SUPER-RESOLUTION FROM UNREGISTERED ALIASED IMAGES WITH UNKNOWN SCALINGS AND SHIFTS

Yigang Peng[*], *Feng Yang*[†], *Qionghai Dai*[*], *Wenli Xu*[*], *Martin Vetterli*[†]

[*]TNList and the Department of Automation,
Tsinghua Univerisity, 100084, Beijing, China
E-mail: pyg07@mails.tsinghua.edu.cn, {qionghaidai, xuwl}@tsinghua.edu.cn
[†]LCAV-School of Computer and Communication Sciences,
Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015, Lausanne, Switzerland
E-mail: {feng.yang, martin.vetterli}@epfl.ch

ABSTRACT

We consider the problem of super-resolution from unregistered aliased images with unknown spatial scaling factors and shifts. Due to the limitation of pixel size in the image sensor, the sampling rate for each image is lower than the Nyquist rate of the scene. Thus, we have aliasing in captured images, which makes it hard to register the low-resolution images and then generate a high-resolution image. To work out this problem, we formulate it as a multichannel sampling and reconstruction problem with unknown parameters, spatial scaling factors and shifts. We can estimate the unknown parameters and then reconstruct the high-resolution image by solving a non-linear least square problem using the variable projection method. Experiments with synthesized 1-D signals and 2-D images show the effectiveness of the proposed algorithm.

Index Terms— super-resolution imaging, multichannel sampling, variable projection method, nonlinear least squares

1. INTRODUCTION

To keep a reasonable noise level and also due to technology limitation, in today's commercial image sensor design, one often chooses a large pixel size. The problem with this design is that it makes the sampling rate of the image sensor lower than the Nyquist rate of the scene they want to capture and we have aliasing in the acquired images.

Can we remove the aliasing and get a high-resolution image? Researchers proposed super-resolution imaging scheme [1]. In this imaging scheme, they take several images of the same scene with slight movement of the camera, and use computational methods to generate a high-resolution image from the low-resolution images. Most super-resolution imaging methods contain two steps: image registration and reconstruction. Vandewalle et al. [2] suggested that a joint consideration of these two steps would provide opportunities to achieve a better global solution. They considered the super-resolution imaging problem as a multichannel sampling problem and proposed several algorithms to jointly register the low-resolution images and generate a high-resolution image. The limitation of their model is that they assumed the camera only had translation motion.

Lu and Vetterli [3] extended such a model by allowing different gains for the acquisition of the images and proposed a fast reconstruction method via overparameterization.

In this paper, we also consider the super-resolution imaging as a multichannel sampling problem. We extend Vandewalle's translation motion model to a more generic motion model by considering spatial scaling factors. (In the rest of the paper, we call spatial scaling factor as scaling for short.) The reason is that in most cases we would change focal lengths or distances between the scene and the camera when capturing the low-resolution images, which makes the images have different scalings. Adding this makes the model more realistic. For simplicity, in the rest of the paper, we call the scalings and shifts as motion parameters. We show that although the non-linear dependence between motion parameters and captured images becomes more complicated, it is still possible to recover the correct motion parameters if we have enough images. We use a variable projection method [4] to estimate the unknown motion parameters and the high-resolution image.

The paper is organized as follows. In Section 2, we describe the super-resolution imaging model, how to formulate this as a multi-channel sampling scheme and the problem we want to solve. We propose a variable projection method to obtain the unknown motion parameters and reconstruct the high-resolution image in Section 3. The experimental results are presented in Section 4, and the conclusion is given in Section 5.

2. IMAGING MODEL AND PROBLEM STATEMENT

In this section, we describe our super-resolution imaging model, formulate the super-resolution imaging as a multichannel sampling problem with unknown motion parameters, and establish the reconstruction problem we want to solve in this paper. To simplify the presentation, we focus our discussion on a one-dimensional (1-D) sensor array, but all the results can be easily generalized to the 2-D case.

