Joint Noise Level Estimation from Personal Photo Collections

YiChang Shih[1,2]   Vivek Kwatra[1]   Troy Chinen[1]   Hui Fang[1]   Sergey Ioffe[1]

[1]Google Research   [2]MIT CSAIL

Abstract

Personal photo albums are heavily biased towards faces of people, but most state-of-the-art algorithms for image denoising and noise estimation do not exploit facial information. We propose a novel technique for jointly estimating noise levels of all face images in a photo collection. Photos in a personal album are likely to contain several faces of the same people. While some of these photos would be clean and high quality, others may be corrupted by noise. Our key idea is to estimate noise levels by comparing multiple images of the same content that differ predominantly in their noise content. Specifically, we compare geometrically and photometrically aligned face images of the same person.

Our estimation algorithm is based on a probabilistic formulation that seeks to maximize the joint probability of estimated noise levels across all images. We propose an approximate solution that decomposes this joint maximization into a two-stage optimization. The first stage determines the relative noise between pairs of images by pooling estimates from corresponding patch pairs in a probabilistic fashion. The second stage then jointly optimizes for all absolute noise parameters by conditioning them upon relative noise levels, which allows for a pairwise factorization of the probability distribution. We evaluate our noise estimation method using quantitative experiments to measure accuracy on synthetic data. Additionally, we employ the estimated noise levels for automatic denoising using "BM3D", and evaluate the quality of denoising on real-world photos through a user study.

1. Introduction

People capture more photos today then ever before, thanks to the rapid proliferation of mobile devices with cameras. A common problem among personal photos is the presence of noise, especially in photos captured in low light using mobile cameras. Recent progress in image denoising has been impressive [2, 3], but many of these methods require accurate image noise levels as input parameters. Fig. 1 shows that these noise parameters can have a significant im-

(a) Noisy input   (b) Our method (σ=14.6)

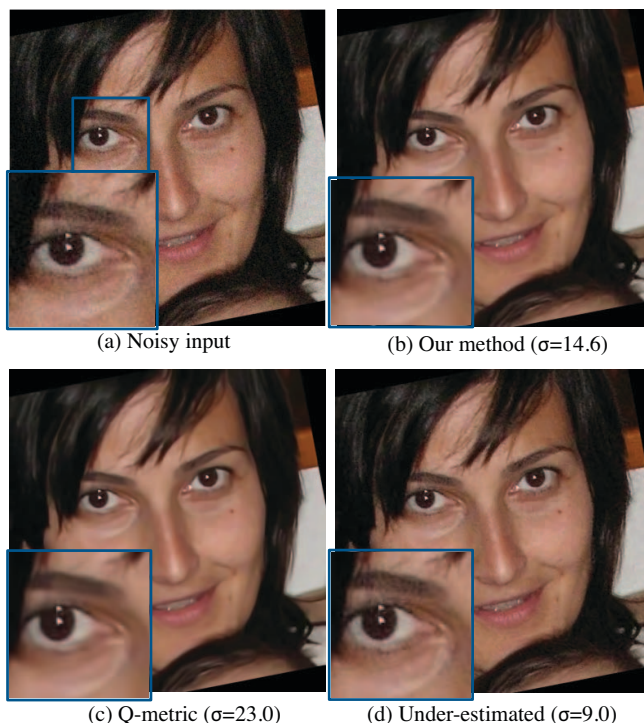(c) Q-metric (σ=23.0)   (d) Under-estimated (σ=9.0)

Figure 1: BM3D denoising using various noise parameters. (a) Noisy input image. (b) BM3D result using our estimated noise parameter. Our result is the sharpest while still being noise-free. (c) BM3D result chosen by Q-metric [15] which over-estimates noise, results in extra smoothing. (d) Insufficiently denoised BM3D result for under-estimated noise.

pact on the denoised result.

Estimating the noise level from a single image is fundamentally ill-posed. Existing methods for noise estimation from a single image [7, 10, 16] often make certain assumptions about the underlying image model. Even if the noise model is appropriate, it can still be challenging to estimate the true noise level, because separating noise from the unknown noise-free reference image remains underconstrained.

Our work is based on the observation that the lack of a noise-free reference image can be dealt with by processing all photos in an album jointly. This is in contrast to previous

methods, which focus on individually estimating noise levels from single images. Most personal photo albums consist of multiple faces of the same people, occurring under different conditions, *e.g.* some images may be taken with high-end cameras or under better lighting conditions, while others may be more noisy. The key idea is to estimate the relative noise between these images, and then treat the comparatively cleaner images as *references* for obtaining absolute noise levels in all images.

We propose a two-stage algorithm to jointly determine the noise levels for all face images in an album. In the first stage, we estimate the most probable relative noise levels between each image pair, and show that this can be done by combining relative variances over corresponding patches in a probabilistic fashion. In the second stage, we employ a pairwise Markov random field, conditioned upon the relative noise levels obtained in the previous step to model the joint probability over all absolute noise levels. This joint optimization is then solved using weighted least squares over a fully connected graph, where each node represents a face image, and each edge represents the relative noise level between a pair of images.

