A FAMILY OF RISK ESTIMATORS AS CRITERIA FOR PSF ESTIMATION: FROM SURE TO GCV

Feng Xue, Lian Xue and Jiaqi Liu

National Key Laboratory of Science and Technology on Test Physics and Numerical Mathematics
Beijing, 100076, China

ABSTRACT

Stein's unbiased risk estimate (SURE) has been proven as a valid criterion for PSF (point spread function) estimation [1], which is essential to blind image deconvolution. In this paper, we develop a family of risk estimators as the criteria for blur identification. We first provide a direct proof of the validity of SURE. From this new perspective, we develop generalized cross validation (GCV) as a novel criterion and interpret it as a variant of SURE. A key advantage of GCV over SURE is that it does not depend on noise variance: we do not need to estimate it in advance.

We also provide a theoretical error analysis for the regularizer approximation within this SURE-type framework, by which we show that the error of PSF estimate is upper bounded by the approximation error. We further introduce a novel adaptive regularizer, which yields more accurate PSF estimate than other choices by extensive experimental tests.

Index Terms— PSF estimation, Stein's unbiased risk estimate (SURE), generalized cross validation (GCV), adaptive regularizer

1. INTRODUCTION

Blind image deconvolution is to recover the original image $\mathbf{x} \in \mathbb{R}^N$ from a blurred and noisy measurement [1]:

$$\mathbf{y} = \mathbf{H}_0\mathbf{x} + \mathbf{b}; \qquad \mu = \mathbf{H}_0\mathbf{x} \tag{1}$$

where $\mathbf{H}_0 \in \mathbb{R}^{N\times N}$ is a ground truth (unknown) convolution matrix constructed by PSF h_0, the vector $\mathbf{b} \in \mathbb{R}^N$ is a zero-mean additive white Gaussian noise with variance σ^2.

In most practical situations, such a problem can be cast by a separate strategy: PSF estimation is followed by non-blind deconvolution [1]. Regarding the PSF estimation, we have shown in [1] that Stein's unbiased risk estimate (SURE) is superior to other methods, e.g. APEX [2] and DL1C [3], in terms of estimation accuracy .

SURE is an unbiased estimate of the expected prediction error (EPE) under additive Gaussian noise assumption. For any linear estimate of μ, denoted by $\widehat{\mu} = \mathbf{U}\mathbf{y}$, EPE is defined as [4]:

$$\mathrm{EPE} = \frac{1}{N}\mathbb{E}\left\{\|\mathbf{U}\mathbf{y} - \mu\|^2\right\} \tag{2}$$

The SURE, given by [1]:

$$\mathrm{SURE} = \frac{1}{N}\left\|\mathbf{U}\mathbf{y} - \mathbf{y}\right\|^2 + \frac{2\sigma^2}{N}\mathrm{Tr}(\mathbf{U}) - \sigma^2 \tag{3}$$

has the same expectation with EPE. See Theorem 3.1 of [1] for the proof.

In our previous work [1], we proved that if the matrix $\mathbf{U}$ is an exact smoother filtering [4]:

$$\mathbf{U} = \mathbf{HSH}^{\mathrm{T}}\left(\mathbf{HSH}^{\mathrm{T}} + \sigma^2\mathbf{I}\right)^{-1} \Longleftrightarrow U(\omega) = \frac{|H(\omega)|^2}{|H(\omega)|^2 + \sigma^2/S(\omega)} \tag{4}$$

the EPE minimization yields exact PSF estimate in magnitude: $|H(\omega)| = |H_0(\omega)|$. Refer to Theorem 2.1 of [1] for the proof. Here, $\mathbf{S} = \mathbb{E}\{\mathbf{x}\mathbf{x}^{\mathrm{T}}\}$, $S(\omega)$ is a power spectral density of $\mathbf{x}$.

