

УДК 621.391; 681.327.12.001.362

И. С. Грузман, И. В. Курилин

(Новосибирск)

**НЕЛИНЕЙНЫЙ АЛГОРИТМ ФИЛЬТРАЦИИ
КУСОЧНО-ПОСТОЯННЫХ МНОГОУРОВНЕВЫХ СИГНАЛОВ ***

Для марковской модели строчных реализаций изображений, представляющей собой аддитивную смесь кусочно-постоянного многоуровневого сигнала и гауссовского шума, синтезирован нелинейный квазиоптимальный рекуррентный алгоритм фильтрации. Его особенностью является то, что гауссовская аппроксимация используется лишь для «части» апостериорного распределения, которое в целом остается негауссовским. Анализ эффективности полученного алгоритма показал, что он практически обеспечивает потенциально достижимую точность оценивания и существенно лучше линейной каузальной оценки, представляющей собой скользящее среднее.

При решении задач выделения контуров, сжатия и сегментации используются негауссовские модели изображений, описывающие характерные свойства контурных линий, например, резкие скачки (перепады) яркости [1–3]. Идеальные перепады яркости можно рассматривать как сигналы, имеющие форму ступеньки случайной величины. Поэтому для описания двумерной функции яркости используем кусочно-постоянное многоуровневое поле. Это поле состоит из областей, внутри которых яркость постоянна и меняется скачком на случайную величину при переходе от одной области к другой. Контуром является линия, разделяющая области с постоянной яркостью. Наблюдаемое изображение представляет собой аддитивную смесь кусочно-постоянного поля и стационарного гауссовского шума, который описывает в модели фон.

Рассмотрим синтез одномерного рекуррентного алгоритма фильтрации кусочно-постоянного сигнала, наблюдаемого на фоне аддитивного шума. Такой алгоритм может быть использован, например, при построении двухэтапных процедур обработки сигналов [4] или изображений [5–7]. Итак, строчные реализации функции яркости модели изображения Z представляют собой сумму двух компонент:

$$Z = \{z_i = y_i + u_i, i = \overline{1, I}\}, \quad (1)$$

* Работа выполнена при поддержке Министерства образования РФ (грант № Г00-4.2-7).

где $Y = \{y_i, i = \overline{1, I}\}$ – кусочно-постоянный многоуровневый сигнал; $U = \{u_i, i = \overline{1, I}\}$ – гауссовский дискретный белый шум с априорно известной дисперсией D_u и нулевым математическим ожиданием: $u_i \in N(0, D_u)$; I – число отсчетов в строке.

Здесь и далее гауссовское распределение будем обозначать $N(m, D)$, где первый аргумент соответствует математическому ожиданию m , а второй – дисперсии D .

Конкретизируем модель кусочно-постоянного многоуровневого сигнала Y . Пусть сигнал Y задан нелинейным разностным уравнением

$$Y = \begin{cases} y_1 = v_1; \\ y_i = y_{i-1}(1 - x_i) + x_i v_i, \quad i = \overline{2, I} \end{cases}, \quad (2)$$

где $X = \{x_i, i = \overline{1, I}\}$ – однородный стационарный бинарный марковский сигнал, принимающий значение 0 или 1; $V = \{v_i, i = \overline{1, I}\}$ – гауссовский дискретный белый шум: $v_i \in N(0, D_v)$. Из (2) следует, что сигнал Y сохраняет постоянный уровень до тех пор, пока сигнал X равен 0. Если сигнал X принимает значение 1, то значение сигнала Y меняется скачком на случайную величину. Сигнал X , определяющий протяженность областей постоянного уровня, задан одношаговой матрицей перехода [8]

$$\pi_x(x_i/x_{i-1}) = \begin{bmatrix} \pi_x(0/0) & \pi_x(0/1) \\ \pi_x(1/0) & \pi_x(1/1) \end{bmatrix} \quad (3)$$

и вероятностями стационарного состояния

$$p(1) = \frac{\pi_x(1/0)}{\pi_x(0/1) + \pi_x(1/0)}, \quad p(0) = 1 - p(1),$$

где $p(1)$ и $p(0)$ – вероятности того, что бинарный сигнал X принимает соответственно значения 1 и 0. Поскольку протяженность однородных областей на изображениях достаточно велика, то для принятой модели (2) вероятность появления 0 должна быть много больше вероятности появления 1. Поэтому скачкообразное изменение уровня сигнала Y является относительно редким событием. Положим, что $\pi_x(0/0) \gg \pi_x(1/0)$ и $\pi_x(0/1) \gg \pi_x(1/1)$. В частности, если вероятность перехода $\pi_x(1/1)$ принять равной 0, то сигнал Y не может 2 раза подряд изменять свое значение.