Quantitatively, we show that our method performs better than Liu et al's method [7] on synthetic noisy data. On real world data, we show how to use it to perform automatic parameter selection for the state-of-the-art BM3D denoising algorithm [3]. We evaluate the effectiveness of our approach with a user study. In particular, we compare against the Q-metric [15] based approach for best image selection from among multiple BM3D results. Fig. 1 demonstrates a denoising result generated by BM3D using our automatically estimated noise parameter and shows comparisons.

2. Related Work

Proper knowledge of image noise level can be crucial for many denoising algorithms. One can obtain the noise level for known cameras if both the EXIF file and raw image are available [4]. But in practice, this information may not always be present, requiring estimation of noise directly from images. Estimating noise levels from a single image relies on assumed image models, such as the piecewise linear model in [7], or needs to restore the clean noise-free image simultaneously with estimation [9], which can be ill-posed. By contrast, we use multiple images of the same subject, which makes the estimation problem relatively well-posed and results in better accuracy.

Rank *et al.* [10] obtain good results by first convolving the image with a Laplacian filter, and then separating noise from the edges for estimation. However their method tends to degrade when the images have more textured regions. To overcome their weakness, Zoran *et al.* [16] exploit scale invariance in natural statistics, and improved the estimation accuracy over textured images by analyzing kurtosis varia-
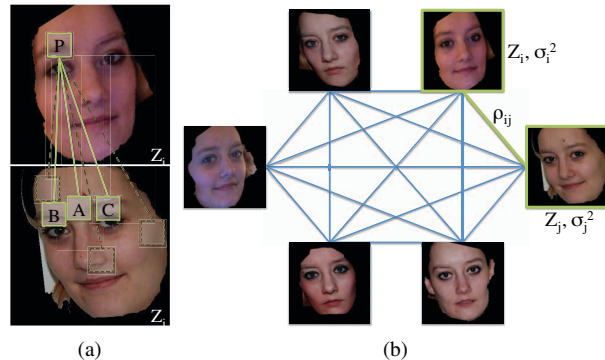

(a)   (b)

Figure 2: Illustration of our algorithm. (a) Patch-based estimation of relative noise ρ_{ij} pools estimates overs a small number of patch pairs: $P - A, P - B, P - C$ in this example. (b) Absolute noise levels σ_i^2 are estimated by jointly optimizing over a fully connected graph.

tion of image statistics. While the above methods rely on the accuracy of natural image modeling, our work aims to take advantage of correlations between images of similar content, specifically faces.

An alternative to noise estimation is to select the parameter by maximizing certain subjective non-reference quality measurement of the denoised output [15]. These subjective measurements can produce high quality results, but require dense sampling of the noise parameter space, which increases the computational cost.

There has been other work on exploiting faces for image enhancement. Joshi *et al.* [5] employ user-specific face databases for image restoration. Shah and Kwatra [13] exploit albums and photo bursts for facial expression enhancement. The ultimate application of our work is also enhancement (by denoising). However, our main contribution lies in estimating the noise.

3. Two-stage Joint Noise Level Estimation

Our method works on multiple face images of the same person captured under various noise levels. Given a collection of n face images $\{Z_i\}_{i=1:n}$ from an album, all containing the user's face, our goal is to estimate the noise levels for all those images jointly[1]. We assume an additive white noise model, *i.e.* the noise is assumed independent of the image content, with zero mean and fixed variance. For now, we treat these as single channel images. Color channels are incorporated by taking the the mean variance across all color channels[2], except for the normalization procedure described in section 3.3. Each observed image Z_i can be

---

[1]We describe collection of face images in section 3.4
[2]This can be improved by considering noise variation across color channels and by pixel intensity, but we leave that for future work.

modeled as:

$$Z_i = X_i + N_i \qquad (1)$$

where X_i is the underlying (unknown) clean image, and N_i is the noise layer, with the noise at pixel p denoted by:

$$\eta_{ip} \sim \mathcal{N}(0, \sigma_i^2). \qquad (2)$$

To determine the noise parameters $\{\sigma_i\}_{i=1:n}$ for all images, we want to maximize the joint probability given face images:

$$\{\sigma_i^2\} = \{\nu_i^*\} = \underset{\{\nu_k\}_{k=1:n}}{\arg\max} \, P(\nu_1, \nu_2, ..., \nu_n | Z_1, Z_2, ..., Z_n) \quad (3)$$

We denote σ_i^2 by ν_i here. While single image noise estimation methods focus on individually modeling $P(\nu_i | Z_i)$, we aim to model the joint distribution over $\{\nu_i\}_{i=1:n}$ given the image set $\{Z_i\}_{i=1:n}$. We further focus on pair-wise interactions between images, where each image pair i, j is used to estimate the relative noise between those images, denoted by ρ_{ij}. This relative noise acts as a latent variable in our formulation, and allows us to simplify the joint noise estimation into a two stage process, as described below.

