Signals and Images Foreground/Background Joint Estimation and Separation

Boujemaa Ait-El-Fquih and Ali Mohammad-Djafari

Laboratoire des Signaux et Systèmes UMR 8506
(CNRS-SUPELEC-UNIV PARIS SUD)
Supélec, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette, France

Abstract. This paper is devoted to a foreground/background joint estimation and separation problem. We first observe that this problem is modeled by a conditionally linear and Gaussian hidden Markov chain (CLGHMC). We next propose a filtering algorithm in the general non-linear and non Gaussian conditionally hidden Markov chain (CHMC), allowing the propagation of the filtering densities associated to the foreground and the background. We then focus on the particular case of our CLGHMC in which these filtering densities are weighted sums of Gaussian distributions; the parameters of each Gaussian are computed by using the Kalman filter algorithm, while the weights are computed by using the particle filter algorithm. We finally perform some simulations to highlight the interest of our method in both signals and images foreground/backgound joint estimation and separation.

Keywords: Foreground/background joint estimation and separation, Conditionally hidden Markov chain models, Kalman filter, Particle filter
PACS: 02.50.Tt, 02.50.Ga, 02.70.Uu

INTRODUCTION

An important problem in signal and image processing consists in jointly estimating and separating a background image $\mathbf{b}$ and a foreground one $\mathbf{f}$ from their superposition $\mathbf{y} = \mathbf{b} + \mathbf{f} + \varepsilon$. The image $\mathbf{b}$ is classically modeled by an homogeneous autoregressive (AR) process while $\mathbf{f}$ is non homogeneous and depending on a binary valued Markov variable $r \in \{r^b, r^f\}$ separating the foreground from the background. $\mathbf{f}|r = r^f$ and $\mathbf{f}|r = r^b$ are then modeled by two AR processes with different parameters. In that end, we develop an adaptive filtering algorithm allowing the restoration of $\mathbf{b}$, $\mathbf{f}$ and r from $\mathbf{y}$.

We first show that our problem is modeled by a CLGHMC, in which the model $((\mathbf{b}, \mathbf{f}), \mathbf{y})$ is a linear and Gaussian hidden Markov chain (LGHMC) conditionally on the discrete Markov chain (MC) r. We next develop a Bayesian filtering algorithm in a general non-linear and non Gaussian CHMC model, allowing the propagation of the filtering distributions associated to $\mathbf{b}$, $\mathbf{f}$ and r. We then return to our CLGHMC model and we observe that the filtering densities associated to $\mathbf{b}$ and $\mathbf{f}$ are Gaussian. Then propagating these densities amounts to propagating their parameters by a Kalman filtering-like algorithm. Furthermore, the propagation of the discrete filtering distributions associated to r reduces to the propagation of their Monte Carlo approximations by using a particle filtering-like algorithm. We finally provide some simulations results intended to show the interest of the proposed method in both signals and images foreground/background joint estimation and separation.

FOREGROUND/BACKGROUND JOINT ESTIMATION AND SEPARATION

Let $b = \{b_n\}_{n \geq 0}$, $f = \{f_n\}_{n \geq 0}$ and $r = \{r_n\}_{n \geq 0}$ be hidden processes; and $y = \{y_n\}_{n \geq 0}$ an observed process. For each n, b_n, f_n and y_n are continuous ($\in \mathbb{R}$) and r_n is discrete. Let note $b_{0:n} = \{b_i\}_{i=0}^n$ and $y_{0:n} = \{y_i\}_{i=0}^n$, and $p(b_n)$, $p(b_{0:n})$ and $p(b_n|y_{0:n})$, say, denote the probability density function (pdf) w.r.t. Lebesgue measure of b_n, the pdf of $b_{0:n}$, and the pdf of b_n conditionally on $y_{0:n}$. Similar notations are considered for f and y. Furthermore, without loss of generality the symbols $p(.)$ associated to the discrete process r define the probability mass functions (pmf) rather than the pdf.

A foreground/background separation problem can be illustrated by the Fig. 1 below.