Байесовская каузальная оценка текущего отсчета \hat{y}_i по наблюдаемым отсчетам $Z_i = \{z_1, z_2, \dots, z_i\}$ сигнала Z равна апостериорному среднему:

$$\hat{y}_i = \langle y_i / Z_i \rangle = \int_{-\infty}^{\infty} y_i w(y_i / Z_i) dy_i, \quad (4)$$

где $\langle \cdot \rangle$ – символ математического ожидания; $w(y_i / Z_i)$ – апостериорное распределение вероятностей (АРВ) отсчета y_i .

Чтобы найти байесовскую оценку (4), необходимо определить АРВ $w(y_i/Z_i)$. Для этого воспользуемся тем, что векторный процесс (X, Y, Z) является марковским. Поэтому совместное АРВ $w(y_i, x_i/Z_i)$ можно вычислить с помощью рекуррентных соотношений [9], а затем усреднить его по «мешающей» компоненте x_i :

$$w(y_i/Z_i) = \sum_{x_i=0}^1 w(y_i, x_i/Z_i). \quad (5)$$

Здесь учтено, что компонента x_i является дискретной (бинарной) случайной величиной. Рекуррентные соотношения для вычисления $w(y_i, x_i/Z_i)$ имеют вид

$$w(y_1, x_1/Z_1) = \frac{1}{C_1} w(y_1, x_1, z_1); \quad (6)$$

$$w(y_i, x_i/Z_i) = \frac{1}{C_i} \sum_{x_{i-1}=0}^1 \int_{-\infty}^{\infty} w(y_{i-1}, x_{i-1}/Z_{i-1}) \times \\ \times w(y_i, x_i, z_i/y_{i-1}, x_{i-1}, z_{i-1}) dy_{i-1}, \quad (7)$$

где $\{C_i, i=\overline{1, I}\}$ – нормирующие константы, определяемые из условия

$$\sum_{x_i=0}^1 \int_{-\infty}^{\infty} w(y_i, x_i/Z_i) dy_i = 1; \quad (8)$$

$w(x_1, y_1, z_1)$ – совместное распределение первого отсчета векторного процесса (X, Y, Z) ; $w(x_i, y_i, z_i/x_{i-1}, y_{i-1}, z_{i-1})$ – плотность вероятности перехода процесса (X, Y, Z) за один шаг из состояния $(x_{i-1}, y_{i-1}, z_{i-1})$ в состояние (x_i, y_i, z_i) ; $Z_1 = \{z_1\}$ – первый отсчет наблюдаемого сигнала (1).

Определим в явном виде АРВ (6). Поскольку распределение y_1 и z_1 не зависит от значения x_1 (см. (2)), то с учетом соотношения (1) АРВ

$$w(y_1, x_1 = k/z_1) = \frac{w(z_1/y_1)w(y_1)p(x_1 = k)}{C_1} = \\ = \frac{N(z_1 - y_1, D_u)N(0, D_v)p(x_1 = k)}{C_1}, \quad k = \overline{0, 1}. \quad (9)$$

Выражение (9) позволяет записать АРВ $w(y_1, x_1/Z_1)$ в виде

$$w(y_1, x_1 = k/Z_1) = \frac{C_1^{(k)}}{C_1} N(m_1, D_1), \quad k = \overline{0, 1}, \quad (10)$$

где $D_1 = ((D_v)^{-1} + (D_u)^{-1})^{-1}$; $m_1 = \frac{D_1}{D_u} z_1$; $C_1^{(k)} = p(x_1 = k) \varphi(z_1, 0, D_v + D_u)$;

$C_1 = C_1^{(0)} + C_1^{(1)}$. Нелинейная функция $\varphi(z_i, m, D) = \frac{1}{\sqrt{2\pi D}} \exp\left(-\frac{(z_i - m)^2}{2D}\right)$ от

значений z_i соответствует гауссовской кривой с параметрами m и D . Выполнив усреднение (9) по x_1 , получим, что АРВ $w(y_1/Z_1)$ является гауссовским $N(m_1, D_1)$, а байесовская оценка первого отсчета y_1 равна m_1 , т. е. линейна относительно наблюдения z_1 . Этот результат является тривиальным, так как в соответствии с принятой моделью случайные величины y_1 и z_1 являются совместно гауссовскими и не зависят от x_1 .