Denoting the sets $\{\nu_i\}$ as $\boldsymbol{\nu}$, $\{Z_i\}$ as $\boldsymbol{Z}$, and $\{\rho_{ij}\}$ as $\boldsymbol{\rho}$, the RHS of Eq. 3 can be written as:

$$\arg\max_{\boldsymbol{\nu}} P(\boldsymbol{\nu}|\boldsymbol{Z}) = \arg\max_{\boldsymbol{\nu}} \int_{\boldsymbol{\rho}} P(\boldsymbol{\nu}|\boldsymbol{\rho}, \boldsymbol{Z}) P(\boldsymbol{\rho}|\boldsymbol{Z}) d\boldsymbol{\rho} \quad (4)$$

Eq. 4 marginalizes over all possible values of ρ_{ij} for all i, j, which makes the optimization intractable. Therefore, instead of marginalizing, we assume a unimodal distribution, allowing Eq. 4 to be approximated via sampling at the most likely $\boldsymbol{\rho}$ conditioned upon $\boldsymbol{Z}$. We can then write Eq. 3 as:

$$\boldsymbol{\nu}^* = \arg\max_{\boldsymbol{\nu}} P(\boldsymbol{\nu}|\boldsymbol{Z}) \approx \arg\max_{\boldsymbol{\nu}} P(\boldsymbol{\nu}|\boldsymbol{Z}, \boldsymbol{\rho}^*) P(\boldsymbol{\rho}^*|\boldsymbol{Z}) \quad (5)$$

$$\text{s.t. } \boldsymbol{\rho}^* = \arg\max_{\boldsymbol{\rho}} P(\boldsymbol{\rho}|\boldsymbol{Z}) \quad (6)$$

Eq. 5 and 6 can be decoupled into a two-stage optimization problem. In the first stage (section 3.1), we perform relative noise estimation between image pairs, where the relative noise ρ_{ij} provides a probabilistic estimate of $\sigma_i^2 - \sigma_j^2$. Once the relative noise between image pairs is obtained, we solve for the absolute noise level estimates σ_i^2 in the second stage (section 3.2) by optimizing over a fully connected graph of all valid images. See Fig. 2 for an illustration.

3.1. Estimating relative noise ρ_{ij}

To optimize Eq. 6, we employ a directed graphical model as illustrated in Fig. 3 to decompose $P(\boldsymbol{\rho}|\boldsymbol{Z})$ into pairwise terms:

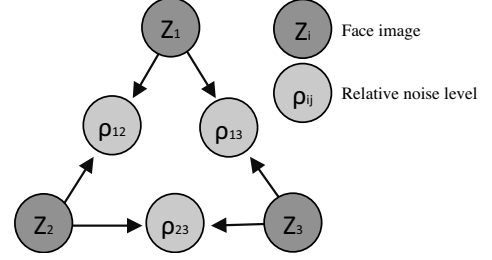$$P(\boldsymbol{\rho}|\boldsymbol{Z}) = \prod_{ij, i<j} P(\rho_{ij}|Z_i, Z_j) \qquad (7)$$

Figure 3: We employ a directed graphical model for $P(\boldsymbol{\rho}|\boldsymbol{Z})$. Pairwise relative noise ρ_{ij} is only dependent on the two images Z_i and Z_j. Therefore, we can maximize the joint probability by individually maximizing $P(\rho_{ij}|Z_i, Z_j)$. Here we show an illustrative example for three face images.

Since ρ_{ij} only depends on the known Z_i and Z_j, we can optimize each $P(\rho_{ij}|Z_i, Z_j)$ term in Eq. 7 independently.

Our hypothesis is that if the face images Z_i and Z_j share the same content, the differences between them should be accounted for by variations in the image capture process, such as changes in lighting conditions, camera viewpoint, facial pose and expression, and noise. To extract the relative noise reliably, we need to factor out the remaining variables. Hence, we coarsely pre-align the faces via an affine transform[3] as described in section 3.4, and compensate for lighting variations as well (described later in section 3.3). Even so, the alignment may not always be perfect, and even if it were, our white noise model may not hold uniformly over the entire image. Therefore, we eschew operating on global images any further, and instead take a patch-based approach to address these issues. As the noise is assumed to be independent of image content in our model, the variance within image patches[4] at locations p and q in images Z_i and Z_j, respectively, can be expressed as:

$$\text{Var}(\mathbf{z}_{ip}) = \text{Var}(\mathbf{x}_{ip}) + \sigma_i^2 + \omega_i \qquad (8)$$

$$\text{Var}(\mathbf{z}_{jq}) = \text{Var}(\mathbf{x}_{jq}) + \sigma_j^2 + \omega_j. \qquad (9)$$

Here $\mathbf{z}_{ip}$ and $\mathbf{z}_{jq}$ denote patch pixels in the observed images, while $\mathbf{x}_{ip}$ and $\mathbf{x}_{jq}$ refer to the corresponding pixels in the unknown clean images. ω_i and ω_j denote an uncertainty (or residual variance) term that models variations between the true noise and our white noise model. If we can find p and q that form a matching pair of patches, i.e. correspond to the same region within the face, then we can assume that $\text{Var}(\mathbf{x}_{ip}) \approx \text{Var}(\mathbf{x}_{jq})$. It follows from Eq. 8 and Eq. 9 that:

$$\rho_{ij} \triangleq \sigma_i^2 - \sigma_j^2 + \omega_{ij} \approx \text{Var}(\mathbf{z}_{ip}) - \text{Var}(\mathbf{z}_{jq}) = \xi_{ij}^{pq}. \quad (10)$$

where the relative uncertainty $\omega_{ij} = \omega_i - \omega_j$ relates the relative noise ρ_{ij} to absolute levels σ_i^2 and σ_j^2, and is used later in section 3.2.