2.1. Bandlimited light intensity field model

We consider the super-resolution imaging model as in Fig. 1. The incoming light intensity field is $g_0(t)$. After passing the lens, which acts as a low-pass filter, it becomes a approximately bandlimited signal $g(t)$.
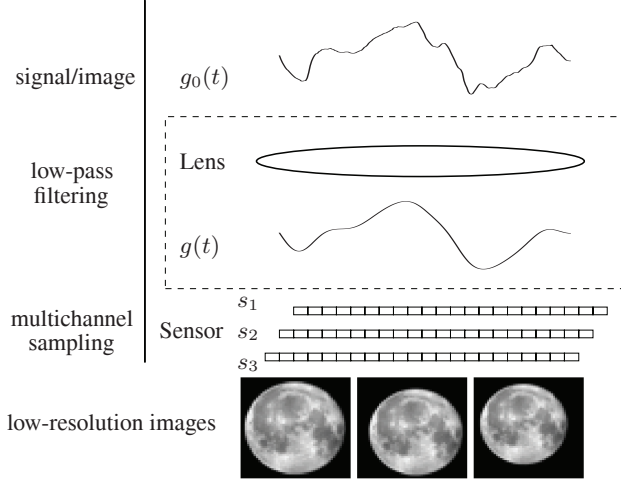
Fig. 1. The super-resolution imaging model. The lens filters the incident light intensity field $g_0(t)$ and generates a smoothed light intensity field $g(t)$. Then we use multiple sensors with different motion parameters to capture the light intensity field $g(t)$ and get several low-resolution images.

The bandlimited signal $g(t)$ can be expressed as [2]:

$$g(t) = \sum_{l=-K}^{K} \beta_l e^{j2\pi l t}, \qquad (1)$$

where β_l are the $L = 2K + 1$ Fourier coefficients of $g(t)$. We can easily check that the period of this signal is 1.

2.2. Multichannel sampling of the light intensity field

We then use the sensors in Fig. 1 to sample the light intensity field $g(t)$. Physically there is only one image sensor in the camera. But through moving the camera, zooming in or out, we can assume that the camera has multiple sensors with different motion parameters.

This can be formulated as a multichannel sampling problem shown in Fig. 2. We sample the signal $g(t)$ using M channels with unknown scalings $\{a_m\}$ and shifts $\{t_m\}$, $1 \leq m \leq M$. After scaling and shifting, the signal $g(t)$ becomes $g(a_m t + t_m)$, which is equal to $g(t) * (a_m \delta(a_m t + t_m))$, where $*$ denotes the convolution. Since the period of $g(t)$ is 1, the sampling interval $T = \frac{1}{N}$. Note that, due to the limitation of the pixel size in the image sensor, the sampling rate is sub-Nyquist. We can write the samples $y_m(n)$, $1 \leq m \leq M$ as,

$$
\begin{aligned}
y_m[n] &= g(a_m \frac{n}{N} + t_m) = \sum_{l=-K}^{K} \beta_l e^{j2\pi l(a_m \frac{n}{N} + t_m)} \\
&= \sum_{l=-K}^{K} \beta_l W^{a_m l n} z_m^l, \qquad (2)
\end{aligned}
$$

where $W = e^{j2\pi/N}$ and $z_m = e^{j2\pi t_m}$.
Let $\boldsymbol{F}^* = [f_{ij}] =$

$$
\begin{bmatrix}
1 & \cdots & 1 & 1 & 1 & \cdots & 1 \\
W^{-K} & \cdots & W^{-1} & 1 & W & \cdots & W^K \\
\vdots & & \vdots & \vdots & \vdots & & \vdots \\
W^{-(N-1)K} & \cdots & W^{-(N-1)} & 1 & W^{N-1} & \cdots & W^{(N-1)K}
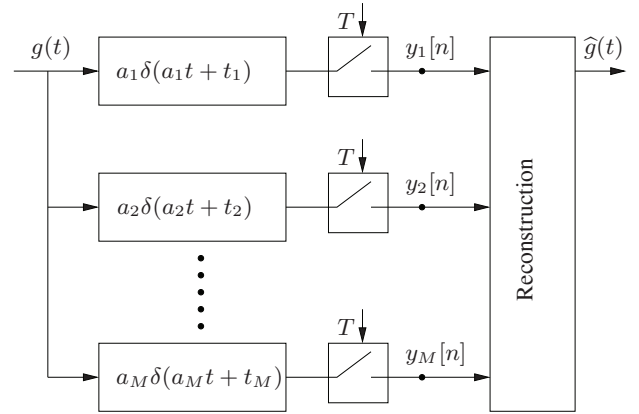\end{bmatrix}
$$

