EMPIRICAL CONDITIONAL MEAN: NONPARAMETRIC ESTIMATOR FOR COMPARAMETRIC EXPOSURE COMPENSATION

Dong Sik Kim, Su Yeon Lee*, and Kiryung Lee***

* School of Electronics and Information Engineering, Hankuk University of Foreign Studies, Kyonggi-do, 449-791, Korea
** LG Electronics Institute of Technology, Seoul, Korea

ABSTRACT

In this paper, a comparametric exposure compensation is conducted using a nonparametric estimator: empirical conditional mean. The Nadaraya-Watson estimator is used to smooth the empirical conditional mean curve especially for the case of small number of samples. The performance of the estimator is compared with those of the polynomial and piecewise-linear fittings. Designing the Nadaraya-Watson estimator is very simple and achieves lower errors than the fitting cases, which require a heavy computational burden of solving equations, without worry about the singular matrix case.

1. INTRODUCTION

In the formation of images in a camera, the exposure is appropriately adjusted to obtain a good image by means of the shutter speed and the aperture setting. In order to optimize the image quality, the exposure of each image could be different from each other depending on the object and the illumination. For example, digital cameras or scanners automatically adjust the exposure and even modify the image histograms. The exposure compensations in aligning the *serial section images* for the 3D reconstruction and in averaging *single particle images* for reducing the noise are important [1, ch. 3].

Employing the *comparametric function*, which represents a relationship between paired images having the same scene [2], we can compensate the exposure of an image with respect to another image, which is assumed as a reference one. If we know the response function of cameras, then we can derive the comparametric function. However, for the case of unknown response functions, we should consider a *parametric* or *nonparametric* estimator for deciding the comparametric function using samples. Most current approaches are using parametric estimators based on models or approximation [2],[3]. Mann [2] introduced sophisticated models for the response functions and consequently for the comparametric functions. In order to precisely describe wider class of comparametric functions without knowledge of the regression model, we may use the polynomial fitting (POL) [4] or the piecewise-linear fitting (PWL) [5, ch. 5]. In employing the parametric estimators, however, we are confronted with the model misspecification problem and the high computational complexity for solving complicate equations.

In a sense of minimizing the error between the reference and compensated images, an optimal predictor is the regression function given by the *conditional mean* (CM) [5, p. 19]. In this paper, a nonparametric estimator for CM, the Nadaraya-Watson estimator (NW) with the *Epanechnikov* kernel [6, p. 119] is considered for the comparametric exposure compensation. Under low computational complexity, this estimator can precisely describe any arbitrary-shaped curves. As a special case of NW, we also consider the *empirical conditional mean* (ECM), of which kernel is given by the indicator function. Using ECM, we can reduce the designing complexity of the parametric estimators, such as PWL.

This paper is organized in the following way. In Section 2, the exposure compensation problem is formulated in the frame work of regression analysis, and ECM and NW are introduced. A convergence analysis is introduced in Section 3 to see the performance with respect to the sample size. Compared with POL and PWL, numerical results with real images are shown in Section 4 for different exposures and sample sizes, and we conclude the findings in the last section.

2. EXPOSURE COMPENSATION BASED ON REGRESSION FUNCTION

An exposure compensation problem in a sense of minimizing the *mean squared error* (*mse*) is formulated in this section. Practically, the predictors are empirically designed by using samples from an unknown underlying distribution function. nonparametric estimators for the predictors are also formulated and discussed with the smoothing issue.

2.1. Optimal Predictor; Regression Function

For a simple regression model, suppose that $U, U_1, \ldots, U_m$ and $V, V_1, \ldots, V_m$ are independent, and identically distributed random variables, respectively, and $V, V_1, \ldots, V_m$ take values

This work was supported by grant No. R01-2005-000-10035-0 from the Basic Research Program of the Korea Science and Engineering Foundation.

from a finite set $\mathcal{V} := \{y_i\} \subset \mathbf{R}$. Suppose that the size of $\mathcal{V}$ is n. Note that U_ℓ and V_ℓ imply the reference and input images, respectively, with an underlying distribution function F, and the elements of $\mathcal{V}$ are the pixel levels. Thus, n could be 256 as an example.