---

[3]These normalization steps introduce a bias into the noise estimation procedure, which we need to correct for, as described later.

[4]We use 16×16 patches in our experiments.

In practice, not all corresponding pairs $p \in Z_i$ and $q \in Z_j$ are equally good matches. We therefore marginalize the relative noise probability over all corresponding patch pairs:

$$P(\rho_{ij}|Z_i, Z_j) = \sum_{p,q} P(\rho_{ij}|p,q)P(p,q). \quad (11)$$

$$P(\rho_{ij}|p,q) \propto \exp(-\|\rho_{ij} - \xi_{ij}^{pq}\|^2)$$

is the probabilistic version of Eq. 10. $P(p,q)$ is the probability that the pair (p,q) forms a good match, modeled by a zero-mean Gaussian distribution over patch similarities:

$$P(p,q) \propto c_{pq} \triangleq \exp(-\|\mathbf{z}_{ip} - \mathbf{z}_{jq}\|^2 \kappa_{pq}). \quad (12)$$

The coefficient $\kappa_{pq} = \frac{0.1}{\min(g_p, g_q)}$, where g_p and g_q denote the average patch gradient magnitudes at p and q. The motivation here is that patches should have higher similarity in smooth regions to be considered a match as compared to textured regions. We optimize Eq. 11 by maximizing the *expected* log probability, obtained by employing the Jensen's inequality [1]:

$$\max \quad -\log \sum_{p,q} P(\rho_{ij}|p,q)P(p,q)$$

$$\leq \quad \max \quad \sum_{p,q} -P(p,q) \log P(\rho_{ij}|p,q)$$

$$= \quad \max \quad \frac{\sum_{p,q} c_{pq} \|\rho_{ij} - \xi_{ij}^{pq}\|^2}{\sum_{p,q} c_{pq}}. \quad (13)$$

which has the weighted least squares solution:

$$\rho_{ij}^* = \frac{\sum_{p,q} c_{pq} \xi_{ij}^{pq}}{\sum_{p,q} c_{pq}}. \quad (14)$$

This formulation bears similarity to kernel-based regression techniques, which use data-dependent weights for adaptive image processing [14, 6]. Note that an exhaustive marginalization would require computing probabilities for all possible patch pairs. In practice, however, it is sufficient to consider only a few patches q (3 in our experiments) in the vicinity of p as the images are already coarsely aligned.

3.2. Estimating absolute noise σ_i^2

Given $\boldsymbol{\rho}^*$ from the previous stage, we estimate the absolute noise levels σ_i^2 for all images by optimizing Eq. 5. We model $P(\boldsymbol{\nu}|\boldsymbol{Z}, \boldsymbol{\rho}^*)$ in Eq. 5 by employing a pair-wise Markov random field defined over $\boldsymbol{\nu}$. Specifically, the joint probability $P(\boldsymbol{\nu}|\boldsymbol{Z}, \boldsymbol{\rho}^*)$ is defined as below:

$$P(\boldsymbol{\nu}|\boldsymbol{Z}, \boldsymbol{\rho}^*) = \prod_{ij} \phi(\nu_i, \nu_j, \rho_{ij}^*, Z_i, Z_j) \quad (15)$$

where ϕ is modeled as a Gaussian distribution that takes the uncertainty of Z_i and Z_j sharing the same content into

| (a) | (b) | (c) |

Figure 4: We normalize (a) relative to (b) by applying the color transform in Eq. 23, resulting in (c). The noise gain g_i from (a) to (c) is 9.02.

account. We introduced this uncertainty in Eq. 10 and take it to subsume alignment errors or mismatches between noise models. Substituting ν for σ^2 and taking the optimal ρ_{ij}^*, we have from Eq. 10:

$$\rho_{ij}^* = \nu_i - \nu_j + \omega_{ij} \quad (16)$$

$$\Rightarrow \quad \phi \propto \exp\left(-w_{ij}\|\nu_i - \nu_j - \rho_{ij}^*\|^2\right) \quad (17)$$

where the weight $w_{ij} = 1/\omega_{ij}$ is higher for a pair of images when the uncertainty between them is lower. We propose a heuristic function for w_{ij} based on the difference between Z_i and Z_j as $w_{ij} = 1/(\|Z_i - Z_j\|_1 + \epsilon)$, where $\epsilon = 0.05$. This entails solving another weighted least squares problem, albeit at the image level over a fully connected graph:

$$\{\sigma_i^2\} = \arg\min_{\{\nu_k\}_{k=1:n}} \sum_{i \neq j} w_{ij} \| \nu_i - \nu_j - \rho_{ij}^* \|^2 \quad (18)$$