Fig. 2. The multichannel sampling scheme, where each channel uniformly samples a shifted and scaled version of a bandlimited signal $g(t)$. The shifts $\{t_m\}$ and scalings $\{a_m\}$ are unknown. Our goal is to reconstruct $g(t)$ from all the samples $y_m[n]$.

be the $N \times L$ inverse discrete Fourier transform (IDFT) matrix [2], which is the Hermitian transpose of the forward Fourier transform matrix $\boldsymbol{F}$, $[f_{ij}^{a_m}]$ denote the pointwise exponential of the matrix $\boldsymbol{F}^*$, $\boldsymbol{D}_{t_m}$ be an $L \times L$ diagonal matrix with its diagonal elements $\boldsymbol{D}_{t_m}(l, l) = z_m^l$ ($-K \leq l \leq K$), and $\boldsymbol{\beta} = [\beta_{-K} \cdots \beta_0 \cdots \beta_K]^T$ be the Fourier coefficients, then the matrix form of (2) is,

$$\boldsymbol{y}_m = \left[f_{ij}^{a_M}\right] \boldsymbol{D}_{t_m} \boldsymbol{\beta}.$$

By stacking the M channels' samples together into a single vector $\boldsymbol{y}$, and $\left[f_{ij}^{a_m}\right] \boldsymbol{D}_{t_m}$ into $F(\boldsymbol{a}, \boldsymbol{t})$, where $\boldsymbol{a} = [a_1 \ldots a_M]^T$ and $\boldsymbol{t} = [t_1 \ldots t_M]^T$, we get

$$
\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_M \end{bmatrix} = \begin{bmatrix} \left[f_{ij}^{a_1}\right] \boldsymbol{D}_{t_1} \\ \vdots \\ \left[f_{ij}^{a_M}\right] \boldsymbol{D}_{t_M} \end{bmatrix} \boldsymbol{\beta} = \boldsymbol{F}(\boldsymbol{a}, \boldsymbol{t}) \boldsymbol{\beta}. \qquad (3)
$$

2.3. Problem statement

We want to reconstruct the signal $g(t)$ from the samples. Due to the sub-Nyquist sampling of each channel, we can not restore $g(t)$ from a single channel. We need to use samples of the M channels to reconstruct $g(t)$. Because $g(t)$ is bandlimited and has L degree of freedom, to recover $g(t)$ is equivalent to reconstruct its L Fourier coefficients $\{\beta_l\}$. This becomes difficult since the scalings $\boldsymbol{a}$ and shifts $\boldsymbol{t}$ are also unknown in our setup. From the above analysis, the samples $\boldsymbol{y}$, scalings $\boldsymbol{a}$ and shifts $\boldsymbol{t}$, and Fourier coefficients $\boldsymbol{\beta}$ have a nonlinear relationship as in (3). So we need to solve equation (3) to obtain all the unknowns.

3. RECONSTRUCTION VIA VARIABLE PROJECTION METHOD

In this section, we show how to simultaneously register low-resolution images and reconstruct a high-resolution image via the variable projection method.
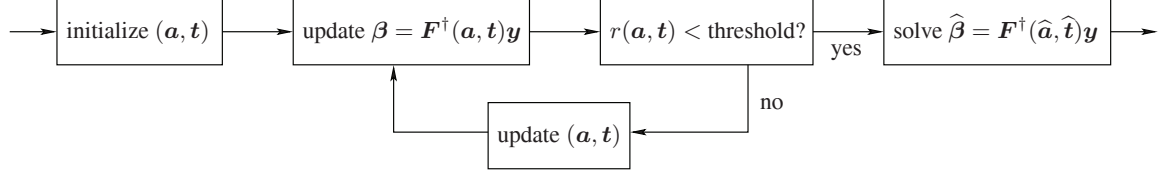
Fig. 3. The diagram of the variable projection method. We estimate the motion parameters $(\boldsymbol{a}, \boldsymbol{t})$ and the Fourier coefficients $\boldsymbol{\beta}$ iteratively. Once the iteration stops, we reconstruct the high-resolution image from the estimated Fourier coefficients $\widehat{\boldsymbol{\beta}}$.

