

Estimating a Small Signal in the Presence of Large Noise

Amy Zhao

xamyzhao@mit.edu

Frédéric Durand

fredod@mit.edu

John Guttag

gutttag@mit.edu

MIT Computer Science and Artificial Intelligence Laboratory

Abstract

Video magnification techniques are useful for visualizing small changes in videos. For instance, Eulerian video magnification has been used to visualize the flow of blood in the human face. Such visualizations have possible applications in remote monitoring or screening for diseases. However, when visualizing blood flow, the signal of interest may be similar in amplitude to the noise in the video. This raises the question of what one is actually seeing in a magnified video: signal or noise? We seek to understand these signal and noise characteristics with the goal of producing informative and accurate visualizations. We present a preliminary algorithm for estimating the signal amplitude in the presence of relatively high noise. We demonstrate that the algorithm can be used to accurately estimate the signal amplitude in an uncompressed simulated video, but is susceptible to compression noise and motion.

1. Introduction

Video magnification can be used to visualize minute changes in videos that would normally not be visible. Studies have indicated that there is a signal associated with blood flow in the skin; this signal is undetectable by the naked eye, but can be measured using video processing techniques and visualized using video magnification. Signals relating to blood flow (e.g., heart rate) have been obtained from videos of human faces by applying temporal and spatial filtering [25] or blind source separation [18] to the intensity signal, or by tracking the motion of the head [4]. Techniques for examining these signals have potential applications in remote monitoring or screening for diseases relating to blood flow.

Wu *et al.* showed that Eulerian linear video magnification can be used to visualize the blood flow signal in human faces [28]. In the context of visualizing blood flow, the signal of interest (that is, the intensity changes reflecting the flow of blood) can have an amplitude that is similar

to or lower than the noise level in the video. It is important to understand the signal and noise characteristics in order to produce accurate and informative visualizations. In this work, we discuss how the signal is affected by the noise that is introduced by the acquisition and the compression processes. We demonstrate a preliminary algorithm for estimating the amplitude of a small signal in the presence of relatively large noise.

2. Related work

Noise Modern images and image sequences are typically affected by noise introduced by the acquisition process, the compression process, or other factors such as motion [15]. CCD noise models usually assume that the acquisition noise level is dependent on the uncorrupted pixel intensity, but is independent between pixels. The relationship between the noise level and mean pixel intensity is described by a noise level function (NLF), which is dependent on the natural behavior of photons, the properties of the camera and some recording parameters [14, 12].

Compression noise can introduce additional artifacts to an image or video. Many modern image and video compression standards apply a transform (commonly the discrete cosine transform, or DCT) in a block-wise fashion to the image and then quantize the transform coefficients [26, 27]. Images and videos compressed in this fashion typically exhibit “blocking” artifacts.

Noise and signal estimation There is a large body of literature regarding noise estimation and denoising techniques for images and videos. The NLF of a CCD camera is easy to estimate from a video or multiple images of a static scene [8, 2, 12]. In the image processing literature, the signal of interest is typically defined as the uncorrupted image; denoising aims to recover this image from a noisy image. A large class of image denoising techniques rely on the concept of averaging to reduce noise levels, e.g., through Gaussian smoothing or anisotropic filtering [17]. There are also deblocking techniques that remove visible

block-based compression artifacts by detecting and smoothing block boundaries [23, 22, 3].

Several key differences make it difficult to directly apply any of these noise estimation and denoising techniques to video magnification. Firstly, most existing denoising algorithms focus on applications with relatively high input SNR's, and may not be accurate enough to estimate a low amplitude signal. Typically, video denoising algorithms are evaluated on image sequences with peak signal-to-noise ratio (PSNR) values in the 10-30 dB range, and obtain improvements in PSNR on the order of 0-10 dB [6, 19]. In contrast, in order to visualize blood flow in the human face using Eulerian video magnification, a magnification factor of 100 has been reported [28]; this suggests that the amplitude of the color variation in the input video is extremely low relative to the image intensity, and that the SNR is frequently much lower than 10 dB.