Note that Eq. 18 is under-determined. For example, adding a constant to all ν_i variables results in the same objective. This issue can be addressed in two ways. The first approach is to manually annotate one or more images as clean images, and force those images Z_{i*} to have a noise level of zero. Eq. 18 then becomes a constrained optimization problem:

$$\min \sum_{i \neq j} w_{ij} \| \nu_i - \nu_j - \rho_{ij}^* \|^2 \quad \text{s.t. } \nu_{i*} = 0, \quad (19)$$

which is equivalent to solving the following linear system:

$$\mathbf{M}\nu = \mathbf{r}, \quad \text{where} \quad (20)$$

$$\mathbf{M}_{i,j} = \begin{cases} -w_{ij} & \text{if } i \neq j \\ \sum_j w_{ij} & \text{if } i = j \end{cases} \quad i, j \notin i^* \quad (21)$$

$$\mathbf{r}_i = \sum_j w_{ij}\rho_{ij}^*, \quad i \notin i^* \quad (22)$$

A second, completely automatic, approach is to arbitrarily choose one image as the zero-noise image, solve the above system, and then offset the computed noise levels by subtracting away the lowest noise value. This is equivalent to choosing the image with the lowest noise as the only clean image. We use the automatic method during the evaluations.

3.3. Color normalization

Eq. 10 assumes that the underlying clean images X_i and X_j have similar variances at matching locations. However, this assumption would be violated if the images were captured under different lighting conditions, *e.g.* see Fig. 4. We address this issue by applying a global color transform to each image in the album, which would bring its mean color into alignment with a randomly chosen reference image. For each image Z_i, we compute the normalized image as:

$$Z_i^n(p) = \mathbf{A_i}Z_i(p) + \mathbf{b_i} \tag{23}$$

at pixel p, where $\mathbf{A_i}$ and $\mathbf{b_i}$ are the 3×3 color transfer matrix and 3×1 color shift vector, respectively, between Z_i and the reference image. The color transforms are computed by matching second-order image statistics in the CIE XYZ space [11].

Noise levels are then estimated using Z_i^n instead of Z_i, but this introduces a gain g_i in the variance due to the scaling caused by $\mathbf{A_i}$, equivalent to the mean of the eigenvalues of $\mathbf{A_i}^T\mathbf{A_i}$. We adjust the noise estimate as: $\sigma_i^2 \leftarrow \sigma_i^2/g_i$.

3.4. Collecting face images

To collect these face images from an album, we employ automatic methods similar to previous works employing facial analysis [5, 13]. We describe some of these methods here briefly for clarity, but do not claim any contributions on this part. Automatic face image collection is done by detecting all the faces and clustering them based on visual similarity, with all faces in a cluster assigned to a single user's faces. We filter the clusters based on facial pose to avoid mixing faces with widely different poses into the same cluster. The largest cluster is selected as the representative user of the album, and the rest are discarded.

The face regions are then segmented out from these images using the Grabcut algorithm [12]. We bootstrap the Grabcut algorithm with the bounding box of the face landmarks, followed by iterative segmentation of the face region. All faces are then coarsely aligned to a single randomly chosen face by warping them via affine transforms, using landmarks as correspondences. Since warping involves linear interpolation from four neighboring pixels, we need to adjust the noise variance of the aligned image as the weighted sum of four i.i.d. variables, *i.e.* $\sigma_i^2 \leftarrow \sum_{k=1}^{4} w_k^2 \sigma_i^2 = \sigma_i^2/4$, assuming equal weights $w_k = 1/4$.

4. Results

We have conducted quantitative as well as qualitative experiments to evaluate our work. Our overall dataset consists of 27 albums, where each album contains between 3 and 179 photos. We show an example album from our dataset in figure 5, as well the estimated noise level and denoised results.