3.1. Sufficient condition for reconstruction

Note that, each channel is sampled at a sub-Nyquist rate, i.e., $N < L$. If we only use a single channel, the number of unknowns is larger than the number of equations. So equation (3) is an underdetermined system. But because of the multichannel sampling, we have MN samples totally, and $L + 2(M - 1)$ unknown variables including Fourier coefficients and motion parameters (Without loss of generality, we can assume $a_1 = 1$ and $t_1 = 0$.). If the number of equations MN is larger than or equal to the number of unknowns $L + 2(M - 1)$, i.e.,

$$MN \geq L + 2(M - 1),$$

equation (3) becomes an overdetermined or determined system. Then we can get the correct parameters and the high-resolution image. Remark that numerical instability might occur if the matrix $\boldsymbol{F}(\boldsymbol{a}, \boldsymbol{t})$ is near singular. As observed in [2], more samples will lead to more accurate solutions.

3.2. Variable projection method

In this section, we describe the variable projection method [4] for estimating the motion parameters and Fourier coefficients. Notice that in equation (3), the two sets of variables, i.e., the motion parameters $(\boldsymbol{a}, \boldsymbol{t})$ and the Fourier coefficients $\boldsymbol{\beta}$, are in separable form. We show that a variable projection method can solve this problem by iteratively estimating Fourier coefficients $\boldsymbol{\beta}$ and motion parameters $(\boldsymbol{a}, \boldsymbol{t})$ as in Fig. 3.

Let $\boldsymbol{F}^{\dagger}(\boldsymbol{a}, \boldsymbol{t})$ denote the Moore-Penrose pseudoinverse of the matrix $\boldsymbol{F}(\boldsymbol{a}, \boldsymbol{t})$, then

$$\boldsymbol{F}^{\dagger}(\boldsymbol{a}, \boldsymbol{t}) = \left(\boldsymbol{F}^{*}(\boldsymbol{a}, \boldsymbol{t})\boldsymbol{F}(\boldsymbol{a}, \boldsymbol{t})\right)^{-1} \boldsymbol{F}^{*}(\boldsymbol{a}, \boldsymbol{t}),$$

where $\boldsymbol{F}^{*}(\boldsymbol{a}, \boldsymbol{t})$ is the Hermitian transpose of $\boldsymbol{F}(\boldsymbol{a}, \boldsymbol{t})$.

From (3), when the motion parameters $(\boldsymbol{a}, \boldsymbol{t})$ are known, we can get the Fourier coefficients $\boldsymbol{\beta}$ by solving a least squares problem and the result is

$$\boldsymbol{\beta} = \boldsymbol{F}^{\dagger}(\boldsymbol{a}, \boldsymbol{t})\boldsymbol{y}. \tag{4}$$

We can define the residual as

$$r(\boldsymbol{a}, \boldsymbol{t}) \stackrel{\text{def}}{=} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{a}, \boldsymbol{t})\boldsymbol{\beta}\|_2^2$$
$$= \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{a}, \boldsymbol{t})\boldsymbol{F}^{\dagger}(\boldsymbol{a}, \boldsymbol{t})\boldsymbol{y}\|_2^2, \tag{5}$$

where (5) follows from (4).

So the variables $(\boldsymbol{a}, \boldsymbol{t})$ can be estimated by solving the following nonlinear least square problem

$$(\widehat{\boldsymbol{a}}, \widehat{\boldsymbol{t}}) = \min_{\boldsymbol{a}, \boldsymbol{t}} r(\boldsymbol{a}, \boldsymbol{t}). \tag{6}$$

Once we get the estimated motion parameters $(\widehat{\boldsymbol{a}}, \widehat{\boldsymbol{t}})$, the estimated Fourier coefficients $\widehat{\boldsymbol{\beta}}$ can be computed from equation (4) and $g(t)$ can be reconstructed using equation (1).

