Image Motion Deblurring Using Minimum Entropy

Farhang Yeganegi
Department of Electrical Engineering
Amirkabir University of Technology
Email:
farhang.yeganegi@gmail.com

Hamidreza Amindavar
Department of Electrical Engineering
Amirkabir University of Technology
Email:
hamidami@aut.ac.ir

Abstract—

I. INTRODUCTION

II. IMAGE DENOISING USING SINGULAR VALUE DECOMPOSITION

One of the matrix decomposition method is based on singular value decomposition (SVD). In this approach, the matrix $A_{m\times n}$ with the rank p can be decomposed to $A = USV^T$ where $U_{m\times m}$ and $V_{n\times n}$ are orthogonal matrices. The columns of U contain the eigenvector of AA^T and the columns of V contain the eigenvector of A^TA. AA^T and A^TA are symmetric and positive semi-definite matrices and have positive or zero eigenvalues. $S_{m\times n}$ is a diagonal matrix whose diameter values are the singular values of A.

$$S_{m\times n} = diag(\sigma_1, ..., \sigma_\mu)$$
$$\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_p > 0 \tag{1}$$
$$\mu = min\{m, n\}$$
$$\sigma_{p+1} = \sigma_{p+2} = ... = \sigma_\mu = 0$$

where σ_1 and σ_p are the largest and smallest nonzero singular values of the matrix A, respectively. Suppose that ξ is i.i.d. Gaussian random variables with zero mean and standard deviation σ denoted by $N(0, \sigma^2)$.

$$\xi = [\xi_1 \quad \xi_2 \quad ... \quad \xi_n] \tag{2}$$

Each ξ_i $(i = 1, 2, ..., n)$, is a $m \times 1$ random variable. If A is corrupted by the noise process: $A' = A + \xi$ and ξ is independent of A, then for $n \longrightarrow \infty$, by the i.i.d. property, we have

$$E(\xi\xi^T/n) = \sigma^2 I_m \tag{3}$$

and due to the independence of ξ and A, we can obtain

$$E(A'A'^T/n) = E(AA^T/n) + \sigma^2 I_m \tag{4}$$

where $E(.)$ means the expectation of a random matrix and I_m is the unit matrix of order m. The SVD of the observation matrix (A') can be written like below

$$A' = U'S'V'^T \tag{5}$$

where $U'_{m\times m}$ and $V'_{n\times n}$ are orthogonal matrices. The columns of U' are composed of eigenvectors of $A'A'^T$ and the columns of V' are composed of eigenvectors of A'^TA'. If $\{\lambda_1, \lambda_2, ..., \lambda_m\}$ are the eigenvalues of AA^T, then using the equation (4), the eigenvalues of matrix $A'A'^T$ will be $\{\lambda_1 + n\sigma^2, \lambda_2 + n\sigma^2, ..., \lambda_m + n\sigma^2\}$. So this brings us to the fact that the matrix U' of the observation matrix (noisy image) are influenced by the standard deviation of the noise directly. The same procedure can be carried out for the matrix V' and the similar result can be obtained. In another word, generally, the noise will expand the whole column space and singular values of the original image. For signals with the rank p, as indicated in equation (1), the sets of singular values $\{\sigma_{p+1}, \sigma_{p+2}, ..., \sigma_\mu\}$ are zero. However, as discussed earlier, when these signals are corrupted by the noise process, the mentioned singular values vary according to the standard deviation of the noise.

The central idea of the denoising purpose is to approximate a matrix derived from the data with another matrix of lower rank. Equation (5) can be written as $(n < m)$

$$A' = \sum_{j=1}^{n} \sigma'_j u'_j v'^T_j \tag{6}$$

where σ'_j, u'_j, and v'_j are singular values of the noisy image, $m \times 1$ column vectors of U', and $n \times 1$ column vectors of V', respectively. Let $k < n$, then

$$\hat{A}' = \sum_{j=1}^{k} \sigma'_j u'_j v'^T_j \tag{7}$$

is a denoised image. The essense of equation (7) is to retain the energy in signal subspace and discard the energy in the noise subspace. The most critical step in this approach is to determine the value k differentiating the signal subspace from the noise subspace.