True Sigma	Estimated Sigma			PSNR		
	Liu et. al.	Q-metric	Our Method	Q-metric	Our Method	Best BM3D
3.97	20.32	15	*4.97*	35.83	*38.90*	39.40
7.79	21.02	19	*8.69*	35.45	*36.58*	36.95
11.47	22.69	27	*12.01*	35.04	*35.71*	35.67
15.01	24.57	35	*15.67*	34.68	*35.03*	34.97
18.46	26.69	31	*18.87*	34.35	*34.40*	34.44
3.97	16.46	7	*5.66*	38.44	*38.90*	39.36
7.86	17.78	11	*9.09*	37.05	*37.32*	37.29
11.69	19.3	15	*13.15*	36.13	*36.22*	36.13
15.48	21.65	19	*16.77*	35.37	*35.41*	36.37
19.25	24.45	27	*20.31*	34.53	*34.65*	34.64
3.99	20.54	11	*10.63*	*37.31*	37.21	39.47
7.86	21.42	15	*12.7*	35.74	*36.11*	36.93
11.64	22x.75	19	*14.84*	34.55	*35.08*	35.24
15.35	24.66	23	*17.83*	33.56	*34.00*	34.02
18.97	26.61	27	*20.93*	32.79	*33.12*	33.10
3.82	16.26	15	*4.02*	38.20	*41.83*	40.24
7.35	17.93	19	*7.79*	37.43	*38.90*	38.46
10.78	19.07	23	*10.68*	36.61	*37.08*	37.20
14.14	21.09	31	*13.47*	*35.58*	35.35	36.06
17.44	22.86	35	*17.32*	*34.63*	34.37	34.90
3.95	20.81	11	*7.43*	39.43	*40.28*	41.10
7.83	21.85	19	*11.18*	37.60	*38.62*	38.78
11.66	22.99	27	*13.5*	36.57	*37.46*	37.42
15.39	25	31	*16.6*	35.63	*36.35*	36.33
19.04	26.52	31	*19.26*	35.06	*35.22*	35.33
4	19.4	11	*7.56*	37.74	*39.09*	39.89
7.96	20.39	15	*9.81*	36.39	*37.60*	37.55
11.86	22.58	15	*12.64*	36.04	*36.26*	36.04
15.67	23.99	19	*15.89*	35.02	*35.15*	35.02
19.41	26.28	23	*19.79*	34.16	*34.19*	34.16

Table 1: Column 2-4: Estimated noise σ from various methods. Each group of rows corresponds to the same original image. Columns 5-7: PSNR for Q-metric, our method, and the best achievable via BM3D. The best numbers are in bold italics. The best BM3D results use discrete sampling of parameter space, which is why we sometimes do better.
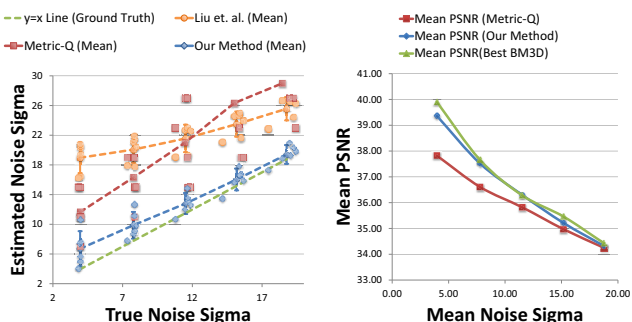

Figure 6: (a) Estimated noise level *vs*. ground truth noise level. The means are computed per true σ value. We don't show error bars for Q-metric since it selects a σ from a discrete set, as opposed to truly estimating it. Our estimates are closest to the ground truth line. (b) PSNR of the outputs denoised by BM3D using our σ *vs*. Q-metric. Our PSNR is close to the best possible PSNRs using BM3D.

4.1. Quantitative ground truth experiments

To evaluate the accuracy of our noise estimation technique, we selected 6 albums from our dataset, and further selected a single test image from each of those albums. We added synthetic white noise with known variance to each test image, and then estimated the noise in the modified image using our algorithm. We compare our results against noise estimates from Liu *et al.*'s [8] single-image method,
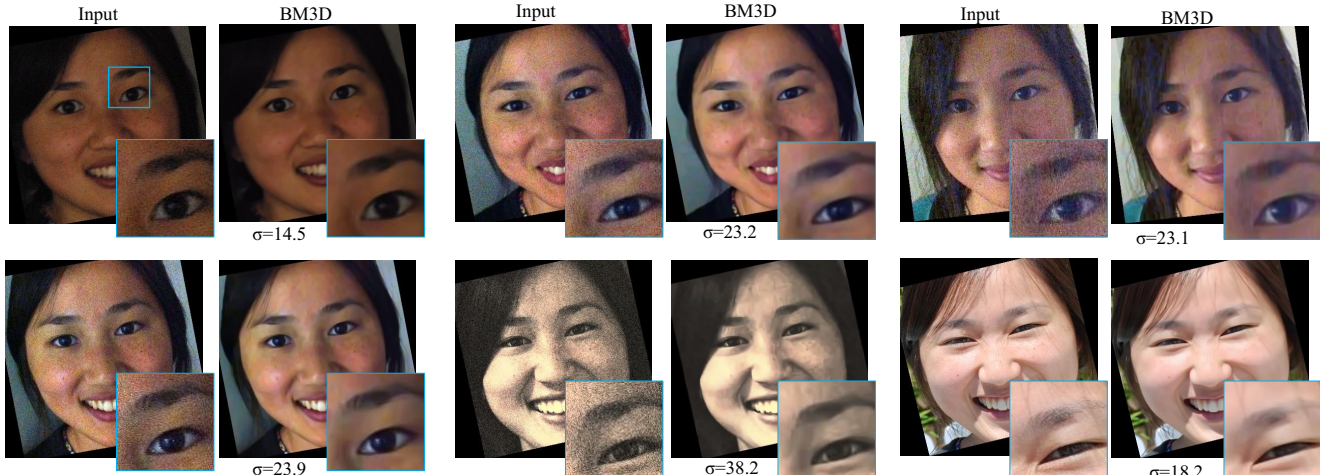
Figure 5: In each of the 6 examples from the same subject, the left image is input; the right is BM3D denoised using the estimated noise level shown.

and the BM3D noise parameters selected by Q-metric [15]. For Liu's method, we used 30 clusters for their piece-wise linear model.