In [1], we used $\lambda\|\omega\|^2$ to approximate the unknown $\sigma^2/S(\omega)$ and finally formulated the PSF estimation as minimization of SURE over both $\mathbf{H}$ and λ. However, there are three main drawbacks with the approach presented in [1].

1. The computation of the SURE (3) requires the knowledge of the noise variance, which is, unfortunately, often unknown in practice.

2. It lacks a theoretical error analysis for the regularization approximation (e.g. the regularizer $\lambda\|\omega\|^2$).

3. The regularizer $\lambda\|\omega\|^2$ used in [1] is not adaptive to various types of images, which deteriorates the accuracy of PSF estimation.

As a continuation of the work [1], this paper is to address the foregoing problems. We derive generalized cross validation (GCV) from the perspective of SURE, and verify it as a novel criterion for PSF estimation. We also provide a theoretical error analysis that directly links the regularization approximations to the PSF estimation accuracy, based on which, we further propose a novel adaptive regularizer to improve the estimation accuracy.

2. GCV — A NOVEL CRITERION FOR PSF ESTIMATION

2.1. Direct proof of SURE criterion

In [1], we demonstrated the effectiveness of SURE by two steps: we first proved the validity of EPE minimization (Theorems 2.1 of [1]), and then, showed the statistical unbiasedness of SURE w.r.t. EPE (Theorems 3.1 of [1]). Instead of the indirect derivations, we now prove the validity of SURE in a straightforward way, as stated in the following ONE theorem.

Theorem 2.1 *If the matrix* $\mathbf{U}$ *is given by* (4)*, the minimization of expected SURE (given as* (3)*) over* $\mathbf{U}$ *yields the exact frequency magnitude of PSF:* $|H(\omega)| = |H_0(\omega)|$.

Proof From (3), the expected value of SURE is:

$$\begin{aligned}\mathbb{E}\{\mathrm{SURE}\} &= \frac{1}{N}\mathbb{E}\Big\{\|\mathbf{U}\mathbf{y}-\mathbf{y}\|^2\Big\} + \frac{2\sigma^2}{N}\mathrm{Tr}(\mathbf{U}) - \sigma^2 \\ &= \frac{1}{N}\mathbb{E}\Big\{\mathrm{Tr}\Big((\mathbf{U}-\mathbf{I})\mathbf{y}\mathbf{y}^{\mathrm{T}}(\mathbf{U}-\mathbf{I})^{\mathrm{T}}\Big)\Big\} + \frac{2\sigma^2}{N}\mathrm{Tr}(\mathbf{U}) - \sigma^2 \\ &= \frac{1}{N}\mathrm{Tr}\Big((\mathbf{U}-\mathbf{I})\underbrace{\mathbb{E}\{\mathbf{y}\mathbf{y}^{\mathrm{T}}\}}_{\mathbf{A}}(\mathbf{U}-\mathbf{I})^{\mathrm{T}}\Big) + \frac{2\sigma^2}{N}\mathrm{Tr}(\mathbf{U}) - \sigma^2 \quad (5)\end{aligned}$$

where $\mathbf{A} = \mathbf{H}_0\mathbf{S}\mathbf{H}_0^{\mathrm{T}} + \sigma^2\mathbf{I}$ by simple derivation with $\mathbf{S} = \mathbb{E}\{\mathbf{x}\mathbf{x}^{\mathrm{T}}\}$. Minimizing expected SURE over $\mathbf{U}$ leads to the minimizer $\mathbf{U}_0 = \mathbf{H}_0\mathbf{S}\mathbf{H}_0^{\mathrm{T}}(\mathbf{H}_0\mathbf{S}\mathbf{H}_0^{\mathrm{T}} + \sigma^2\mathbf{I})^{-1}$.