III. MULTIRESOLUTION ANALYSIS AND IMAGE DENOISING USING WAVELET TRANSFORM

The idea of multiresolution analysis is to decompose the $L^2(R)$ space into the embedded subspaces with the following ingredients: ($f(t) \in L^2(R)$ is a continous-time function)

1- Such subspaces should have the following property

$$0 \subset ... \subset V_0 \subset V_1 \subset ... \subset V_j \subset ... \subset L^2(R) \tag{8}$$

2- A sequence of complementary subspaces, W_j, such that $V_j + W_j = V_{j+1}$ and $V_j \bigcap W_j = \{0\}$. This is written as

$$V_j \oplus W_j = V_{j+1} \tag{9}$$

3- A scaling (dilation) law:
If $f(t) \in V_j$ then $f(2t) \in V_{j+1}$

4- A shift (translation) law:
If $f(t) \in V_j$ then $f(t-k) \in V_j \quad k$ integer

5- V_0 and W_0 have a shift-invariant bases

$$V_0: \quad \{\phi(t-k): \quad |k| < \infty\} \tag{10}$$

$$W_0: \quad \{\psi(t-k): \quad |k| < \infty\} \tag{11}$$

We know that $V_0 \subset V_1$. So any function in V_0 can be written as a combination of the basic functions for V_1. In particular, since $\phi(t) \in V_0$ we can write

$$\phi(t) = 2\sum_{k=0}^{N} h_0(k)\phi(2t-k) \tag{12}$$

This is the Refinement equation. We also know that $W_0 = V_1 - V_0$. This means that any function in W_0 can also be written as a combination of the basic functions for V_1. Since $\psi(t) \in W_0$, we can write

$$\psi(t) = 2\sum_{k=0}^{N} h_1(k)\phi(2t-k) \tag{13}$$

This is the wavelet equation. h_0 and h_1 are lowpass and highpass filters, respectively. As a result, V_0 gives us the coarse approximation of the input function and W_j gives us the leve j of detail of the input function. Consequently, multiresolution representation can be written

$$L^2(R) = V_0 \oplus W_0 \oplus W_1 \oplus W_2 \oplus ... \tag{14}$$

Using equation (9), we can obtain that

$$V_j = V_0 \oplus W_1 \oplus W_2 \oplus ... \oplus W_{j-1} \tag{15}$$

This means that for an input function $f \in L^2(R)$ we have

$$f_j(t) = \sum_k a_{0k}\phi(t-k) + \sum_j \sum_k d_{jk}\psi_{jk}(t) \tag{16}$$

which describes the multiresolution representation up to level j.

The solution to the equation (12) may not always exist. The existence of the solution will depend on the discrete-time filter h_0. Using the cascade algorithm, if the iteration of the equation (12) converges, then the solution will be given by

$$\phi(t) = \lim_{i \to \infty} \phi^{(i)}(t) \tag{17}$$

where $\phi(t)$ is called father wavelet. The wavelet function ($\psi(t)$) will be obtained using equations (13) and (17) and is called mother wavelet. Equation (13) produces different wavelet families like Daubechies, Haar, Symlets, Coiflets, etc. Selecting specific $\phi(t)$ and $\psi(t)$ according to the assigned wavelet families and put them in equation (16) will result in wavelet decomposition. As a result, wavelet decomposition separates the input function into approximation and details subbands. Implementation of the wavelet decomposition is possible by using Mallat pyramid algorithm which is illustrated below

$$a_{j-1}(n) = \sqrt{2}\sum_k h_0(k-2n)a_j(k) \tag{18}$$

$$d_{j-1}(n) = \sqrt{2}\sum_k h_1(k-2n)a_j(k) \tag{19}$$

where the approximation and detail coefficients of the specific level can be obtained by the coefficients of the previous level. In another word, by applying one level wavelet decomposition, two subbands are created (approximation and detail subbands). Then using the Mallat pyramid algorithm, the coefficients of the higher level wavelet decomposition can be obtained. For the reconstruction procedure, again the Mallat pyramid algorithm is used which is indicated below

$$a_j(n) = \sqrt{2}\sum_k a_{j-1}(k)h_0(n-2k) + \sqrt{2}\sum_k d_{j-1}(k)h_1(n-2k) \tag{20}$$

where the coefficients of the previous levels can be obtained by the coefficients of the higher levels.