Выразим плотность вероятности перехода через условные распределения:

$$w(x_i, y_i, z_i/x_{i-1}, y_{i-1}, z_{i-1}) = w(z_i/x_i, y_i, x_{i-1}, y_{i-1}, z_{i-1}) \times \\ \times w(y_i/x_i, x_{i-1}, y_{i-1}, z_{i-1}) p(x_i/x_{i-1}, y_{i-1}, z_{i-1}). \quad (11)$$

С учетом независимости отсчетов белого шума и соотношения (1) первый сомножитель в (11) имеет вид

$$w(z_i/x_i, y_i, x_{i-1}, y_{i-1}, z_{i-1}) = w(z_i/y_i) = N(z_i - y_i, D_u).$$

Второй сомножитель в (11) полностью определяется нелинейным разностным уравнением (2):

$$w(y_i/x_i, x_{i-1}, y_{i-1}, z_{i-1}) = w(y_i/x_i, y_{i-1}) = \begin{cases} \delta(y_i - y_{i-1}) & \text{при } x_i = 0; \\ N(0, D_v) & \text{при } x_i = 1. \end{cases}$$

Поскольку вероятность значения x_i бинарного марковского сигнала X не зависит от величин y_{i-1} и z_{i-1} , то третий сомножитель в (11) равен одношаговой матрице перехода (3):

$$p(x_i/x_{i-1}, y_{i-1}, z_{i-1}) = p(x_i/x_{i-1}) = \pi_x(x_i/x_{i-1}).$$

Отсюда получим АРВ $w(y_i, x_i/Z_i)$ для второго шага:

$$w(y_2, x_2 = k/Z_2) = \frac{C_2^{(k)}}{C_2} N(m_2^{(k)}, D_2^{(k)}), \quad k = \overline{0, 1}, \quad (12)$$

где $D_2^{(1)} = D_1$; $D_2^{(0)} = ((D_1)^{-1} + (D_u)^{-1})^{-1}$; $m_2^{(1)} = \frac{D_1}{D_u} z_2$; $m_2^{(0)} = D_2^{(0)} \left(\frac{m_1}{D_1} + \frac{z_2}{D_u} \right)$;

$$C_2^{(1)} = \frac{\varphi(z_2, 0, D_u + D_v)}{C_1} \sum_{l=0}^1 C_1^{(l)} \pi_x(1/l); \\ C_2^{(0)} = \frac{\varphi(z_2, m_1, D_1 + D_u)}{C_1} \sum_{l=0}^1 C_1^{(l)} \pi_x(0/l); \quad (13)$$

$$C_2 = C_2^{(0)} + C_2^{(1)}. \quad (14)$$

Используя равенства (12) с учетом (4), легко найти байесовскую оценку

$$\hat{y}_2 = \frac{1}{C_2} \sum_{k=0}^1 C_2^{(k)} m_2^{(k)}.$$

Из этого соотношения следует, что оценка сигнала Y на втором шаге нелинейная, так как весовые коэффициенты, вычисляемые по формулам (13) и (14), являются нелинейными функциями от значений наблюдаемого сигнала Z .

Из нелинейного разностного уравнения (2) и полученных соотношений для вычисления АРВ на первом и втором шагах следует, что

$$w(y_i, x_i = 1/Z_i) = w(y_i/z_i, x_i = 1) p(x_i = 1/Z_i), \quad (15)$$

где $w(y_i/z_i, x_i = 1) = N(m_i^{(1)}, D_1)$ – гауссовское распределение вероятностей

с математическим ожиданием $m_i^{(1)} = \frac{D_1}{D_u} z_i$; $p(x_i = 1/Z_i) = \frac{C_i^{(1)}}{C_i}$ – апостериор-

ная вероятность того, что i -й отсчет бинарного сигнала X равен 1. Легко убедиться, что весовой коэффициент $C_i^{(1)}$ определяется рекуррентным соотношением:

$$C_i^{(1)} = \frac{\varphi(z_i, 0, D_u + D_v)}{C_{i-1}} \sum_{l=0}^1 C_{i-1}^{(l)} \pi_x(1/l), \quad (16)$$