On the other hand, if we restrict the linear processing to $\mathbf{U} = \mathbf{H}\mathbf{S}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{S}\mathbf{H}^{\mathrm{T}} + \sigma^2\mathbf{I})^{-1}$, then, we have:

$$\mathbf{H}\mathbf{S}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{S}\mathbf{H}^{\mathrm{T}} + \sigma^2\mathbf{I})^{-1} = \mathbf{H}_0\mathbf{S}\mathbf{H}_0^{\mathrm{T}}(\mathbf{H}_0\mathbf{S}\mathbf{H}_0^{\mathrm{T}} + \sigma^2\mathbf{I})^{-1}$$

which yields $\mathbf{H}\mathbf{S}\mathbf{H}^{\mathrm{T}} = \mathbf{H}_0\mathbf{S}\mathbf{H}_0^{\mathrm{T}}$. In frequency domain, it is equivalent to $|H(\omega)| = |H_0(\omega)|$. ■

This theorem directly verifies SURE without resorting to EPE (2). We will see later that GCV can be developed and verified by slightly modifying the proof above.

2.2. Complementary smoother filtering — a new perspective of SURE

Note that Eq.(5) can be rewritten as:

$$\mathbb{E}\{\mathrm{SURE}\} = \frac{1}{N}\mathrm{Tr}\Big(\mathbf{M}\mathbf{A}\mathbf{M}^{\mathrm{T}}\Big) - \frac{2\sigma^2}{N}\mathrm{Tr}(\mathbf{M}) + \sigma^2 \quad (6)$$

where $\mathbf{M} = \mathbf{I} - \mathbf{U}$. Now, the expected SURE becomes a functional of $\mathbf{M}$, instead of $\mathbf{U}$. The minimizer of expected SURE is $\mathbf{M}_0 = \sigma^2\mathbf{A}^{-\mathrm{T}}$, i.e.,

$$\mathbf{M}_0 = \sigma^2(\mathbf{H}_0\mathbf{S}\mathbf{H}_0^{\mathrm{T}} + \sigma^2\mathbf{I})^{-1} \Longleftrightarrow M_0(\omega) = \frac{\sigma^2}{|H_0(\omega)|^2 S(\omega) + \sigma^2}$$

Thus, we obtain the following theorem.

Theorem 2.2 *If we restrict the linear processing to*

$$\mathbf{M} = \sigma^2(\mathbf{H}\mathbf{S}\mathbf{H}^{\mathrm{T}} + \sigma^2\mathbf{I})^{-1} \Longleftrightarrow M(\omega) = \frac{\sigma^2}{|H(\omega)|^2 S(\omega) + \sigma^2}, \quad (7)$$

the minimization of expected SURE (6) *over* $\mathbf{M}$ *yields the exact frequency magnitude of PSF:* $|H(\omega)| = |H_0(\omega)|$.

As a counterpart of Theorem 2.1, Theorem 2.2 implies that the SURE minimization is essentially equivalent to matching the filtering $M(\omega)$ in frequency domain. Recalling the definition of exact smoother filtering by Eq.(4), we have a connection that $M(\omega) + U(\omega) = 1$ for $\forall\omega$ or $\mathbf{M} + \mathbf{U} = \mathbf{I}$. Hence, we call $M(\omega)$ as *complementary smoother filtering*.

2.3. GCV — a variant of SURE

From the discussions above, we can see that the expected SURE can be a functional of either $\mathbf{U}$ or $\mathbf{M}$. Both interpretations, however, are essentially different, since $\mathbf{M}$ is linearly proportional to $\mathbf{A}^{-\mathrm{T}}$ (as shown in Eq.(7)), whereas $\mathbf{U}$ not. We now relax the expression of $\mathbf{M}$, assuming that the proportional factor can be any constant α, not necessarily σ^2:

$$\mathbf{M} = \alpha\mathbf{A}^{-\mathrm{T}} = \alpha(\mathbf{H}_0\mathbf{S}\mathbf{H}_0^{\mathrm{T}} + \sigma^2\mathbf{I})^{-1}$$

It is easy to verify that it is a minimizer of

$$J = \mathrm{Tr}\Big(\mathbf{M}\mathbf{A}\mathbf{M}^{\mathrm{T}}\Big) - 2\alpha\cdot\mathrm{Tr}(\mathbf{M}) + C$$

where C is any constant. Here, J can be regarded as a Lagrangian formulation of the following constrained optimization problem:

$$\min_{\mathbf{M}} \mathrm{Tr}\Big(\mathbf{M}\mathbf{A}\mathbf{M}^{\mathrm{T}}\Big) \qquad \text{s.t.} \quad \mathrm{Tr}(\mathbf{M}) = 1$$

with the undetermined Lagrangian multiplier 2α.

