number of multivariate Gaussians parametrized by their means $\mu_k \in \mathbb{R}^n$ and covariance matrices $\Sigma_k \in \mathbb{R}^{n\times n}$, $k = 1,...,K$, each patch f_i (for $i = 1,...,I$) can be estimated by maximizing the log *a posteriori* probability $p(f_i|y_i^1,...,y_i^J,\Sigma_k)$, which is equivalent to solving

$$(\hat{f}_i,\hat{k}_i) = \arg\min_{f,k} \sum_{j=1}^{J}\left(\frac{1}{\sigma_j^2}||U_i f - y_i^j||^2\right) + f^T\Sigma_k^{-1}f + \log|\Sigma_k|. \quad (1)$$

The best estimate from each Gaussian is found using a simple linear filter

$$f_i^k = \sum_{j=1}^{J} W_{k,i}^j y_i^j,\ k = 1,...,K \quad (2)$$

where $W_{k,i}^j = \Sigma_k\left(\sum_{l=1}^J \frac{1}{\sigma_l^2}(U_i^l)^T U_i^l \Sigma_k + Id\right)^{-1}(U_i^j)^T$.

For each patch, the solution $\hat{k}_i$ of (1) is found by simply evaluating the functional at f_i^k (from (2)) for $k = 1,...,K$, and choosing $\hat{k}_i$ to be the one that yields the smallest value. Then the minimizing patch for (1) is

$$\hat{f}_i = f_i^{\hat{k}_i}. \quad (3)$$

The patches $\{\hat{f}_i\}_{i=1}^I$ are overlapping and averaged at each location to produce the final result, $\hat{f}$. So each pixel will typically be estimated by a mixture of Gaussians.

Once the model selection $\hat{k}_i$ and signal estimation $\hat{f}_i$ are known for each patch, setting $C_k = \{i|\hat{k}_i = k\}$ for $k = 1,...,K$, the new Gaussian parameters are computed by

$$\hat{\mu}_k = \frac{1}{|C_k|}\sum_{i\in C_k}\hat{f}_i \ \text{ and }\ \hat{\Sigma}_k = \frac{1}{|C_k|}\sum_{i\in C_k}(\hat{f}_i - \hat{\mu}_k)^T(\hat{f}_i - \hat{\mu}_k). \quad (4)$$

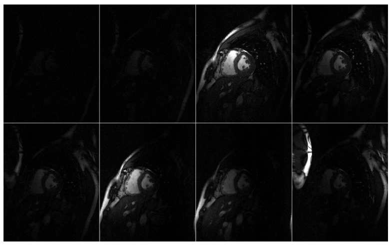
The steps (1) and (4) are iterated until convergence, thus the initialization of the parameters $(\mu_k,\Sigma_k)_{k=1,...,K}$ is critical. We follow the initialization proposed in [1] based on clustering edges at fixed orientations.

For each $k \in \{1,...,K\}$, the solution f_i^k in (2) can also be interpreted as a sparse patch-based representation. Specfically, $f_i^k = B_k a_i^k$ where

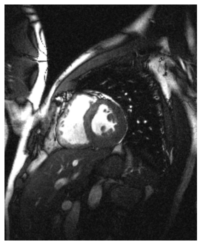
$$a_i^k = \arg\min_{a\in\mathbb{R}^n}\left(\sum_{j=1}^{J}\frac{1}{\sigma_j^2}||U_i^j B_k a - y_i^j||^2 + \sum_{m=1}^{n}\frac{|a[m]|^2}{\lambda_m^k}\right). \quad (5)$$

Here B_k is the orthonormal PCA basis diagonalizing the covariance matrix Σ_k corresponding to the k^{th} multivariate Gaussian, and the λ_m^k are the corresponding eigenvalues in decreasing order. The weighted l_2 norm in (5) favors dominant eigenvectors, and since the eigenvalues rapidly decrease in magnitude [1], it induces sparsity. Furthermore, the PCA basis B_k and the eigenvalues incorporate information from the data, so this model promotes collaborative filtering where the dictionary is learned from the data.

The model is flexible, and the level of smoothing can also be adapted to each patch. A number of experimental results demonstrate the success of the model, some of which are included below.



(a) 8 coil MRI data, co. Mark Griswold.



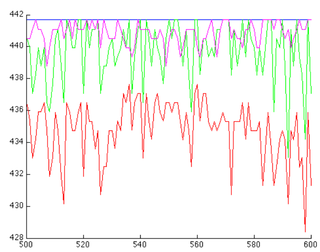
(b) Our fusion result.

(c) blurry image, PSNR=27.5454

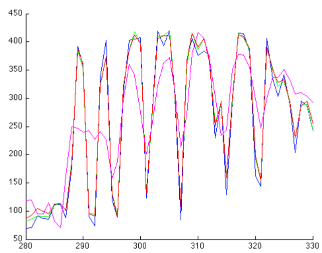
(d) noisy image, PSNR=29.2852

(e) original image

(f) fusion result, PSNR=35.7534



(g) flat region of the sky



(h) lettering on the plane

Figure 1: (a-b) fusing 8 coil MRI data; (c-f) fusing a noisy/blurry pair; (g-h) typical cross-sections of the magnitude of the intensity for the blurry/noisy pair (c)-(d): blue=original, green=fused result (PSNR=35.7534), red=denoised only (PSNR=33.4498), magenta=deblurred only (PSNR=31.3776).

[1] Guoshen Yu, Guillermo Sapiro, and Stephan Mallat. Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity. *IEEE Trans. Image Process.*, 21(5): 2481– 2499, 2012.