We consider a map η as $x \mapsto \eta(x)$, where $x \in \mathcal{V}$ and $\eta(x) \in \mathbf{R}$. We call η a predictor for U using the input V. Here, $\eta(V)$ implies the compensated image. Let $D(\eta)$ denote the expected prediction error [5, p. 18] for a given predictor η, and define D as,

$$D(\eta) := E\{[U - \eta(V)]^2\}.$$

An optimal predictor in a sense of minimizing *mse* achieving the *optimum* $D^* := \min_\eta D(\eta)$ is the regression function of U on V [6]. Denoting such a predictor as η^*, $D(\eta^*) = D^*$ holds, and η^* is given by CM,

$$\eta^*(v) := E\{U \mid V = v\}, \text{ for } v \in \mathcal{V},$$

and the optimum D^* is equal to the expectation of the conditional variance, $D^* = E\{\mathrm{Var}\{U \mid V\}\}$ [7]. As a function of the optimum D^*, $D(\eta)$ can be rewritten by

$$D(\eta) = D^* + E\left\{[\eta^*(V) - \eta(V)]^2\right\} \tag{1}$$

for a given η. Hence, calculating only $E\left\{[\eta^*(V) - \eta(V)]^2\right\}$ is enough to check the performance of η. Note that predictors yield small D can be regarded as good predictors.

Instead of the average error D obtained by the underlying distribution function F, we usually consider the empirically averaged error, which is calculated by the samples under F. Using m independent observations of $\{(V_\ell, U_\ell)\}$, the empirical error D_m, which is usually considered as an unbiased estimate of D for a fixed η, is defined as

$$D_m(\eta) := \frac{1}{m}\sum_{\ell=1}^{m}[U_\ell - \eta(V_\ell)]^2.$$

Note that, for a fixed η, $E\{D_m(\eta)\} = D(\eta)$, and $D_m(\eta) \to D(\eta)$, almost surely, from the *law of large numbers*.

2.2. Empirical Conditional Mean

In order to find η^* with an unknown F, an inductive method can be considered based on minimizing $D_m(\eta)$. An empirically optimal predictor η_m^* that achieves the *empirical minimum*, $\min_\eta D_m(\eta) = D_m(\eta_m^*)$, is given by a nonparametric estimator defined as

$$\eta_m^*(v) := \sum_{\ell=1}^{m} I_{\{V_\ell\}}(v)U_\ell \Big/ \sum_{\ell=1}^{m} I_{\{V_\ell\}}(v), \tag{2}$$

if $\sum_{\ell=1}^{m} I_{\{V_\ell\}}(v) \neq 0$, and $\eta_m^*(v) := 0$ otherwise, for $v \in \mathcal{V}$. Here, for a set $S \subset \mathbf{R}$, $I_S(x) = 1$ if $x \in S$, and $I_S(x) = 0$ otherwise. We call this estimator ECM.

We now reformulate the empirical error D_m as a function of η_m^*. D_m can be expanded as follows:

$$\begin{aligned} D_m&(\eta) \\ &= \frac{1}{m}\sum_{i=1}^{n}\sum_{\ell=1}^{m}\Big[[I_{\{V_\ell\}}(y_i)\cdot U_\ell - \eta_m^*(y_i)]^2 \\ &\quad + [\eta_m^*(y_i) - \eta(y_i)]^2 \\ &\quad + 2[I_{\{V_\ell\}}(y_i)\cdot U_\ell - \eta_m^*(y_i)]\cdot[\eta_m^*(y_i) - \eta(y_i)]\Big] \end{aligned} \tag{3}$$

Here, $\sum_{\ell=1}^{m}[I_{\{V_\ell\}}(y_i)\cdot U_\ell - \eta_m^*(y_i)]\cdot[\eta_m^*(y_i) - \eta(y_i)] = 0$ holds from (2). Hence, from the definition of $D_m(\eta_m^*)$ and (3), we obtain the following relationship:

$$D_m(\eta) = D_m(\eta_m^*) + \Delta_m(\eta), \tag{4}$$

where $\Delta_m(\eta) := m^{-1}\sum_{i=1}^{n} m_i[\eta_m^*(y_i) - \eta(y_i)]^2$ and $m_i := \sum_{\ell=1}^{m} I_{\{V_\ell\}}(y_i)$. Here, m_i is the number of elements that are equal to y_i in $\{V_\ell\}$, and Δ_m is a function of ECM. (4) implies that minimizing the empirical error $D_m(\eta)$ with respect to predictors η, is equivalent to minimizing $\Delta_m(\eta)$. Hence, even when we design parametric estimators, such as POL and PWL, we may minimize Δ_m using ECM and can reduce the designing complexity.