где начальные значения $C_1^{(l)}$, $l = \overline{0,1}$, вычисляются по формуле (10). К сожалению, не удастся получить соотношения для вычисления АРВ $w(y_i, x_i/Z_i)$ при $x_i = 0$ и коэффициента $C_i^{(0)}$, аналогичные (12)–(14), так как уже на третьем шаге АРВ

$$w(y_3, x_3 = 0/Z_3) = \frac{1}{C_3} \sum_{l=0}^1 C_3^{(0,l)} N(m_3^{(0,l)}, D_3^{(0,l)}), \quad (17)$$

т. е. представляет собой взвешенную сумму двух гауссовских распределений. Здесь

$$D_3^{(0,l)} = ((D_2^{(l)})^{-1} + (D_u)^{-1})^{-1}; \quad m_3^{(0,l)} = D_2^{(0,l)} \left(\frac{m_2^{(l)}}{D_1} + \frac{z_3}{D_u} \right);$$

$$C_3^{(0,l)} = \frac{C_2^{(l)}}{C_2} \varphi(z_3, m_2^{(l)}, D_2^{(l)} + D_u) \pi_x(0/l), \quad l = \overline{0,1}.$$

Соотношение (17) нетрудно получить, подставив (12)–(14) в рекуррентное соотношение (7). На каждом последующем шаге число слагаемых в формуле

для вычисления АРВ $w(y_i, x_i/Z_i)$ при $x_i = 0$ будет удваиваться. По этой причине оптимальный алгоритм оценки сигнала Y будет практически нереализуемым. Чтобы вычислительная сложность алгоритма не возрастала от шага к шагу, необходимо, чтобы апостериорные распределения на третьем и последующем шагах имели такую же структуру, как и АРВ на втором шаге. Для этого аппроксимируем (17) следующим выражением:

$$w(y_3, x_3 = 0/Z_3) = \frac{C_3^{(0)}}{C_3} N(m_3^{(0)}, D_3^{(0)}),$$

где весовой коэффициент $C_3^{(0)} = \sum_{l=0}^1 C_3^{(0,l)}$, а математическое ожидание $m_3^{(0)}$ и дисперсия $D_3^{(0)}$ определяются из условия равенства первых двух моментов гауссовского распределения $N(m_3^{(0)}, D_3^{(0)})$ и нормированного на $\frac{C_3}{C_3^{(0)}}$ АРВ $w(y_3, x_3 = 0/Z_3)$:

$$m_3^{(0)} = \frac{1}{C_3^{(0)}} \sum_{l=0}^1 C_3^{(0,l)} m_3^{(0,l)};$$

$$D_3^{(0)} = \frac{1}{C_3^{(0)}} \sum_{l=0}^1 C_3^{(0,l)} ((m_3^{(0,l)})^2 + D_3^{(0,l)}) - (m_3^{(0)})^2.$$

Здесь коэффициент нормировки $\frac{C_3}{C_3^{(0)}}$ выбирался из условия

$$\int_{-\infty}^{\infty} w(y_3, x_3 = 0/Z_3) dy_3 = 1.$$

Выполняя аналогичную аппроксимацию на каждом последующем шаге, легко показать, что АРВ $w(y_i, x_i/Z_i)$ для $i=3, 4, \dots$ определяется следующим приближенным выражением:

$$w(y_i, x_i = k/Z_i) = \frac{C_i^{(k)}}{C_i} N(m_i^{(k)}, D_i^{(k)}), \quad k = \overline{0, 1},$$

где

$$D_i^{(0,l)} = ((D_{i-1}^{(l)})^{-1} + (D_u)^{-1})^{-1}; \quad m_i^{(0,l)} = D_i^{(0,l)} \left(\frac{m_{i-1}^{(0,l)}}{D_{i-1}^{(l)}} + \frac{z_i}{D_u} \right);$$

$$C_i^{(0,l)} = \frac{C_{i-1}^{(l)}}{C_{i-1}} \varphi(z_i, m_{i-1}^{(l)}, D_{i-1}^{(l)} + D_u) \pi_x(0/l); \quad C_i^{(0)} = \sum_{l=0}^1 C_i^{(0,l)};$$

$$m_i^{(0)} = \frac{1}{C_i^{(0)}} \sum_{l=0}^1 C_i^{(0,l)} m_i^{(0,l)}; \quad D_i^{(0)} = \frac{1}{C_i^{(0)}} \sum_{l=0}^1 C_i^{(0,l)} ((m_i^{(0,l)})^2 + D_i^{(0,l)}) - (m_i^{(0)})^2.$$