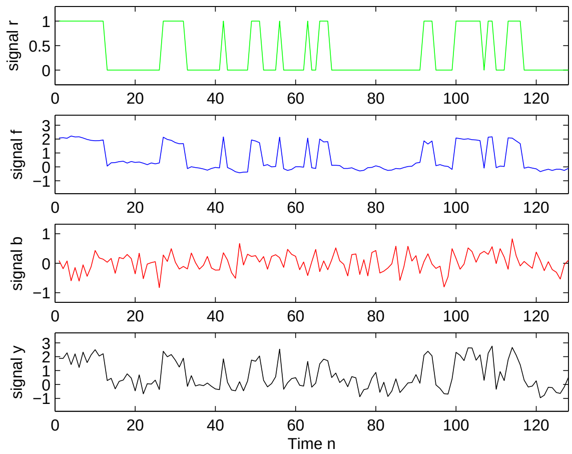

FIGURE 1. A foreground/background separation problem.

Indeed,

- The MC r takes its values in a binary set $\Omega = \{r^b, r^f\}$.
- The foreground signal f is a Gaussian AR process conditionally on the MC r. By assuming $r_n = k'$ and $r_{n+1} = k$ we have

$$f_{n+1} = a_n^f \delta(k - k') f_n + (1 - \delta(k - k')) m_n^f(k) + \varepsilon_n^f(k), \quad (1)$$

  with $\varepsilon^f(k) = \{\varepsilon_n^f(k)\}_n$ is a Gaussian independent process and independent of the Gaussian initial state f_0, and $\varepsilon_n^f(k) \sim \mathcal{N}(0, v_n^f(k))$. The variable $m_n^f(k)$ is deterministic and $\delta(.)$ is the symbol of Kronecker. Let us remark that the foreground signal is a Gaussian homogeneous AR process with $p(f_{n+1}|f_n, r_{n:n+1}) \sim \mathcal{N}(a_n^f f_n, v_n^f(r_n))$ if $r_{n+1} = r_n$, and an independent Gaussian process with $p(f_{n+1}|f_n, r_{n:n+1}) \sim \mathcal{N}(m_n^f(r_{n+1}), v_n^f(r_{n+1}))$ if $r_{n+1} \neq r_n$.
- The background signal b is modeled by a Gaussian AR process with

$$b_{n+1} = a_n^b b_n + \varepsilon_n^b, \quad (2)$$

with $\varepsilon^b = \{\varepsilon_n^b\}_n$ is a Gaussian independent process and independent of the Gaussian initial state b_0, and $\varepsilon_n^b \sim \mathcal{N}(0, v_n^b)$.

- The observed signal y is viewed as a noised superposition of b and f with

$$y_n = b_n + f_n + \varepsilon_n^y, \tag{3}$$

and $\varepsilon^y = \{\varepsilon_n^y\}_n$ is a Gaussian independent process and independent of b and f, and $\varepsilon_n^y \sim \mathcal{N}(0, v_n^y)$.

Conditional Markovian structure of the model (1)-(3)

Let us consider the continuous augmented state $\mathbf{x}_n = [b_n, f_n]^T$. From (1)-(3) one can easily check that

$$\begin{aligned} p(\mathbf{x}_{n+1}|\mathbf{x}_{0:n}, r_{0:n+1}) &= p(\mathbf{x}_{n+1}|\mathbf{x}_n, r_{n+1}, r_n), &(4)\\ &\sim \mathcal{N}(\mathbf{F}_n(r_{n+1}, r_n)\mathbf{x}_n + \mathbf{b}_n(r_{n+1}, r_n), \mathbf{Q}_n(r_{n+1})); &(5)\\ p(y_{0:n}|\mathbf{x}_{0:n}, r_{0:n}) &= \prod_{k=0}^{n} p(y_k|\mathbf{x}_k, r_k), &(6)\\ &\sim \prod_{k=0}^{n} \mathcal{N}(\mathbf{h}\mathbf{x}_k, v_k^y), &(7) \end{aligned}$$

with $\mathbf{h} = [1, 1]$, $\mathbf{F}_n(k, k') = \begin{bmatrix} a_n^b & 0 \\ 0 & a_n^f \delta(k - k') \end{bmatrix}$, $\mathbf{Q}_n(k) = \begin{bmatrix} v_n^b & 0 \\ 0 & v_n^f(k) \end{bmatrix}$ and $\mathbf{b}_n(k, k') = \begin{bmatrix} 0 \\ (1 - \delta(k - k'))m_n^f(k) \end{bmatrix}$.