We can extend the above model and algorithm to 2-D images with little effort. A bandlimited image is modeled as

$$g(\boldsymbol{x}) = \sum_{k_1=-K_1}^{K_1} \sum_{k_2=-K_2}^{K_2} \beta_{\boldsymbol{k}} e^{j2\pi \boldsymbol{k}^T \boldsymbol{x}}, \tag{7}$$

with $\boldsymbol{k} = [k_1 \ k_2]^T$ and $\boldsymbol{x} = [x_1 \ x_2]^T$. Suppose the sampling rates for the two dimensions are N_1 and N_2 respectively. Then, the mth image is:

$$y_m[n_1, n_2] = g(a_{m,1}\frac{n_1}{N_1} + t_{m,1}, a_{m,2}\frac{n_2}{N_2} + t_{m,2}), \tag{8}$$

where $\{a_{m,1}\}$ and $\{a_{m,2}\}$ are the vertical and horizontal scalings, $\{t_{m,1}\}$, $\{t_{m,2}\}$ are the vertical and horizontal shifts. By stacking the image into a vector, we can get a similar formula as equation (3), and the rest of the analysis follows what has been done for 1-D signals.

4. EXPERIMENTAL RESULTS

In this section, we show the experimental results on synthesized 1-D signals and 2-D images. In the above model, the signals are assumed to be periodic. This is not the case in most real scenarios. However, we can assume that the scalings and shifts are small and neglect the differences between the signals due to their aperiodicity.

4.1. 1-D signals

We generate 1-D signals with a decay rate of $1/(|\omega| + 1)$ in the frequency domain. Such a model is often used for natural images [5, 6]. We set the sampling rate $N = 67$, the number of channels $M = 2$, and the number of Fourier coefficients $L = 111$. We also add additive white Gaussian noise with variance σ^2 to the measurements. Without loss of generality, we set the scaling $a_1 = 1$ and shift $t_1 = 0$. The ground truth of the scaling a_2 and shift t_2 are set to 1.05 and 2, where the unit of the shift is pixel size, i.e., $1/N$. If the error of the scaling is smaller than 10^{-3} and the error of the shift is smaller than 10^{-2}, a trial is considered to be successful. We repeat 300 independent simulations. In each simulation, the initialization of a_2 and t_2 are randomly chosen from the interval $[1.04, \ 1.06]$ and $[1, \ 3]$ separately. Fig. 4 (a), (b), and (c) show the mean absolute error (MAE) of the scaling estimation, shift estimation, and the success rate as a function of the signal-to-noise ratio (SNR), which is defined as

$$\text{SNR} = 10 \log_{10} \frac{1}{N\sigma^2} \sum_{n} y_m^2[n].$$

We can see that the higher the SNR, the better the estimation, and when the SNR is lower than 35dB, the estimation performance drops rapidly.
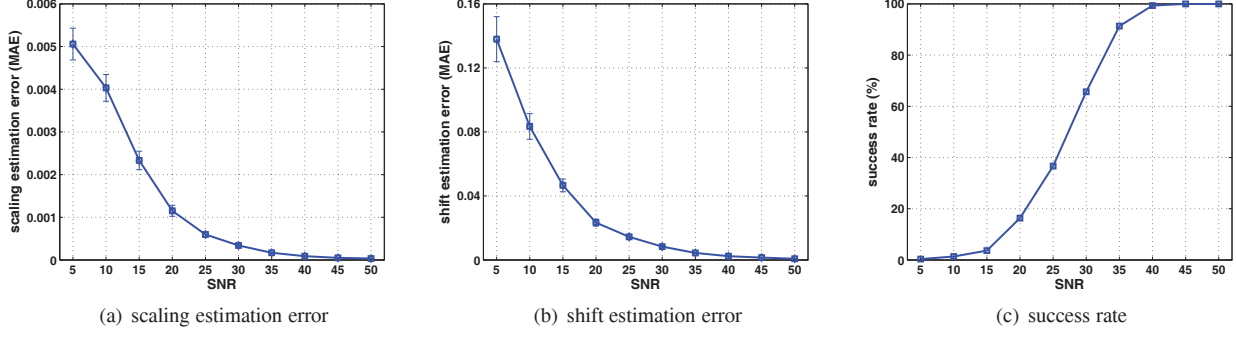
Fig. 4. Simulation results on 1-D signals as a function of the signal-to-noise ratio (SNR). The mean absolute error (MAE) for estimating (a) the scaling parameter and (b) the shift parameter. (c) Success rate under different SNRs.
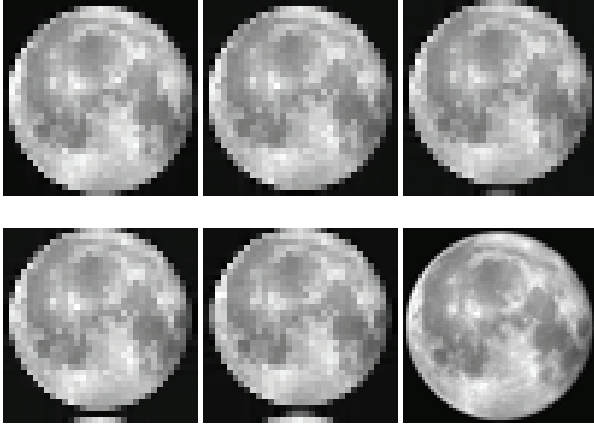