2.3. Local Polynomial Kernel Estimators

In order to smooth the ECM curve especially for small m, we use the local polynomial kernel estimator [6, p. 119] of the area of statistical learning. We denote the estimator as $\hat{\eta}(v; p, \lambda)$ with the polynomial order p and the bandwidth λ. Among the local polynomial kernel estimators, NW ($p = 0$) is given by

$$\hat{\eta}(v; 0, \lambda) = \sum_{\ell=1}^{m} K_\lambda(V_\ell - v)U_\ell \Big/ \sum_{\ell=1}^{m} K_\lambda(V_\ell - v),$$

for $v \in \mathcal{V}$, where we employ the *Epanechnikov* quadratic kernel [5, p. 167]:

$$K_\lambda(v) = \begin{cases} (3/4)[1 - (v/\lambda)^2], & \text{if } |v/\lambda| \le 1 \\ 0, & \text{otherwise,} \end{cases}$$

where λ approaches zero, but at a rate slower than m^{-1} for the consistency.

3. CONVERGENCE RATES

If we consider an absolutely continuous distribution for the underlying distribution F, then from [6, pp. 125-129] $\hat{\eta}$ with the kernel having the bandwidth of order $m^{-1/5}$ converges to η^* at a rate of $m^{-4/5}$ in the *mean integrated squared error* sense [6, p. 35]. Here, the analysis is conditionally obtained for a given input. However, for the predictor η_m designed with the random pairs $\{(U_\ell, V_\ell)\}$, we should consider the *mean mean squared error* (*mmse*):

$$E\left\{E\{[\eta^*(V) - \eta_m(V)]^2 \mid U_1, \ldots, U_m, V_1, \ldots, V_m\}\right\} \tag{5}$$

to see the convergency of η_m. In the sense of *mmse* or the expectation of D_m, we can find some convergence analyses only for several specific regression models in the literature [7]. The affine function is the first-order Taylor approximation to η^* [7, p. 356] and is equivalent to the comparametric function in the *affine correction* (AC) when the response function has the power-law type [2, Proposition III.3]. This function is appropriate for estimating the regression function, which is based on the simple linear regression model [5, ch. 3], and can sometimes outperform the cases of nonparametric, or nonlinear models, especially in situations with small numbers of samples, low signal-to-noise ratio, or sparse data. From [7, p. 363], the bias of D_m is given by $D_1 \cdot 2m^{-1}$, which shows the rate $O(m^{-1})$. Here, $D_1 \geq D^*$ and equality holds for the simple linear regression model.

Under the discrete distribution assumption for a general class of regression models as in (1), from [8], we can explicitly derive the bias of $D_m(\eta_m^*)$, and obtain the following relationship, for any constant r,

$$D^* - E\{D_m(\eta_m^*)\} = c_\infty \cdot nm^{-1} + o(m^{-r}),$$

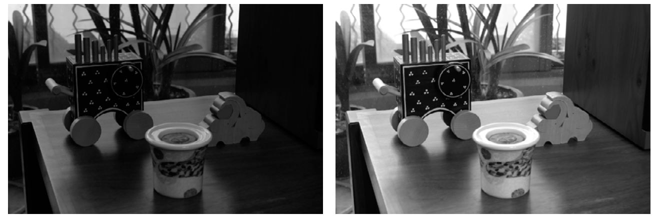
where $c_\infty := n^{-1} \sum_{i=1}^{n} E\{[U - \eta^*(V)]^2 \mid V = y_i\}$. We can also derive the bias of $D(\eta_m^*)$, and obtain a relationship:

$$E\{D(\eta_m^*)\} - D^* = c_\infty \cdot nm^{-1} + o(m^{-1}),$$

which is equal to (5) and results in the convergence $\eta_m^* \rightarrow \eta^*$ in the *mmse* sense at a rate of m^{-1}. Note that the rate m^{-1} is as fast as that of the usual parametric estimators. Further, for the equiprobable case as $\Pr\{V = y_i\} = 1/n$, $c_\infty = D^*$ holds. Hence, we can have a bias of the form $D^* \cdot nm^{-1}$, which implies the importance of the ratio n/m in obtaining a good estimator rather than the sample size m alone. Note that the bias $D^* \cdot nm^{-1}$ could be greater than the affine function case showing $D_1 \cdot 2m^{-1}$ for relatively small m. However, the bias of ECM is usually smaller than the affine function case since $D^* \ll D_1$, e.g., 13.5dB < 22.3dB as shown in the following section. Hence, even though the sample size is small, the nonparametric estimators, ECM and NW, can show a good performance compared to the parametric cases.

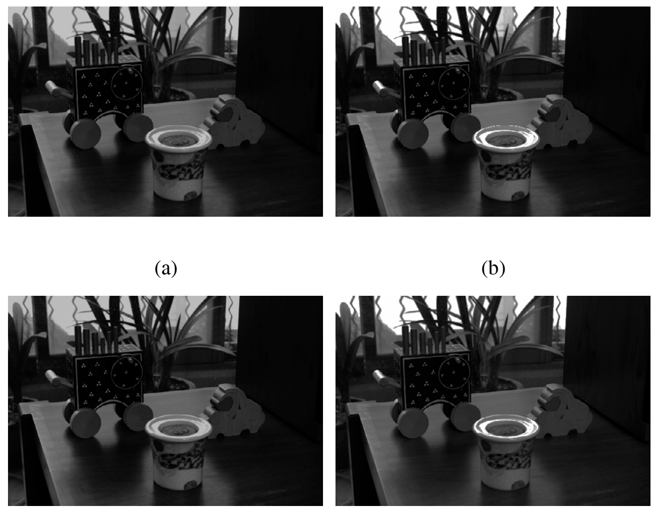
4. NUMERICAL RESULTS

In this section, we numerically observe the performance of ECM and NW comparing to the parametric methods; POL, PWL of Candocia [3], AC, and the *preferred correction* (PC) of Mann [2]. In the PWL case, we should construct a linear equation of which matrix size is given by the number of segments, and solve the equation avoiding the singular case. However, designing ECM requires only n multiplications as shown in (2). Further, the parametric methods can be designed using ECM by minimizing Δ_m from (4). Since the input has finite n values from the assumption, we can save the number of multiplications by gathering U_ℓs having the same V_ℓ. Note that this notion is achieved by using ECM.



(a) (b)

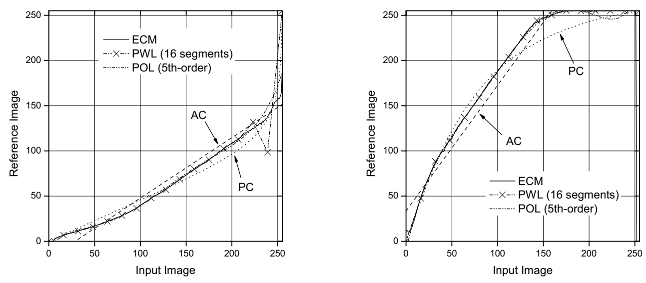
Fig. 1. Differently exposed images (600 × 400, 8 b/pixel, Nikon D70s). (a) Image I (1/2.5 sec, F11). (b) Image II (1.3 sec, F11, error=35.19dB).



(a) (b) (c) (d)

Fig. 2. Compensated images and errors in decibel (reference: Image I, input: Image II). (a) ECM: 14.69B. (b) 16-segments PWL [2]: 23.32dB. (c) 5th-order POL: 15.55dB. (d) PC: 23.46dB.

Fig. 1 shows two images obtained under different exposures on the same scene, and Fig. 2 shows results of the compensated images with the prediction error in decibel. We notice that the ECM case shows the best result, where a linear interpolation is applied to the $m_i = 0$ case. We could also do a similar compensation for differently illuminated images. The predictor curves of Fig. 2 are depicted in Fig. 3(a), where the AC yields the error of 21.44dB. At the right part of the comparametric plot of PWL (Fig. 3(a)), we may notice a large error, which is caused by the equal-spaced knots along the input image axis. On the other hand, if we use Images I and II as the input and reference ones, respectively, then PWL is as good as the ECM case (Fig. 3(b)). Here, the errors are given by 13.58, 13.86, 13.84, 24.62, and 18.02dB for the ECM, POL, PWL, AC, and PC cases, respectively. A consistent result could be found in Fig. 4, where a series of differently



(a) (b)

Fig. 3. Predictors for the comparametric compensation. (a) Reference: Image I, input: Image II (Fig. 2), (b) Reference: Image II, input: Image I.