Here, α can also be understood as a scaling factor, which can be removed by evaluating the ratio:

$$\min_{\mathbf{M}} \frac{\mathrm{Tr}\Big(\mathbf{M}\mathbf{A}\mathbf{M}^{\mathrm{T}}\Big)}{\Big(\mathrm{Tr}(\mathbf{M})\Big)^2}$$

From (5), we have $\mathrm{Tr}\Big(\mathbf{M}\mathbf{A}\mathbf{M}^{\mathrm{T}}\Big) = \mathbb{E}\Big\{\|\mathbf{M}\mathbf{y}\|^2\Big\}$, thus, the ratio becomes:

$$\min_{\mathbf{M}} \mathbb{E}\left\{\frac{\|\mathbf{M}\mathbf{y}\|^2}{\Big(\mathrm{Tr}(\mathbf{M})\Big)^2}\right\} \quad (8)$$

It happens to be *generalized cross validation* (GCV), except for a constant factor N, compared to the standard GCV form [5]. It is easy to obtain the following theorem.

Theorem 2.3 *If the equality* (7) *holds, minimization of expected GCV* (8) *over* $\mathbf{M}$ *yields the exact frequency magnitude of PSF:* $|H(\omega)| = |H_0(\omega)|$.

From the discussions above, we can see that GCV can also be regarded as a variant of SURE. Both of them belong to the same family of risk estimators. Compared to SURE (3), a key advantage of GCV (8) is that it does not depend on noise variance σ^2: we do not need to estimate σ^2 in advance.

3. ERROR ANALYSIS FOR THE REGULARIZATION APPROXIMATION

3.1. Error analysis for PSF estimation

We now treat both SURE and GCV as funcitonals of $\mathbf{M}$. Note that the exact $M(\omega)$ in (7) cannot be used in practice, since $\sigma^2/S(\omega)$ is unknown. It is crucial to find a good regularization term $\lambda R(\omega)$ to approximate $\sigma^2/S(\omega)$, i.e.,

$$\mathbf{M}_{\mathbf{H},\lambda} = \lambda\left(\mathbf{H}\mathbf{R}^{-1}\mathbf{H}^{\mathrm{T}} + \lambda\mathbf{I}\right)^{-1} \Longleftrightarrow M_{H,\lambda}(\omega) = \frac{\lambda R(\omega)}{|H(\omega)|^2 + \lambda R(\omega)} \quad (9)$$

where λ is a regularization parameter.

Proposition 3.1 *Considering the minimization of expected SURE or expected GCV of* $\mathbf{M}_{\mathbf{H},\lambda}$*:*

$$\min_{\mathbf{H},\lambda} \frac{1}{N}\mathrm{Tr}\left(\mathbf{M}_{\mathbf{H},\lambda}\mathbf{A}\mathbf{M}_{\mathbf{H},\lambda}^{\mathrm{T}}\right) - \frac{2\sigma^2}{N}\mathrm{Tr}(\mathbf{M}_{\mathbf{H},\lambda}) + \sigma^2 \quad (10)$$

or

$$\min_{\mathbf{H},\lambda} \mathbb{E}\left\{\frac{\left\|\mathbf{M}_{\mathbf{H},\lambda}\mathbf{y}\right\|^2}{\left(\mathrm{Tr}(\mathbf{M}_{\mathbf{H},\lambda})\right)^2}\right\} \quad (11)$$

where $\mathbf{M}_{\mathbf{H},\lambda}$ *is defined as* (9)*, the estimation error of* $|H_0(\omega)|$ *satisfies the following inequality*

$$\left||H(\omega)|^2 - |H_0(\omega)|^2\right| \le C \cdot \delta_\omega$$

where δ_ω *denotes the approximation error of regularizer:* $\left|\lambda R(\omega) - \frac{\sigma^2}{S(\omega)}\right| \le \delta_\omega$ *for* $\forall\omega$*.*