As we can see, the process $\mathbf{x}$ is a linear and Gaussian MC conditionally on r, and y is an independent Gaussian process conditionally on $\mathbf{x}$. Therefore, the model $(\mathbf{x}, r, y)$ is a CLGHMC. Such models are used in many applications like for instance in speech enhancement [1] or object tracking [2]. The interest of CLGHMC is twofold.

- As far as modeling is concerned, CLGHMC $(\mathbf{x}, r, y)$ are more general then the classical LGHMC in the sens that none of the chains $\mathbf{x}$ or $(\mathbf{x}, y)$ need to be a MC.
- However, as far as restoration is concerned, in spite of their complexity, the CLGHMC models still enable to derive efficient Bayesian algorithms thanks to the conditional Markovian structure of $\mathbf{x}$ ($\mathbf{x}|r$ is a MC). Let us consider for instance the filtering problem. It consists in recursively computing the posterior distribution

$$p(\mathbf{x}_n, r_{0:n}|y_{0:n}) = p(\mathbf{x}_n|r_{0:n}, y_{0:n})p(r_{0:n}|y_{0:n}); \tag{8}$$

the distributions of interest $p(\mathbf{x}_n, r_n|y_{0:n})$, $p(\mathbf{x}_n|y_{0:n})$ and $p(r_n|y_{0:n})$ are then obtained by marginalization; and finally, the filtering estimate $\widehat{\mathbf{x}}_{n|n}$ of $\mathbf{x}_n$ and $\widehat{r}_{n|n}$ of r_n are given by the conditional mean (CM) of $p(\mathbf{x}_n|y_{0:n})$ and the maximum *a posteriori* (MAP) of $p(r_n|y_{0:n})$ respectively:

$$\widehat{\mathbf{x}}_{n|n} = \sum_{r_{0:n}} \left[\int \mathbf{x}_n p(\mathbf{x}_n, r_{0:n}|y_{0:n}) d\mathbf{x}_n \right] \tag{9}$$

$$\widehat{r}_{n|n} \;=\; \arg\max p(r_n|y_{0:n}). \tag{10}$$

Now, given $r_{0:n}$, $p(\mathbf{x}_n|r_{0:n}, y_{0:n})$ is a Gaussian distribution whose the parameters can be evaluated exactly by using the Kalman Filter (KF) [3] algorithm. On the other hand, the discrete distribution $p(r_{0:n}|y_{0:n})$ can be approximated by a sequential Monte Carlo method (see eg. [4]).

The practical computation of (9) and (10) will be detailed below in this paper. Before that, we develop in the following section a Bayesian filtering algorithm in the general CHMC, i.e. in a non-linear and non Gaussian model (in which only (4) and (6) hold but not (5) nor (7)).

BAYESIAN FILTERING IN GENERAL CHMC

Assume that we are given (4) and (6). The following proposition propagates the posterior pdf (8) from which the estimates (9) and (10) are next computed. The proof is omitted due to lack of space.

Proposition 1 *$p(\mathbf{x}_n|r_{0:n}, y_{0:n})$ and $p(r_{0:n}|y_{0:n})$ can be propagated by the following steps:*

- *Conditional prediction.*

$$p(\mathbf{x}_n|r_{0:n}, y_{0:n-1}) = \int p(\mathbf{x}_n|\mathbf{x}_{n-1}, r_n, r_{n-1}) p(\mathbf{x}_{n-1}|r_{0:n-1}, y_{0:n-1}) d\mathbf{x}_{n-1} \tag{11}$$