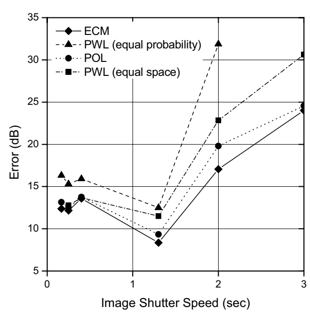
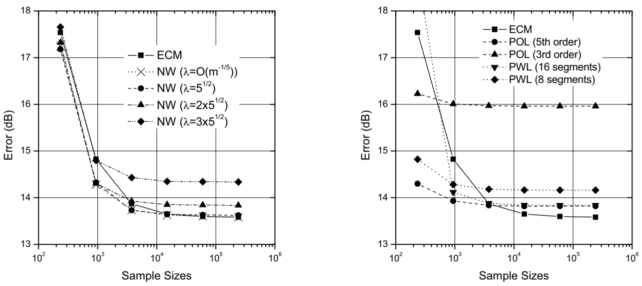


Fig. 4. Exposure compensation for differently exposed images (reference image shutter speed: 1/1.3 sec), where PWL: 16 segments and POL: 5th order.

exposed images are compensated with respect to a reference image having the shutter speed of $1/1.3$ sec. Fig. 4 shows that the compensating the right three images are more difficult than the left cases. Here, PWL shows even worse results than the ECM and POL cases. From the results of Fig. 3(a) and Fig. 4, we may notice that the performance of PWL is not good when the exposure of the reference image is less than that of the input image. Further, we should avoid the singular case by carefully assigning the segment knots. However, the compensation based on ECM or NW can simply achieve the minimum error without worry about the singular case.

ECM and NW are compared with POL and PWL in Fig. 5 for different sample sizes. As shown in Fig. 5(a), smoothing the ECM curve in NW reduces the prediction error for relatively small m, which is consistent with the POL and PWL cases as in Fig. 5(b). However, for relatively large sample size case, we have quite large biases. As shown in Fig. 5(a), the NW with a bandwidth of $\lambda = O(m^{-1/5})$ can achieve a good performance over a wide range of sample sizes.



(a) (b)

Fig. 5. Comparison of ECM, NW, POL, and PWL with respect to different sample sizes (reference: Image II and input: Image I).

5. CONCLUSION

In this paper, we conducted the comparametric exposure compensation based on the empirical conditional mean. We may notice that the compensation error can be minimized without solving any complicate equations. An overfitting problem in using the empirical conditional mean could be alleviated by introducing the Nadaraya-Watson estimator. Further research in conjunction with the registration to perform a joint optimization is being successfully conducted based on the Lucas-Kanade algorithm.

6. REFERENCES

[1] J. Frank, *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*, Academic, San Diego, CA, 1996.

[2] S. Mann, “Comparametric equations with practical applications in quantigraphic image processing,” *IEEE Trans. Image Processing*, vol. IP-9, no. 8, pp. 1389–1406, Aug. 2000.

[3] F. M. Candocia, “Jointly registering images in domain and range by piecewise linear comparametric analysis,” *IEEE Trans. Image Processing*, vol. IP-12, no. 4, pp. 409–419, April 2003.

[4] T. Mitsunaga and S. K. Nayar, “Comparametric equations with practical applications in quantigraphic image processing,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 23–25, June 1999.

[5] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, NY, 2001.

[6] M. P. Wand and M. C. Jones, *Kernel Smoothing*, Chapman & Hall, London, 1995.

[7] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Ye, *Probability & Statistics for Engineers & Scientists*, 7th. ed., Prentice Hall, NJ, 2002.

[8] D. S. Kim and M. R. Bell, “Upper bounds on empirically optimal quantizers,” *IEEE Trans. Inform. Theory*, vol. IT-49, no. 4, pp. 1037–1046, April 2003.