Proof Both expected SURE and expected GCV can be reformulated and simplified as:

$$\min_{\mathbf{M}} \mathrm{Tr}\left(\mathbf{M}\mathbf{A}\mathbf{M}^{\mathrm{T}}\right) - 2\alpha\left(\mathrm{Tr}(\mathbf{M}) - 1\right)$$

which can be expressed in frequency domain:

$$\min_{V} \underbrace{\sum_{\omega}\left(V(\omega)+1\right)^{-2}\cdot\left(V_0(\omega)+1\right) - 2\frac{\alpha}{\sigma^2}\left(V(\omega)+1\right)^{-1}}_{J(V)}$$

where $V(\omega) = |H(\omega)|^2\frac{1}{\lambda R(\omega)}$ and $V_0(\omega) = |H_0(\omega)|^2\frac{S(\omega)}{\sigma^2}$. Here, we consider the minimization of J over $V(\omega)$, which is equivalent to over $M(\omega)$, since $M(\omega) = (V(\omega)+1)^{-1}$.

Taking the differentiation of J w.r.t. $V(\omega)$, and setting it to zero, the optimal $V^\star(\omega)$ should satisfy:

$$\sum_{\omega}\left(V^\star(\omega)+1\right)^{-3}\cdot\left(\frac{\alpha}{\sigma^2}V^\star(\omega) + \frac{\alpha}{\sigma^2} - V_0(\omega) - 1\right) = 0$$

To fix the parameter α, we consider the ideal case where the exact form $\lambda R(\omega) = \sigma^2/S(\omega)$ is applied, then, $|H^\star(\omega)| = |H_0(\omega)|$ and $V^\star(\omega) = V_0(\omega)$ by Theorems 2.1 and 2.3. Then the optimal condition becomes:

$$\sum_{\omega}\left(V_0(\omega)+1\right)^{-3}\cdot\left(\beta V_0(\omega)+\beta\right) = \beta\sum_{\omega}\left(V_0(\omega)+1\right)^{-2} = 0$$

where $\beta = \frac{\alpha}{\sigma^2} - 1$. Since $\sum_\omega\left(V_0(\omega)+1\right)^{-2} > 0$, then $\beta = 0$, i.e. $\alpha = \sigma^2$. Thus, the optimal condition reduces to:

$$\sum_{\omega}\left(V^\star(\omega)+1\right)^{-3}\cdot\left(V^\star(\omega) - V_0(\omega)\right) = 0$$

It yields that $V^\star(\omega) = V_0(\omega)$ for $\forall\omega$, from which we obtain:

$$|H^\star(\omega)|^2 = |H_0(\omega)|^2\lambda R(\omega)\frac{S(\omega)}{\sigma^2}, \quad \forall\omega$$

Thus,

$$\begin{aligned}\left||H^\star(\omega)|^2 - |H_0(\omega)|^2\right| &= |H_0(\omega)|^2\frac{S(\omega)}{\sigma^2}\left|\Delta\omega\right| \\ &\le \underbrace{\max_{\omega}\left(|H_0(\omega)|^2\frac{S(\omega)}{\sigma^2}\right)}_{C}\cdot\underbrace{\max_{\omega}\left|\Delta\omega\right|}_{\delta_\omega}, \quad \forall\omega\end{aligned}$$