Consider the matrix $A_{m \times n}$ as an input image. The idea of the denoising purpose using wavelet transform is as follows: (one level denoising)

1- Apply one level wavelet decomposition on the input imge A.
2- The image is split into 4 subbands, namely the HH, HL, LH, and LL subbands. The HH subband gives the diagonal details of the image; the HL subband gives the horizontal features while the LH subband represents the vertical structure. The LL subband is the approximation subband consisting of low frequency components.
3- Selecting an assigned threshold for the denoising purpose. The choise of the threshold plays an important role in the denoising purpose. Care should be taken so as to preserve the edges of the denoised image.
4- Apply the selected threshold on the details subbands. In case of hard thresholding we just turn to zeros those coefficients which have absolute values less than the

threshold value. In case of soft thresholding we additionally pull down the absolute values of wavelet coefficients.

IV. THEORY

Consider the orthogonal wavelet transform. So, as a result, $\psi_{jk}(t)$ form an orthonormal basis for $L^2(R)$. Also, suppose that $f(t) \in L^2(R)$ is a $1D$ signal. Using multiresolution analysis, we have

$$f(t) = \sum_k a_{0k}\phi(t-k) + \sum_j \sum_k d_{jk}\psi_{jk}(t) \tag{21}$$

Using the orthonormal property, the coefficients d_{jk} will be obtained

$$d_{jk} = \int_{-\infty}^{\infty} f(t)\psi_{jk}(t)dt \tag{22}$$

In this paper, our desire is

$$d_{1k} = \int_{-\infty}^{\infty} f(t)\psi_{1k}(t)dt \tag{23}$$

Also, we know that:

$$||\psi_{jk}(t)||_2^2 = 1 \tag{24}$$

Now suppose that the input signal is corrupted by the noise process

$$\overline{f(t)} = f(t) + n(t) \tag{25}$$

where $\overline{f(t)}$ and $n(t)$ are noisy signal and i.i.d. Gaussian random variables with zero mean and standard deviation σ denoted by $N(0,\sigma^2)$, respectively. Using equation (23) and substituting $\overline{f(t)}$ with $f(t)$ we will obtain

$$\overline{d_{1k}} = d_{1k} + e_k \tag{26}$$

where $\overline{d_{1k}}$ are the coefficients of the detail subband of the noisy signal. e_k is equal to

$$e_k = \int_{-\infty}^{\infty} n(t)\psi_{1k}(t)dt \tag{27}$$

Now we should figure out the mean and the variance of the random process $\overline{d_{1k}}$. However, to achieve this goal, first the mean and the variance of the random process e_k should be obtained. The mean of e_k is zero because $E(n(t))$ is zero in the following equation

$$E\big(e_k\big) = \int_{-\infty}^{\infty} E\big(n(t)\big)\psi_{1k}(t)dt = 0 \tag{28}$$

Which states that the mean of the random process e_k does not depend on k. In order to obtain the variance of e_k, the following procudures are carried out

$$E\big(e_k^2\big) = E\Big(\int_{-\infty}^{\infty} n(t)\psi_{1k}(t)dt \int_{-\infty}^{\infty} n(\tau)\psi_{1k}(\tau)d\tau\Big) \tag{29}$$

$$E\big(e_k^2\big) = E\Big(\int_{-\infty}^{\infty}\int_{\infty}^{\infty} n(t)n(\tau)\psi_{1k}(t)\psi_{1k}(\tau)dtd\tau\Big) \tag{30}$$

$$E\big(e_k^2\big) = \int_{-\infty}^{\infty}\int_{\infty}^{\infty} E\big(n(t)n(\tau)\big)\psi_{1k}(t)\psi_{1k}(\tau)dtd\tau \tag{31}$$

$$E\big(e_k^2\big) = \int_{-\infty}^{\infty}\int_{\infty}^{\infty} \sigma^2\delta(t-\tau)\psi_{1k}(t)\psi_{1k}(\tau)dtd\tau \tag{32}$$

$$E\big(e_k^2\big) = \sigma^2 \int_{-\infty}^{\infty} \psi_{1k}^2(t)dt \tag{33}$$