Secondly, video magnification is affected by noise that may be hard to model and/or correlated with the signal. For instance, when visualizing blood flow, the color change signal is commonly corrupted by a motion signal with a similar temporal signature. Many state-of-the art video denoising techniques are capable of removing the types of noise commonly introduced by imaging devices (e.g., additive, multiplicative or structural noise [5, 13]), but may not be applicable to more complex noise models.

3. Signal Estimation for Video Magnification

We discuss the signal and noise in the context of visualizing blood flow using the Eulerian linear video magnification technique described in [28]. Specifically, we focus on the problem of estimating the amplitude of a color change signal in a video that has been corrupted by acquisition noise of a similar amplitude, as well as compression noise introduced by the Motion JPEG (MJPEG) compression process. The MJPEG scheme is less commonly used than other video compression schemes such as H.264/AVC or MPEG-4. However, we choose to study it because it compresses each frame independently according to the JPEG compression scheme, and thus does not introduce any inter-frame correlations.

In [25], the green channel was observed to have the strongest signal pertaining to blood flow. We focus on the Y or luminance channel, as it contains much of the green channel information, and is not subjected to the subsampling process that is applied to the chroma channels during MJPEG encoding [26].

3.1. Assumptions

We assume that the input video depicts a stationary scene containing a small color change signal. We assume that the signal: (1) is smooth and varies slowly in space, so that

we can apply a small spatial averaging filter without significantly altering the signal characteristics; and (2) has a small enough amplitude as to not affect the intensity-dependent noise level. Earlier, we discussed the difficulty of modeling temporally correlated noise terms such as motion. We assume for now that the subject of the video has been mechanically stabilized such that the effect of motion is negligible.

In the image and video denoising literatures, the acquisition noise is commonly assumed to be additive zero-mean Gaussian noise with a variance that is proportional to the luminance of the scene [24]. We assume in our model that each video frame is affected by such acquisition noise, as well as by JPEG compression.

We further assume that the video scene is smooth, so that pixels within small neighborhoods exhibit similar noise levels. We can enforce this assumption in real videos by excluding parts of the scene that contain high gradients.

3.2. Signal and Noise Model

We model an uncompressed video frame as follows:

$$I = I_0 + \phi + n_{acq}, \quad (1)$$

where I_0 denotes the mean observed image intensity, ϕ the signal of interest and n_{acq} the zero-mean Gaussian noise that is introduced during the acquisition process. These terms are assumed to be independent. Similarly, we model a compressed video frame as:

$$I_c = \mathcal{Q}\{I_0 + \phi + n_{acq}\}, \quad (2)$$

where $\mathcal{Q}\{\cdot\}$ represents the quantization operator that is applied during the JPEG compression process [26].

In JPEG encoding, the image is first divided into non-overlapping 8×8 blocks. The discrete cosine transform (DCT) of each block is computed according to the formula:

$$H(k, l) = \frac{1}{4} C(k) C(l) \sum_{x=0}^7 \sum_{y=0}^7 I(x, y) \cdot \cos\left(\frac{(2x+1)k\pi}{16}\right) \cos\left(\frac{(2y+1)l\pi}{16}\right),$$

$$\text{where } C(k), C(l) = \frac{1}{\sqrt{2}} \text{ for } k, l = 0,$$

$$C(k), C(l) = 1 \text{ otherwise.}$$

$H(k, l)$ describes the DCT coefficient value at location k, l . Each DCT coefficient is then quantized using an application-defined quantization table Q as follows:

$$H_Q(k, l) = Q(k, l) \cdot \text{round}\left(\frac{H(k, l)}{Q(k, l)}\right).$$

The quantization operation is lossy and may be modeled differently depending on the image content. DCT coefficients that are smaller than the corresponding quantization step are quantized to zero and result in a reduction in image energy, while larger DCT coefficients contribute noise that may be modeled as additive random noise [9, 21].