where $\Delta\omega = \lambda R(\omega) - \frac{\sigma^2}{S(\omega)}$. ■

3.2. Choices of regularizer $R(\omega)$ and the comparisons

Proposition 3.1 states that the error of PSF estimation is upper bounded by the approximation error of regularizer. It is crucial to choose a good regularizer $R(\omega)$, s.t. δ_ω is very small. In [1], we proposed the following approximation:

$$M_{H,\lambda}(\omega) = \frac{\lambda\|\omega\|^2}{|H(\omega)|^2 + \lambda\|\omega\|^2} \quad (12)$$

where $R(\omega) = \|\omega\|^2$. However, this approximation cannot cope with a wide range of natural images, since it may not be a sufficiently accurate approximate of $\sigma^2/S(\omega)$.

To make the regularizer $R(\omega)$ more flexible, we now propose the following adaptive one:

$$M_{H,\lambda,\nu}(\omega) = \frac{\lambda\|\omega\|^\nu}{|H(\omega)|^2 + \lambda\|\omega\|^\nu} \quad (13)$$

where ν controls the varying speed. We finally formulate the PSF estimation as the minimization of either SURE (10) or GCV (11) over $\mathbf{H}$, λ and ν. The flowchart is shown in Fig.1.

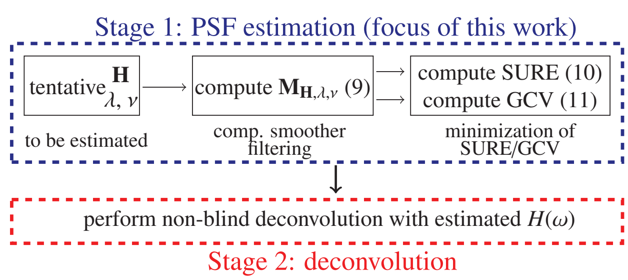

Fig. 1. The flowchart of PSF estimation: joint minimization of SURE/GCV over $\mathbf{H}$, λ and ν, as shown in (10) and (11).

By **Proposition 3.1**, we expect the proposed adaptive regularizer (13) to better approximate $\sigma^2/S(\omega)$, and thus, obtain more accurate PSF estimate. See Section 4 for further discussions.

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

4.1. Experimental setting

Now, we exemplify the proposed method (shown in Fig.1) with typical Gaussian kernel: $K \cdot \exp\left(-\frac{i^2+j^2}{2s^2}\right)$. we apply the Gaussian function with true blur size $s_0 = 2.0$ to blur test image *Coco* (shown in Fig.4), and corrupt it with noise of variances $\sigma^2 = 1$, 5, 10 and 50, respectively. Then, we use the proposed approach to estimate the blur size s.

4.2. SURE/GCV minimization: optimal ν

Taking $\sigma^2 = 1$ for example, Figs. 2–3 show the estimated ν and s by SURE/GCV minimization. We also compare it with the work of [1], where ν is fixed as 2 (i.e., $\|\omega\|^2$).

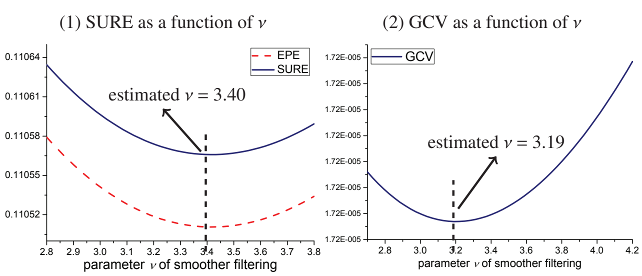

Fig. 2. Optimization of ν by SURE and GCV.

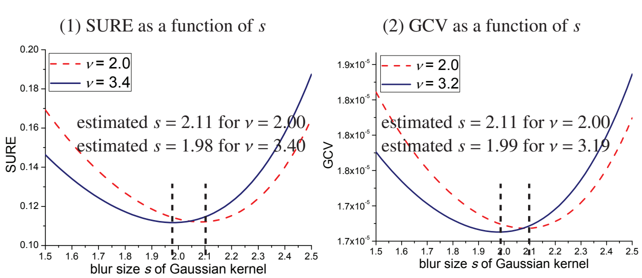

Fig. 3. Optimization of s by SURE and GCV.