We report the estimated noise σ (std. dev.) in Table 1 and plot them in Fig. 6 as a function of the true noise σ. We used $\sigma = 4, 8, 12, 16, 20$ to simulate the noise, but report the true noise as the actual measured σ of the difference between simulated and original images. The test images along with some denoised results are shown in Fig. 9. We consistently perform better than both Liu's method as well as Q-metric in terms of accuracy of the estimated σ.

Since image denoising can directly benefit from our method, we also evaluate denoising results by their signal-to-noise ratio (PSNR) relative to the ground truth image. We chose BM3D, which is the state-of-art image denoising algorithm. We compare against Q-metric as an alternative noise parameter selection method, as well as against the best PSNR achieved by BM3D, which serves as an upper bound on the PSNR. We clearly beat Q-metric and are very close to the upper bound (see Fig. 6). The quality of noise estimation improves with increasing noise, since at smaller noise values, it is harder to distinguish between noise and other image variations. On the other hand, there are more gains in terms of PSNR at lower noise values, due to reduced smoothing.

Note that Q-metric requires running BM3D multiple times at different σ values. Our method on the other hand, runs BM3D only once based on the estimated σ, which is a performance advantage in addition to the better accuracy that we achieve. Even though our estimation procedure has an overhead, the computation is amortized over all images and can be alleviated by subsampling patches.

4.2. Qualitative user studies

To understand the perceptual improvement contributed by our method, we conducted a user study to compare BM3D results using our noise estimates *vs.* Q-metric based selection. Our dataset consisted of 71 images. We presented 3 users with task of selecting the more preferable result between ours and Q-metric, shown in randomized order. They were also able to see the original noisy image. We selected the majority choice as the preferred method. Table 2 shows the outcome of the study: our method is preferred for 41% of images as opposed to 24% for Q-metric. For 76% of the images, our method were considered better than or equal to Q-metric. Fig. 7 shows some denoising results using our estimation method with BM3D, while Fig. 8 shows some comparisons with Q-metric.

Q-metric	Ours	Equally good
24%	**41%**	35%

Table 2: User study result shows our method is preferred by majority. See section 4.2 for details.

Runtime Relative noise estimation for a single image pair takes 1 sec. (C++, 3.2 GHz, 32GB RAM). For N photos, this is done $O(N^2)$ times but is parallelizable. Solving Eqs. 20 - 22 is relatively negligible (10-100ms) using Octave.

5. Conclusion, Limitations and Future work

We have proposed a novel approach for jointly estimating noise level for all images in a photo album by using faces as reference points. Our qualitative and quantitative experiments demonstrate the accuracy of our estimates and its effectiveness in controlling denoising parameters for BM3D. Our user studies show that we do better than the

Figure 7: Denoising results using BM3D with our noise estimates. Please zoom-in to see details.

state-of-the-art (Liu. *et al*., Q-metric) and *as well as* BM3D can do in terms of PSNR. Our method is particularly suitable for photo album services in social networks.

There are several limitations and avenues for future work. Acne, makeup, or aging might change the face appearance of the user. It will be interesting to combine single image noise estimation and our method to further improve the noise estimation accuracy, *e.g*. using the results from single image estimation in Eq. 19 as a soft constraint. During denoising, we show results on faces only, but once noise has been estimated, it could be used to denoise the whole image. Our work focuses on facial denoising, but the idea of jointly denoising a group of photos of the same entity could be applied in other settings, such as popular landmarks.

Acknowledgements

References

[1] http://en.wikipedia.org/wiki/jensen's_inequality.