We assumed earlier that the video scene and the signal are smooth in space. Under the additional assumption that the noise level is small relative to the quantization step size, each frame contains little energy in the DCT components corresponding to high spatial frequencies. In the JPEG standard, the luminance quantization table applies more aggressive quantization to high-frequency DCT components [1]. This means that high-frequency DCT components are typically quantized to 0, while the low frequency components contribute quantization noise. We observed this effect in simulated videos compressed using moderate compression levels. We model this as follows:

$$I = \mathcal{Q}_L\{I_0 + \phi + n_{acq}\} + n_Q,$$

where $\mathcal{Q}_L\{\cdot\}$ represents the spatial lowpass filtering effect of applying quantization to small DCT coefficients, and n_Q represents the quantization noise that is added. Under the assumption that the mean image I_0 and the signal ϕ contain little energy in the high-frequency DCT components, we may rewrite this as:

$$I = I_0 + \phi + n_Q + \mathcal{Q}_L\{n_{acq}\}. \quad (3)$$

3.3. Signal Estimation Algorithm

Local spatial averaging is used in many image denoising algorithms to reduce noise levels [7]. We present an algorithm that examines the effect of spatial averaging on the acquisition noise variance in a video, and extrapolates this relationship to estimate the theoretical variance of the video when the acquisition noise variance has been reduced to zero; the variance of the video should then be comprised of the signal variance.

Let us first consider the case of an uncompressed video. Under the assumption that the signal varies slowly in space, we may apply a spatial averaging filter to reduce the standard deviation of Gaussian noise without significantly affecting the signal. An averaging filter (or, box filter) M of size $m \times m$ attenuates the variance of zero-mean Gaussian noise by a factor of $\frac{1}{m^2}$ [16]. From Eq. 2, we may write the variance over time of the filtered frame I_m as:

$$\begin{aligned} \text{Var}[I_m]_t &= \text{Var}[M * (I_0 + \phi + n_{acq})]_t, \text{ or} \\ \sigma_{I_m}^2 &= \sigma_\phi^2 + \frac{\sigma_{n_{acq}}^2}{m^2}. \end{aligned} \quad (4)$$

This relationship is valid under the assumption that each of the terms I_0 , ϕ , n_{acq} are independent, and that n_{acq} is spatially uncorrelated.

In our algorithm, we choose several box filter sizes and measure the variance $\sigma_{I_{m_i}}^2$ for each m_i . We then extrapolate a linear fit of $\sigma_{I_m}^2$ against $\frac{1}{m^2}$ to the point at which $\frac{\sigma_{n_{acq}}^2}{m^2} = 0$ and $\sigma_{I_m}^2 = \sigma_\phi^2$; the intercept of the linear fit is an estimate of the signal variance.

The compression process introduces complications that require some adjustments to the above algorithm. As we discussed earlier, $\mathcal{Q}_L\{\cdot\}$ sets the high-frequency DCT coefficients of n_{acq} to 0. Applying small spatial filters have little effect on the variance of this compressed noise term since the high spatial frequency energy that would have been removed by the averaging filter is already zero; we confirmed this effect in simulations. Using a larger filter size affects the energy in the remaining spatial frequencies. In simulations, we found that for $m \geq 8$,

$$\text{Var}[M * (\phi + n_Q + \mathcal{Q}_L\{n_{acq}\})] \approx \text{Var}[M * (\phi + n_Q + n_{acq})].$$

The resulting variance over time of the filtered image, $\sigma_{I_{c,m}}^2$, is approximated by:

$$\sigma_{I_{c,m}}^2 = \sigma_\phi^2 + \frac{\sigma_{n_Q}^2 + \sigma_{n_{acq}}^2}{m^2}. \quad (5)$$