Fig. 5. Super-resolution reconstruction for 2-D images. The first five images are low-resolution images with resolution 33×33. The reconstructed high-resolution image with a resolution of 63×63 is the bottom right one.

4.2. 2-D images

We also implement our algorithm for 2-D images. The goal is to reconstruct an image with spatial resolution 63×63 from five 33×33 low-resolution images. In the simulation, we consider the case that the scalings in both dimensions of each image are the same. The scaling $a_{1,1} = a_{1,2} = a_1 = 1$, and shift parameters $t_{1,1} = t_{1,2} = 0$. Thus, we have to estimate the remaining 12 unknown motion parameters. The ground truth of the scaling $a_{m,1} = a_{m,2} = a_m$ and shift $t_{m,1}$ and $t_{m,2}$ for $m = 2, ..., 5$ are uniformly distributed in the interval $[1, 1.05]$, $[0, 1/N]$, and $[0, 1/N]$, respectively. Fig. 5 shows the five low-resolution images and the reconstructed high-resolution image which is the bottom right image.

5. CONCLUSION

In this paper, we considered super-resolution imaging from multiple unregistered aliased images with unknown scalings and shifts. We formulated it as a multichannel sampling problem. The unknown motion parameters and Fourier coefficients are separable in the nonlinear least square problem, so we could use a variable projection method to solve it. The method is also applicable to signals in a finite-dimensional Hilbert space. At present, the computational complexity is high because it is a large-scale nonlinear least square problem. In the future, it is interesting to investigate how to solve this problem more efficiently, and extend the algorithm to deal with more generic motion parameters including rotation [7]. It is also interesting to apply the algorithm to satellite imaging [8] and microscopy imaging [9], for example.

6. REFERENCES

[1] Peyman Milanfar, "Super-resolution imaging," *CRC Press, Roca Raton*, 2010.

[2] Patrick Vandewalle, Luciano Sbaiz, Joos Vandewalle, and Martin Vetterli, "Super-resolution from unregistered and totally aliased signals using subspace methods," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, part 2, pp. 3687–3703, 2007.

[3] Yue M. Lu and Martin Vetterli, "Multichannel sampling with unknown gains and offsets: A fast reconstruction algorithm," *Proc. Allerton Conference on Communication, Control and Computing, Monticello*, 2010.

[4] Gene Golub and Victor Pereyra, "Separable nonlinear least squares: The variable projection method and its applications," *Inverse Problems*, vol. 19, no. 2, pp. R1–R26, 2003.

[5] David J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am. A*, vol. 4, no. 12, 1987.

[6] Antonio Torralba and Aude Oliva, "Statistics of natural image categories," *Network: Computation in Neural Systems*, vol. 14, no. 3, pp. 391–412.

[7] Yu He, Kim-Hui Yap, Li Chen, and Lap-Pui Chau, "A nonlinear least square technique for simultaneous image registration and super-resolution," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2830–2841, 2007.

[8] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21 – 36, 2003.

[9] X.-Y. Zhai, E. I. Christensen, and A. Andreasen, "Low noise electron microscopy by merging multiple images digitized from conventional films with reference to the mouse kidney," *Microscopy and Microanalysis*, vol. 12, no. 3, pp. 255–261, 2006.