Using equations (24) and (33), we will obtain that

$$E\big(e_k^2\big) = \sigma^2 \tag{34}$$

where illustrates the fact that the variance of the random process e_k does not depend on k. Using equations (26), (28), and (34) the mean and the variance of the random process $\overline{d_{1k}}$ will be obtained

$$E\big(\overline{d_{1k}}\big) = d_{1k} \tag{35}$$

$$\sigma^2_{\overline{d_{1k}}} = E\big(e_k^2\big) = \sigma^2 \tag{36}$$

The same result will be obtained if we use a discrete sequence $f(n)$ instead of $f(t)$. From now on, $f(n)$ is considered as an FIR input signal which is equal to zero out of the intervel $n = [0, N-1]$ and its noisy version is $\overline{f(n)}$. So $f(n)$ and $\overline{f(n)}$ are $1 \times N$ input vectors. For the denoising purpose, first the detail subband should be split into some equally size windows W_i where their properties are illustrated below $(i = 1, 2, \ldots, M)$

1- For $i \neq j$, we have

$$W_i \cap W_j = \phi \tag{37}$$

2- And also we should have

$$W_1 \cup W_2 \cup \cdots \cup W_M = \text{detail subband} \tag{38}$$

Suppose that the length of each window is l. Also we know that the length of the detail subband (for one level decomposition) is approximately equal to $\frac{N}{2}$. As a result we should have

$$\frac{N}{2} \approx Ml \tag{39}$$

Now the SVD is applied to each window as a next step for the denoising purpose. Each window has only one singular value, using the fact that each of them is a $1 \times l$ vector. The singular value of each window is denoted like below

$$\text{singular value}\{W_i\} = \alpha_i \tag{40}$$

Also, the result of applying SVD to each window is as follow

$$W_i = U_i S_i V_i^T \tag{41}$$

where U_i, S_i, and V_i are

$$U_i = [u_i]_{1\times 1} \tag{42}$$

$$S_i = [\alpha_i \quad 0 \quad \dots \quad 0]_{1\times l} \tag{43}$$

$$V_i = [v_1^i \quad \dots \quad v_l^i]_{l\times l} \tag{44}$$

u_i is equal to the eigenvector of $W_i W_i^T$. Each α_i is equal to the singular value of each window W_i. Also each columns of V_i are composed of the eigenvectors of $W_i^T W_i$. Equation (41) can also be written as

$$W_i = u_i \alpha_i (v_1^i)^T \tag{45}$$

Our decision rule for the denoising purpose is

$$median\{\alpha_i\} \lesseqgtr \beta\sigma \tag{46}$$

where β is a random number that should be optimized and σ is the standard deviation of the Gaussian additive noise. $median\{\alpha_i\}$ can be expressed by the following equation

$$median\{\alpha_i\} = \alpha_i \tag{47}$$

where used the fact that each window has only one singular value. Equation (46) separates the windows W_i into two groups which are illustrated below

Group 1: This group is composed of the windows that have the following property

$$\alpha_i > \beta\sigma \tag{48}$$

which illustrates the fact that window has edge information.

Group 2: This group is composed of the windows that have the following property

$$\alpha_i < \beta\sigma \tag{49}$$

which illustrates the fact that window does not have edge information.

For the ease of notation, a new parameter γ is defined such that

$$\gamma = \begin{cases} 1 & \text{if window belongs to group 1} \\ 0 & \text{if window belongs to group 2} \end{cases} \tag{50}$$

The final step for the denoising purpose is to determine the threshold value. In this paper, the threshold value is expressed by

$$threshold = \begin{cases} \sigma & \text{if } \gamma = 1 \\ \beta\sigma & \text{if } \gamma = 0 \end{cases} \tag{51}$$

where γ is obtained using equation (50) and this threshold is applied to the singular value of each window. For windows that belong to group 1, if $\alpha_i > \sigma$, then we keep that α_i and if not, then we turn α_i to zero. For windows that belong to group 2, if $\alpha_i > \beta\sigma$, then we keep that α_i and if not, then we turn α_i to zero.

We separate the problem into two parts which $\beta > 1$ and $\beta < 1$. In both cases, windows that belong to group 2 are equal to zero because they are treated as a pure noise. For the case $\beta > 1$, all of the windows that belong to group 1 will be retained, however, for the case $\beta < 1$ some of them can be eliminated and some of them can be retained and it totally depends on the values of β and each α_i.