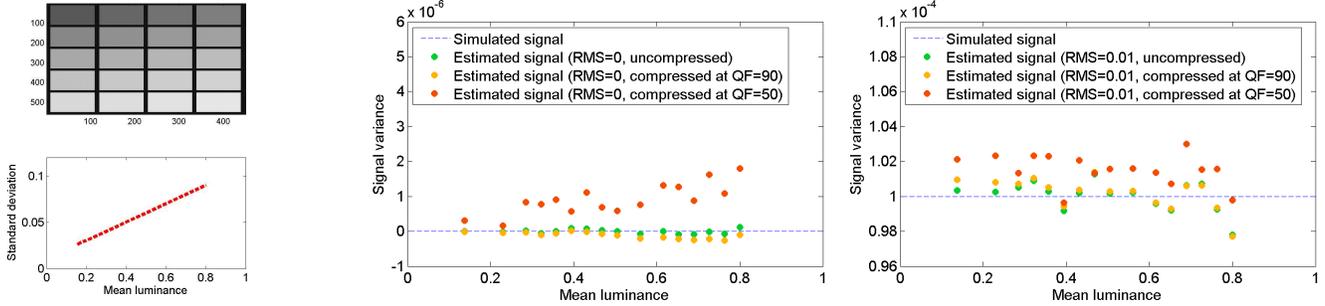
This model assumes that the effect of $\mathcal{Q}_L\{\cdot\}$ can be ignored for $m \geq 8$, that n_{acq} and n_Q are spatially uncorrelated, and that ϕ , n_{acq} and n_Q are independent. We discuss these limitations later.

Similarly to the uncompressed case, we can apply several filters of varying sizes, and estimate a linear fit of $\sigma_{I_{c,m}}^2$ against $\frac{1}{m^2}$. The intercept of the fit can be used as an estimate for σ_ϕ^2 .

4. Experimental Results

We demonstrate that we can use our algorithm to estimate the variance of a simple sinusoidal signal in the presence of acquisition noise. We first simulated two videos of a stationary color calibration target. One video contains no signal (Fig. 1b); the other has a small sinusoidal signal added to each pixel (Fig. 1c). We added zero-mean Gaussian acquisition noise with a linear NLF to both videos (Fig. 1a). The videos were then compressed using various levels of compression. We are able to estimate the signal variance fairly accurately from the uncompressed videos, as well as the videos compressed with a high quality factor (QF; a higher value indicates less aggressive compression). Our predicted signal suffers from a large degree of error when a lower compression quality setting is used. Fig. 1c indicates that even for a high compression quality video, our estimate becomes inaccurate for pixels with high noise levels.

We applied our signal estimation algorithm to several real videos captured with a Panasonic Lumix DMC-G2 in



(a) Simulated video frame (top) and simulated NLF (bottom). (b) Mean (over space) estimated signal variance per mean (over time) luminance, from simulated videos containing no signal. (c) Mean (over space) estimated signal variance per mean (over time) luminance, from simulated videos containing a spatially invariant, temporally varying sine wave with $RMS = 0.01$.

Figure 1: Estimated signal variance from a simulated color calibration target. To assess the effects of compression, we saved the videos using MJPEG encoding and specified the video quality factor (QF).

MJPEG format. The frames were converted to linear RGB space and then to YCbCr space. Fig. 2 shows our results for a stationary color calibration target, which should contain no signal. Our signal estimates in Fig. 2b contain a similar error to the compression noise observed in our simulation results (although the error in our simulations appears more discontinuous, likely due to the more discretized pixel intensity values in the simulation). We use the maximum of this error to conservatively estimate the noise floor of our signal estimates in the subsequent videos. We took videos of a human hand and human feet under the same lighting conditions as in Fig. 2. We used a noise floor estimate of 1×10^{-5} . Fig. 3 show that the regions of the hand supplied most directly by arterial flow [11] exhibit a signal with an amplitude that is above the noise floor. The feet (Fig. 4) appear to exhibit a strong signal everywhere; however, when processing this video using Eulerian linear video magnification, it was difficult to visualize a signal at the subject’s heart rate. It is likely that this high signal estimate is actually caused by motion rather than blood flow in the feet, as some motion can be observed in the video. In both videos, the signal estimate appears to be correlated with the mean luminance (Figs. 3b and 4b); this indicates that our estimate may be affected by the motion of specularities in the videos.