- *Conditional Innovation.*

$$p(y_n|r_{0:n}, y_{0:n-1}) = \int p(y_n|\mathbf{x}_n, r_n) p(\mathbf{x}_n|r_{0:n}, y_{0:n-1}) d\mathbf{x}_n \tag{12}$$

- *Filtering of $r_{0:n}$.*

$$p(r_{0:n}|y_{0:n}) = \frac{p(r_n|r_{n-1}) p(y_n|r_{0:n}, y_{0:n-1})}{p(y_n|y_{0:n-1})} p(r_{0:n-1}|y_{0:n-1}) \tag{13}$$

- *Conditional Filtering.*

$$p(\mathbf{x}_n|r_{0:n}, y_{0:n}) = \frac{p(y_n|\mathbf{x}_n, r_n) p(\mathbf{x}_n|r_{0:n}, y_{0:n-1})}{p(y_n|r_{0:n}, y_{0:n-1})}. \tag{14}$$

Then, $p(\mathbf{x}_n, r_{0:n}|y_{0:n})$ is computed as a product of (13) and (14); the estimates $\widehat{\mathbf{x}}_{n|n}$ and $\widehat{r}_{n|n}$ are deduced by using (9) and (10) respectively.

BAYESIAN FILTERING IN CLGHMC

Let us turn to our foreground/background separation problem, which indeed, is modeled by the CLGHMC ((5),(7)). Recall that $(\mathbf{x}, y)$ is a LGHMC conditionally on r. Then given r, all conditional distributions of $\mathbf{x}$ and y are Gaussian, in particular the conditional prediction pdf $p(\mathbf{x}_n|r_{0:n}, y_{0:n-1})$ and the conditional filtering pdf $p(\mathbf{x}_n|r_{0:n}, y_{0:n})$. Propagating these pdf amounts then to propagating their parameters. Furthermore, the propagation of the pmf $p(r_{0:n}|y_{0:n})$ reduces to that of its Monte Carlo approximation $\{r_{0:n}^{(s)}, w_n^{(s)}\}_{s=1}^{S \gg 1}$,

$$p(r_{0:n}|y_{0:n}) \approx \sum_{s=1}^{S} w_n^{(s)} \delta(r_{0:n} - r_{0:n}^{(s)}), \text{ with } \sum_{s=1}^{S} w_n^{(s)} = 1. \tag{15}$$

This can be done by the particle filtering (PF) algorithm [4] which does not recalled here due to lack of space.

Joint Estimation and Separation of r, b and f

Following (15) we have

$$p(r_n|y_{0:n}) \approx \sum_{s=1}^{S} w_n^{(s)} \delta(r_n - r_n^{(s)}). \tag{16}$$

Then the MAP $\widehat{r}_{n|n}$ (10) of r_n can be approximated as

$$\hat{r}_{n|n} \approx r_n^{(s)}, \text{ with } s = \arg \max_{s' \in \{1,\cdots,S\}} w_n^{(s')}. \tag{17}$$

As far as the filtering estimates $\widehat{b}_{n|n}$ of b_n and $\widehat{f}_{n|n}$ of f_n are concerned, let us first remark that these quantities can be computed by marginalization from $\widehat{\mathbf{x}}_{n|n}$ (9).

Now, by using (15) one can easily check that

$$p(\mathbf{x}_n|y_{0:n}) \approx \sum_{s=1}^{S} w_n^{(s)} p(\mathbf{x}_n|r_{0:n}^{(s)}, y_{0:n}). \tag{18}$$

Let us note

$$p(\mathbf{x}_n|r_{0:n}^{(s)}, y_{0:n}) \sim \mathcal{N}(\widehat{\mathbf{x}}_{n|n}^{(s)}, \widehat{\mathbf{P}}_{n|n}^{(s)}). \tag{19}$$

Then the pdf $p(\mathbf{x}_n|y_{0:n})$ in (18) is approximated by a weighted sum of Gaussian densities, each one can be computed exactly by a KF; the CM (9) can thus be approximated as

$$\widehat{\mathbf{x}}_{n|n} \approx \sum_{s=1}^{S} w_n^{(s)} \widehat{\mathbf{x}}_{n|n}^{(s)}. \tag{20}$$

Consequently, the algorithm of Prop. 1 which propagates $p(\mathbf{x}_n|y_{0:n})$ and $p(r_{0:n}|y_{0:n})$ in a general CHMC model, reduces to an algorithm allowing the propagation of a mixture of Gaussian densities (18) and of a discrete distribution (15). A similar algorithm has been introduced in [5] under the name of a *Mixture Kalman Filter* and whose the area of applications is very wide (see eg. [5] and the references therein). This algorithm is summarized in the following proposition. The proof is omitted for want of space.