[2] A. Buades, B. Coll, and J. Morel. A non-local algorithm for image denoising. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
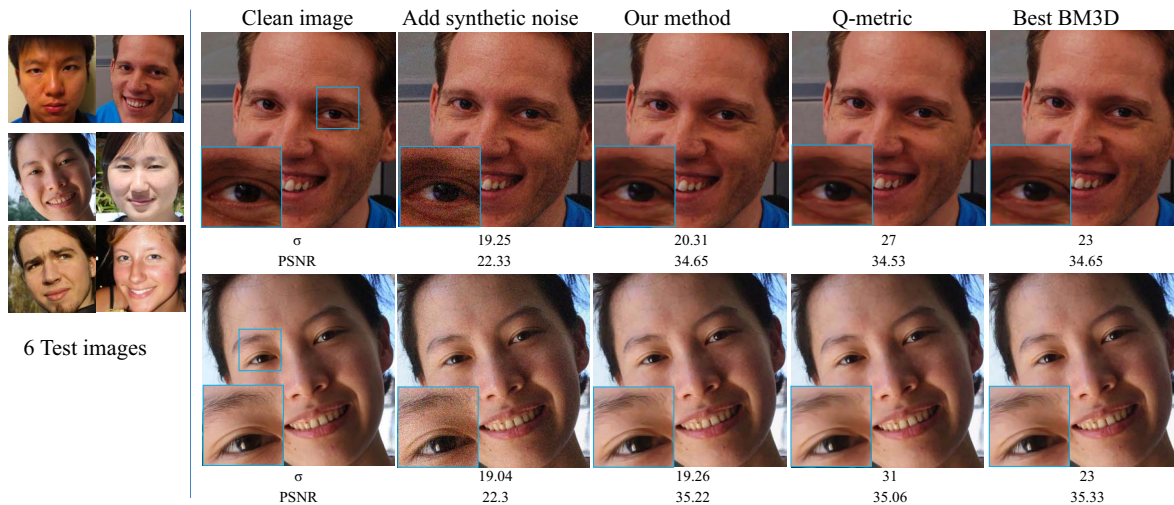
Figure 9: Test images used in our quantitative experiments, along with two examples of denoised results using BM3D with our estimates, Q-metric, and the best achievable by BM3D. Our results are close to the best possible.
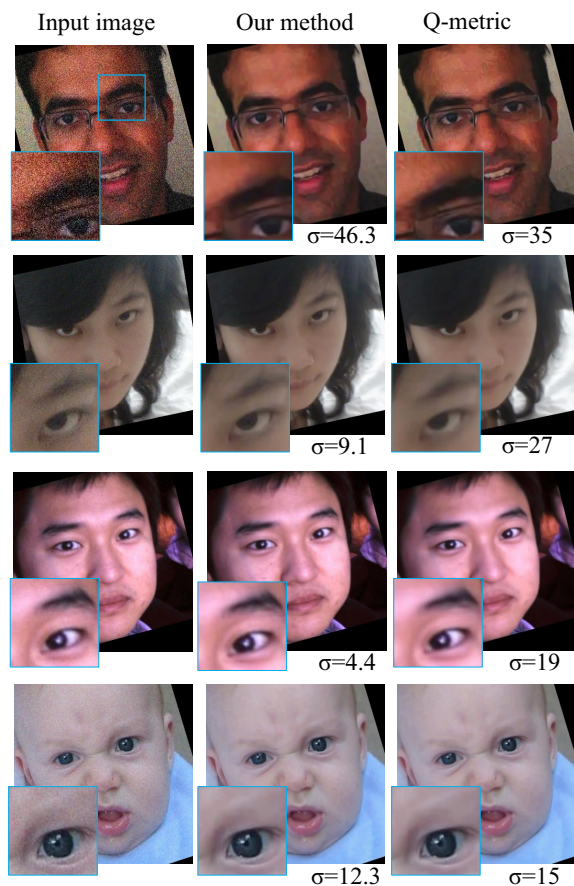

Figure 8: Comparisons between our results (BM3D using our σ) and Q-metric based BM3D. Our results are sharper and without residual noise.

[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8), 2007.

[4] S. Hasinoff, F. Durand, and W. Freeman. Noise-optimal capture for high dynamic range photography. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[5] N. Joshi, W. Matusik, E. Adelson, and D. Kriegman. Personal photo enhancement using example images. *ACM Transactions on Graphics*, 29(2), 2010.

[6] V. Kwatra, M. Han, and S. Dai. Shadow removal for aerial imagery by information theoretic intrinsic image analysis. In *International Conference on Computational Photography*, 2012.

[7] C. Liu, R. Szeliski, S. Kang, C. Zitnick, and W. Freeman. Automatic estimation and removal of noise from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2), 2008.

[8] Z. Liu, Z. Zhang, and Y. Shan. Image-based surface detail transfer. *IEEE Computer Graphics and Applications*, 24(3), 2004.

[9] J. Portilla. Full blind denoising through noise covariance estimation using gaussian scale mixtures in the wavelet domain. In *International Conference on Image Processing*, volume 2, 2004.

[10] K. Rank, M. Lendl, and R. Unbehauen. Estimation of image noise variance. In *IEE Proceedings on Vision, Image and Signal Processing*, volume 146, pages 80–84, 1999.

[11] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley. Color transfer between images. *IEEE Computer Graphics and Applications*, 21(5), 2001.

[12] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM Transactions on Graphics*, volume 23, 2004.

[13] R. Shah and V. Kwatra. All smiles: automatic photo enhancement by facial expression analysis. In *European Conference on Visual Media Production*, 2012.

[14] H. Takeda, S. Farsiu, and P. Milanfar. Kernel regression for image processing and reconstruction. *IEEE Transactions on Image Processing*, 16(2), 2007.

[15] X. Zhu and P. Milanfar. Automatic parameter selection for denoising algorithms using a no-reference measure of image content. *IEEE Transactions on Image Processing*, 19(12), 2010.

[16] D. Zoran and Y. Weiss. Scale invariance and noise in natural images. In *IEEE International Conference on Computer Vision*, 2009.