5. Limitations

Compression noise Figs. 1 and 2 indicate that the accuracy of our signal estimates is affected by the degree of compression (and thus quantization) applied to the video. We roughly estimated the quantization noise level from a video of a static scene with no signal, using a simplistic model of the compression process (Equations 3 and 5). Our model does not account for the spatial correlations that are

introduced by the compression process, or the dependence of the quantization noise on the image content [21]. The effect of compression might be described more accurately using existing models. Quantization noise in the DCT domain has been modeled as uniformly distributed; it has also been modeled based on a Laplacian model for DCT coefficients [21]. Some authors have proposed the use of different distributions depending on the quantized image content [9, 21]. Further studies on the effect of quantization noise should also be conducted by capturing uncompressed real videos.

Motion Motion is often observed to be a confounding factor when attempting to visualize blood flow using video magnification. Because it often has a similar temporal signature to the blood flow signal, it cannot be removed from the video via simple temporal filtering. Figs. 3 and 4 imply that our signal estimate may have been corrupted by the motion of specularities in the video. It may be possible to physically stabilize the video subject using mechanical restraints. Alternatively, future work can examine the effect of motion on our model and our estimates. For instance, it may be possible to use motion estimation techniques such as optical flow to estimate the amount of variance in each pixel that is caused by motion.

Camera processes There are camera processes other than compression that may affect the accuracy of our model. For example, many demosaicing strategies perform spatial interpolation [20] that introduces spatial correlations. Modeling these correlations requires the consideration of covariance terms in Equations 4 and 5. The quantization process that occurs when the image is exported from the camera

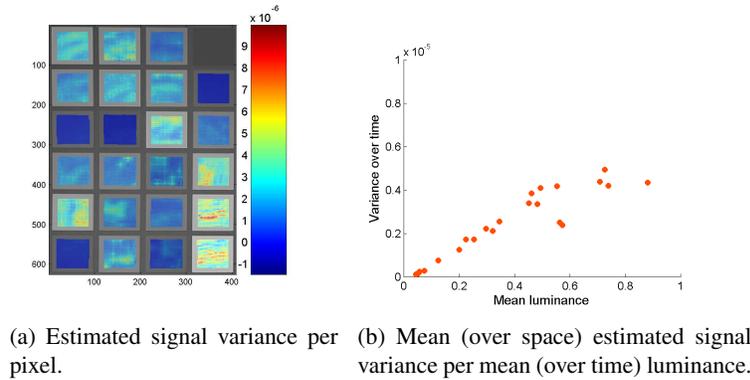


Figure 2: Estimated signal variance for a stationary color calibration target. The upper right region of the card was excluded from the visualization as it appeared to have extremely high noise levels that interfere with the visualization scale.