Proposition 2 *Let us consider the CLGHMC ((5),(7)). We assume the notations (19),*

$$\begin{aligned} p(\mathbf{x}_n|r_n, r^{(s)}_{0:n-1}, y_{0:n-1}) &\sim \mathcal{N}(\widehat{\mathbf{x}}^{(s)}_{n|n-1}(r_n), \widehat{\mathbf{P}}^{(s)}_{n|n-1}(r_n)), \\ p(y_n|r_n, r^{(s)}_{0:n-1}, y_{0:n-1}) &\sim \mathcal{N}(\widehat{y}^{(s)}_{n|n-1}(r_n), \widehat{P}^{y(s)}_{n|n-1}(r_n)). \end{aligned}$$

Then $\{r^{(s)}_n, w^{(s)}_n\}_{s=1}^{S}$ *and* $(\widehat{\mathbf{x}}^{(s)}_{n|n}, \widehat{\mathbf{P}}^{(s)}_{n|n})$ *can be propagated by the following steps:*

- *Conditional prediction. For* $s = 1, \cdots, S$ *and* $r_n \in \Omega$,

$$\begin{aligned} \widehat{\mathbf{x}}^{(s)}_{n|n-1}(r_n) &= \mathbf{F}_{n-1}(r_n, r^{(s)}_{n-1})\widehat{\mathbf{x}}^{(s)}_{n-1|n-1} + \mathbf{b}_{n-1}(r_n, r^{(s)}_{n-1}) \\ \widehat{\mathbf{P}}^{(s)}_{n|n-1}(r_n) &= \mathbf{F}_{n-1}(r_n, r^{(s)}_{n-1})\widehat{\mathbf{P}}^{(s)}_{n-1|n-1}(\mathbf{F}_{n-1}(r_n, r^{(s)}_{n-1}))^T + \mathbf{Q}_{n-1}(r_n) \end{aligned}$$

- *Conditional Innovation. For* $s = 1, \cdots, S$ *and* $r_n \in \Omega$,

$$\begin{aligned} \widehat{y}^{(s)}_{n|n-1}(r_n) &= \mathbf{h}\widehat{\mathbf{x}}^{(s)}_{n|n-1}(r_n) \\ \widehat{P}^{y(s)}_{n|n-1}(r_n) &= \mathbf{h}\widehat{\mathbf{P}}^{(s)}_{n|n-1}(r_n)\mathbf{h}^T + v^y_n \end{aligned}$$

- *Filtering of* $r_{0:n}$ *(PF). Compute the discrete distribution* $\{r^{(s)}_{0:n}, w^{(s)}_n\}_{s=1}^{S}$ *by PF [4].*
- *Conditional Filtering. For* $s = 1, \cdots, S$,

$$\begin{aligned} \mathbf{K}^{(s)}_n &= \widehat{\mathbf{P}}^{(s)}_{n|n-1}(r^{(s)}_n)\mathbf{h}^T(\widehat{P}^{y(s)}_{n|n-1}(r^{(s)}_n))^{-1} \\ \widehat{\mathbf{x}}^{(s)}_{n|n} &= \widehat{\mathbf{x}}^{(s)}_{n|n-1}(r^{(s)}_n) + \mathbf{K}^{(s)}_n[y_n - \widehat{y}^{(s)}_{n|n-1}(r^{(s)}_n)] \\ \widehat{\mathbf{P}}^{(s)}_{n|n} &= \widehat{\mathbf{P}}^{(s)}_{n|n-1}(r^{(s)}_n) - \mathbf{K}^{(s)}_n\mathbf{h}\widehat{\mathbf{P}}^{(s)}_{n|n-1}(r^{(s)}_n) \end{aligned}$$

We eventually approximated the CM $\widehat{\mathbf{x}}_{n|n}$ *and the MAP* $\widehat{r}_{n|n}$ *by (20) and (17) respectively.*

SIMULATIONS

Two experiences are performed in this section: a simulated 1-D model and a real images separation example.