Table 1 reports the comparisons between the work of [1] and our proposed approaches. We can see that the proposed adaptive regularizer yields more accurate estimate. Fig.4 shows a visual example, where we use our developed SURE-LET algorithm [6] to perform non-blind deconvolution.

Table 1. The estimated blur size s for *Coco*

σ^2	1	5	10	50
work of [1]	2.11	2.10	2.12	2.17
proposed SURE (10)	1.98	2.03	2.03	2.06
proposed GCV (11)	1.99	2.02	2.03	2.05

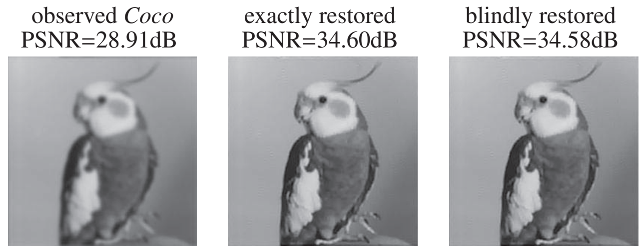

Fig. 4. Restoration of *Coco*: true $s_0 = 2.0$, estimated $s = 2.02$.

4.3. Application to real image

Finally, we apply the proposed method to real image *Fruit* captured by digital camera, shown in Fig.5. We assume the out-of-focus blur as Gaussian function. It can be easily seen from Fig.5 that our proposed approach achieves significant improvement of visual quality.

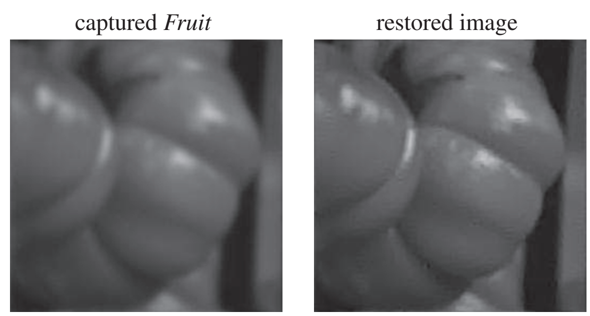

Fig. 5. Restoration of *Fruit*: the estimated blur size of Gaussian kernel is $s = 2.81$ by the proposed SURE/GCV method.

5. CONCLUSIONS

Proceeding with our previous work, this paper directly proved the validity of SURE as a criterion for PSF estimation, without referring to the true prediction error. We developed and verified GCV as a novel criterion, which avoids to estimate noise variance in advance.

We also provided an error analysis to directly link the regularization approximation to PSF estimation accuracy, which highlights the significance of selection of regularizer. In particular, we proposed an adaptive regularizer, to obtain more accurate estimate of PSF.

6. ACKNOWLEDGMENTS

The work was supported by the National Natural Science Foundation of China under Grant No. 61401013.

7. REFERENCES

[1] F. Xue and T. Blu, "A novel SURE-based criterion for parametric PSF estimation," *IEEE Trans. on Image Process.*, vol. 24, no. 2, pp. 595–607, 2015.

[2] A.S. Carasso, "The APEX method in image sharpening and the use of low exponent Lévy stable laws," *SIAM J. on Appl. Math.*, pp. 593–618, 2002.

[3] F. Chen and J.L. Ma, "An empirical identification method of Gaussian blur parameter for image deblurring," *IEEE Trans. on Signal Process.*, vol. 57, no. 7, pp. 2467–2478, 2009.

[4] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning*, vol. 2, Springer, 2009.

[5] G.H Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.

[6] F. Xue, F. Luisier, and T. Blu, "Multi-Wiener SURE-LET deconvolution," *IEEE Trans. on Image Process.*, vol. 22, no. 5, pp. 1954–1968, 2013.