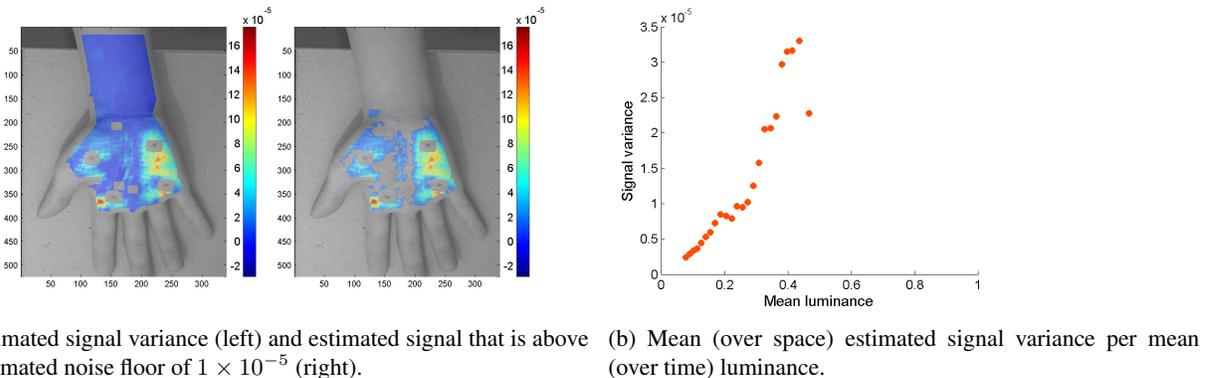


Figure 3: A healthy female human hand. Video was taken under similar conditions to Fig. 2.

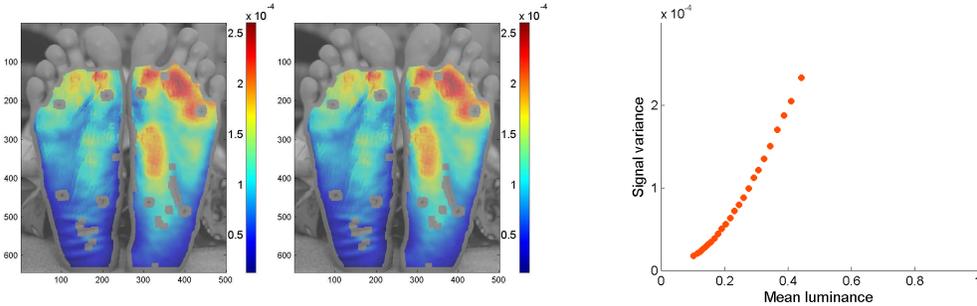
(typically to 8-16 bits per channel) introduces additional quantization noise [10]. The accuracy of our model and camera can likely be improved by accounting for these effects.

6. Conclusion

Video magnification can be used to produce informative visualizations of human blood flow from consumer-grade videos. In these videos, the signal of interest may be very small compared to the noise level. We presented a model and an algorithm for estimating the signal in this context. We showed that our algorithm can be used to estimate the signal variance in a simulated video that is corrupted by zero-mean Gaussian noise and compressed at a high quality factor. However, we also demonstrated that our estimates are highly affected by aggressive compression as well as motion. Further work is required to assess the performance of our algorithm in the presence of compression and motion.

References

- [1] Digital compression and coding of continuous-tone still images, part 1, requirements and guidelines. *ISO/IEC JTC1 Draft International Standard 10918-1*, 1991.
- [2] A. Amer and E. Dubois. Fast and reliable structure-oriented video noise estimation. *Circuits and Systems for Video Technology, IEEE Transactions on*, 15(1):113–118, 2005.
- [3] A. Z. Averbuch, A. Schclar, and D. L. Donoho. Deblocking of block-transform compressed images using weighted sums of symmetrically aligned pixels. *Image Processing, IEEE Transactions on*, 14(2):200–212, 2005.
- [4] G. Balakrishnan, F. Durand, and J. Guttag. Detecting pulse from head motions in video. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3430–3437. IEEE, 2013.
- [5] J. E. Bioucas-Dias and M. A. T. Figueiredo. Multiplicative noise removal using variable splitting and constrained optimization. *Image Processing, IEEE Transactions on*, 19(7):1720–1730, 2010.
- [6] J. C. Brailean, R. P. Kleihorst, S. Efstratiadis, A. K. Katsaggelos, and R. L. Lagendijk. Noise reduction filters for dy-



(a) Estimated signal variance (left) and estimated signal that is above the estimated noise floor of 1×10^{-5} (right). (b) Mean (over space) estimated signal variance per mean (over time) luminance.