• A simulated 1-D model. The aim is to separate the signals b, f and r plotted of in Fig. 1. The parameters considered in this example are $r_n \in \{0,1\}$, $p(r_0 = 0) = p(r_0 = 1) = .5$, $\{p(r_n = k|r_{n-1} = k')\}_{(k,k')=(0,0)}^{(1,1)} = \begin{bmatrix} .9 & .1 \\ .2 & .8 \end{bmatrix}$, $a^f_n = .97$, $m^f_n(k) = 2 \times \delta(k-1)$, $v^f_n(k) = .01$, $a^b_n = .1$, $v^b_n = .1$, $v^y_n = .05$. The particles number is $S = 300$ in the PF algorithm. The method's performance is measured by the empirical standard deviation (std) for b_n and f_n:

$$std = \sqrt{\frac{1}{128}\sum_{n=0}^{127} \mid i_n - \widehat{i}_{n|n} \mid^2}, \quad i = b, f,$$

and by the binary error rate (BER) in % for r_n:

$$BER(\%) = \frac{100}{128} \times \sum_{n=0}^{127} \delta(r_n - \widehat{r}_{n|n}).$$

Our algorithm gives good results as we can see in Fig. 2.

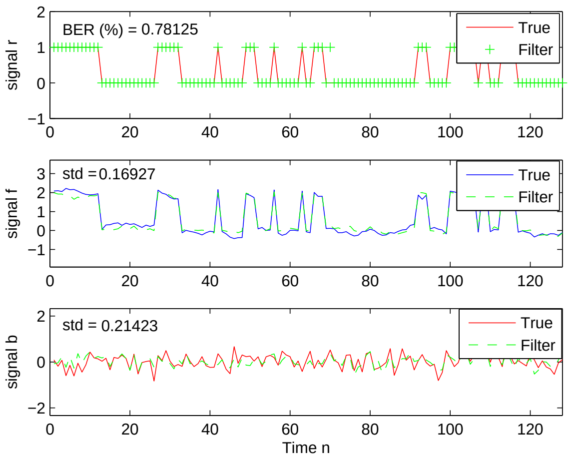

FIGURE 2. Separation of the 1-D signals of Fig. 1.

• Real images separation. From an $n_r \times n_c$ observed image Y (Fig. 3), the aim is to restore 3 hidden images: a background B, a foreground F and a discrete image R. Indeed, the idea is to consider the image Y as a concatenation of n_r 1-D signals $y^{row} = y_{1:n_c}$; restoring the images B, F and R from Y amounts then to restoring 3 1-D signals b^{row}, f^{row} and r^{row} from y^{row} for $row = 1, \cdots, n_r$. The algorithm is initialized by the following parameters[1]: $p(b_{n+1}|b_n) \sim \mathcal{N}(.5b_n, 2.5)$; $p(f_{n+1}|f_n, r_{n+1} = k, r_n = k') \sim \mathcal{N}(.99\delta(k-k')f_n + (1-\delta(k-k'))k, 2.5)$ and $p(y_n|b_n, f_n) \sim \mathcal{N}(b_n + f_n, .7)$. The particles number in the PF algorithm is $S = 10$. As we can see in Figs. 3 and 4 the foreground and the background are well separated. Finally, let us remark that one can also adopt the same idea by processing Y column by column rather than row by row.

---

[1] The parameters of the MC r can be computed from Y by using the K-means algorithm (among other methods).