In equations (35) and (36), the mean and the variance of each Gaussian random variable $\overline{d_{1k}}$ are obtained. Now we proceed our work to obtain the mean and the variance of each singular value of each window. For the mathematical convenience, our decision rule in equation (46) will become

$$median\{\alpha_i^2\} \lesseqgtr \beta^2\sigma^2 \tag{52}$$

Also, the above equation can be written

$$\lambda_i \lesseqgtr \beta^2\sigma^2 \tag{53}$$

where λ_i is equal to the eigenvalue of each window. Suppose that the window W_i is represented by

$$W_i = [x_{i1} \quad \dots \quad x_{il}]_{1\times l} \tag{54}$$

The eigenvalue of each window can be calculated by the following equation

$$det\big(\lambda_i - W_i W_i^T\big) = 0 \tag{55}$$

where λ_i will be obtained

$$\lambda_i = \sum_{j=1}^{l} x_{ij}^2 \tag{56}$$

Each x_{ij} is a Gaussian random variable with the mean d_{1ij} and standard deviation σ. Now, instead of α_i, our interest is to obtain the mean and the variance of λ_i because our decision rule was changed into equation (53). In order to obtain the mean of λ_i, the following procedures are carried out

$$E\big(\lambda_i\big) = E\big(\sum_{j=1}^{l} x_{ij}^2\big) \tag{57}$$

Using the fact that each x_{ij} is independent, so equation (57) will be

$$E\big(\lambda_i\big) = \sum_{j=1}^{l} E\big(x_{ij}^2\big) \tag{58}$$

If we denote d_{1ij} with μ_{ij}, $E\big(x_{ij}^2\big)$ will be

$$E\big(x_{ij}^2\big) = \sigma^2 + \mu_{ij}^2 \tag{59}$$

As a result, the mean of λ_i will be obtained

$$E\big(\lambda_i\big) = l\sigma^2 + \sum_{j=1}^{l} \mu_{ij}^2 \tag{60}$$

Following procedures are carried out to obtain the variance of λ_i

$$var\big(\lambda_i\big) = var\big(\sum_{j=1}^{l} x_{ij}^2\big) \tag{61}$$

Using the fact that each x_{ij} is independent, so equation (61) will be

$$var\big(\lambda_i\big) = \sum_{j=1}^{l} var\big(x_{ij}^2\big) \tag{62}$$

where $var\big(x_{ij}^2\big)$ is

$$var\big(x_{ij}^2\big) = E\big(x_{ij}^4\big) - E^2\big(x_{ij}^2\big) \tag{63}$$

Also using equation (59), equation (63) will be

$$var\big(x_{ij}^2\big) = E\big(x_{ij}^4\big) - \big(\sigma^2 + \mu_{ij}^2\big)^2 \tag{64}$$

$E\big(x_{ij}^4\big)$ can be expressed by

$$E\big(x_{ij}^4\big) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x_{ij}^4 exp\big(-\frac{1}{2\sigma^2}(x_{ij} - \mu_{ij})^2\big) dx_{ij} \tag{65}$$

which is calculated and the result is illustrated below

$$E\big(x_{ij}^4\big) = \mu_{ij}^4 + 6\sqrt{2\pi}\mu_{ij}^2\sigma^3 + 3\sqrt{2\pi}\sigma^5 \tag{66}$$

Using equation (66) and (64), $var\big(x_{ij}^2\big)$ can be obtained

$$var\big(x_{ij}^2\big) = 3\sqrt{2\pi}\sigma^5 - \sigma^4 + 6\sqrt{2\pi}\mu_{ij}^2\sigma^3 - 2\sigma^2\mu_{ij}^2 \tag{67}$$

As a result, $var\big(\lambda_i\big)$ will be

$$var\big(\lambda_i\big) = l\big(3\sqrt{2\pi}\sigma^5 - \sigma^4\big) + \big(6\sqrt{2\pi}\sigma^3 - 2\sigma^2\big) \sum_{j=1}^{l} \mu_{ij}^2 \tag{68}$$

Equations (60) and (68) are the mean and the variance of λ_i, respectively.

Before that the signal is corrupted by the noise process, λ_i of each window was

$$\lambda_i = \sum_{j=1}^{l} \mu_{ij}^2 \tag{69}$$

Comparison between equation (69) and equation (60) will lead us to the fact that equation (53) and equivalently equation (46) can be considered as an accurate decision rule for the denoising purpose. Consequently, all of our interpretations and results were correct. In this paper, we prove our denoising scheme for $1D$ input signal, however, it will also be correct for $2D$ input signal like image.