Figure 4: A healthy pair of male human feet. Video was taken under similar conditions to Fig. 2.

- dynamic image sequences: a review. *Proceedings of the IEEE*, 83(9):1272–1292, 1995.
- [7] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. 2:60–65, 2005.
- [8] G. E. Healey and R. Kondepudy. Radiometric ccd camera calibration and noise estimation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(3):267–276, 1994.
- [9] A. Ichigaya, M. Kurozumi, N. Hara, Y. Nishida, and E. Nakasu. A method of estimating coding psnr using quantized dct coefficients. *Circuits and Systems for Video Technology, IEEE Transactions on*, 16(2):251–259, 2006.
- [10] K. Irie, A. E. McKinnon, K. Unsworth, and I. M. Woodhead. A technique for evaluation of ccd video-camera noise. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(2):280–284, 2008.
- [11] D. B. Jenkins. *Hollinshead’s functional anatomy of the limbs and back*. Elsevier Health Sciences, 2008.
- [12] M. Kobayashi, T. Okabe, and Y. Sato. Detecting forgery from static-scene video based on inconsistency in noise level functions. *Information Forensics and Security, IEEE Transactions on*, 5(4):883–892, 2010.
- [13] C. Liu and W. T. Freeman. A high-quality video denoising algorithm based on reliable motion estimation. In *Computer Vision–ECCV 2010*, pages 706–719. Springer, 2010.
- [14] C. Liu, W. T. Freeman, R. Szeliski, and S. B. Kang. Noise estimation from a single image. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 901–908. IEEE, 2006.
- [15] M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian. Video denoising, deblocking, and enhancement through separable 4-d nonlocal spatiotemporal transforms. *Image Processing, IEEE Transactions on*, 21(9):3952–3966, 2012.
- [16] P. M. Narendra. A separable median filter for image noise smoothing. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (1):20–29, 1981.
- [17] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(7):629–639, 1990.
- [18] M.-Z. Poh, D. J. McDuff, and R. W. Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics express*, 18(10):10762–10774, 2010.
- [19] M. Protter and M. Elad. Image sequence denoising via sparse and redundant representations. *Image Processing, IEEE Transactions on*, 18(1):27–35, 2009.
- [20] R. Ramanath, W. E. Snyder, G. L. Bilbro, and W. A. Sander. Demosaicking methods for bayer color arrays. *Journal of Electronic imaging*, 11(3):306–315, 2002.
- [21] M. A. Robertson and R. L. Stevenson. Dct quantization noise in compressed images. *Circuits and Systems for Video Technology, IEEE Transactions on*, 15(1):27–38, 2005.
- [22] S.-Y. Shih, C.-R. Chang, and Y.-L. Lin. A near optimal deblocking filter for h. 264 advanced video coding. In *Design Automation, 2006. Asia and South Pacific Conference on*, pages 6–pp. IEEE, 2006.
- [23] S.-C. Tai, Y.-Y. Chen, and S.-F. Sheu. Deblocking filter for low bit rate mpeg-4 video. *Circuits and Systems for Video Technology, IEEE Transactions on*, 15(6):733–741, 2005.
- [24] Y. Tsin, V. Ramesh, and T. Kanade. Statistical calibration of ccd imaging process. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 480–487. IEEE, 2001.
- [25] W. Verkrusse, L. O. Svaasand, and J. S. Nelson. Remote plethysmographic imaging using ambient light. *Optics express*, 16(26):21434–21445, 2008.
- [26] G. K. Wallace. The jpeg still picture compression standard. *Communications of the ACM*, 34(4):30–44, 1991.
- [27] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra. Overview of the h. 264/avc video coding standard. *Circuits and Systems for Video Technology, IEEE Transactions on*, 13(7):560–576, 2003.
- [28] H.-Y. Wu, M. Rubinstein, E. Shih, J. V. Guttag, F. Durand, and W. T. Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph.*, 31(4):65, 2012.