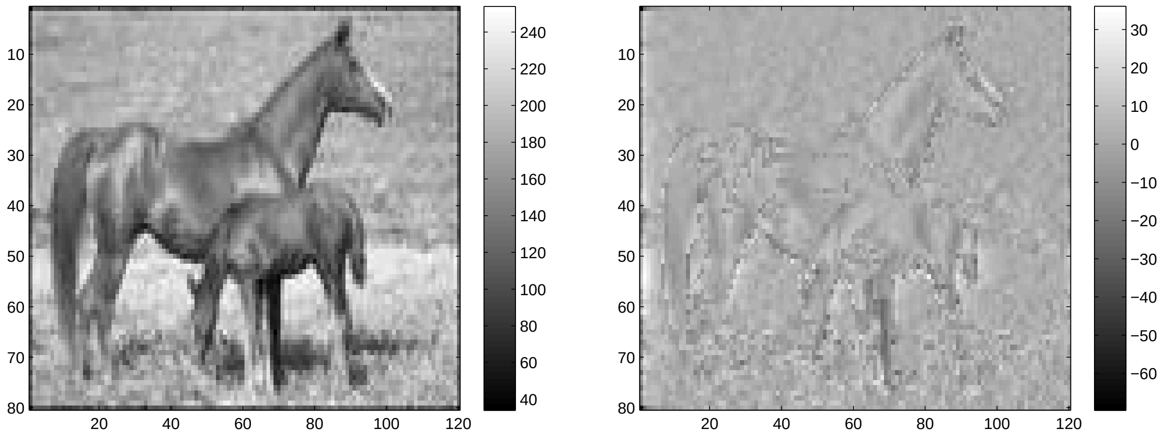

FIGURE 3. The observed image Y (left) and the background restored image B (right).

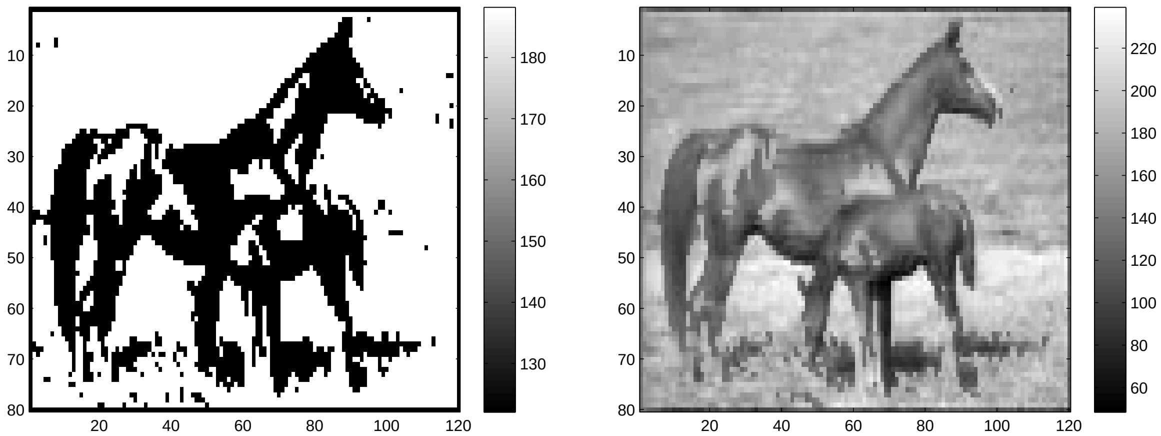

FIGURE 4. The discrete restored image R (left) and the foreground restored image F (right).

REFERENCES

1. J. Vermaak, C. Andrieu, A. Doucet, and S. Godsill, *IEEE Trans. on Speech and Audio Processing* **10**, 173– 85 (2002).
2. D. Angelova, and L. Mihaylova, *IEEE Trans. on Signal Processing* **56**, 825–32 (2008).
3. R. E. Kalman, *J. Basic Eng., Trans. ASME, Series D* **82**, 35–45 (1960).
4. A. Doucet, N. de Freitas, and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*, Statistics for Engineering and Information Science, Springer Verlag, New York, 2001.
5. R. Chen, and S. Liu, *J. Royal Stat. Soc. B* (2000).