АРВ для $k=1$ вычисляется по формуле (15), а нормирующая константа $C_i = C_i^{(0)} + C_i^{(1)}$. Выражения для вычисления АРВ $w(y_i, x_i/Z_i)$ на первом и втором шагах приведены выше. Байесовская оценка сигнала Y на i -м шаге с учетом (5) определяется следующим соотношением:

$$\hat{y}_i = \frac{1}{C_i} \sum_{k=0}^1 C_i^{(k)} m_i^{(k)}. \quad (18)$$

Таким образом, получен нелинейный рекуррентный алгоритм фильтрации кусочно-постоянного многоуровневого сигнала, наблюдаемого на фоне аддитивного шума. Его особенностью является то, что гауссовская аппроксимация используется лишь для части апостериорного распределения (при $x_i = 0$). В целом апостериорное распределение остается негауссовским. Алгоритм является квазиоптимальным, поэтому существует вопрос определения потенциально достижимой точности оценки сигнала Y . Для ее вычисления положим, что моменты скачкообразного изменения уровня сигнала Y заранее известны. Тогда оценке подлежат уровни, которые принимает сигнал Y . Среднеквадратическая ошибка рекуррентной каузальной оценки сигнала постоянного уровня, наблюдаемого на фоне аддитивного белого гауссовского шума, определяется рекуррентными соотношениями [10]:

$$D_{\varepsilon,1} = ((D_v)^{-1} + (D_u)^{-1})^{-1}; \quad D_{\varepsilon,m} = ((D_{\varepsilon,m-1})^{-1} + (D_u)^{-1})^{-1}, \quad m = \overline{2, M},$$

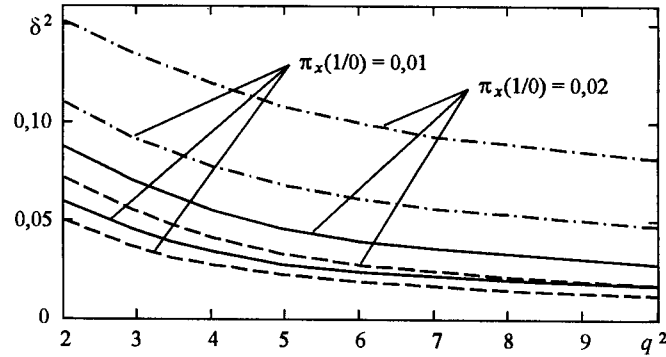
где M – длительность интервала, на котором сигнал Y сохраняет постоянное значение. Среднее значение дисперсии ошибки на интервале длительностью M определяется соотношением:

$$D_{\varepsilon,M} = \frac{1}{M} \sum_{m=1}^M D_{\varepsilon,m}.$$

В соответствии с моделью (2) начало интервала определяется моментом появления единицы в бинарном сигнале X , а его длительность – числом нулей между двумя единицами. Параметр M является случайной величиной, распределение которой полностью определяется матрицей перехода (3) однородного стационарного бинарного сигнала X :

$$P(M=1) = \frac{p(1)\pi_x(1/1)}{\sum_{L=1}^{\infty} P(M=L)}; \quad (19)$$

$$P(M=J) = \frac{p(1)(\pi_x(0/1))^{J-1} \pi_x(1/0)}{\sum_{L=1}^{\infty} P(M=L)}, \quad J = \overline{2, \infty}.$$



Зависимости нормированной среднеквадратической ошибки оценивания от отношения сигнал/шум

С учетом (19) относительная среднеквадратическая ошибка каузальной рекуррентной оценки сигнала Y при известных моментах скачкообразного изменения уровня, усредненная по M , имеет вид

$$\delta^2 = \frac{\langle D_{\epsilon, M} \rangle}{D_v} = \frac{\sum_{J=1}^{\infty} P(M=J) \left(\frac{1}{M} \sum_{m=1}^M D_{\epsilon, m} \right)}{D_v}. \quad (20)$$

Очевидно, что δ^2 является границей снизу для полученного алгоритма фильтрации.

На рисунке приведены зависимости относительной среднеквадратической ошибки фильтрации сигнала Y от отношения сигнал/шум $q^2 = \frac{D_v}{D_u}$ для

двух значений условных вероятностей $\pi_x(1/0) = \pi_x(1/1)$. Сплошные линии соответствуют нелинейной квазиоптимальной оценке (18), штрихпунктирные – линейной каузальной оценке, представляющей собой скользящее среднее

$$\tilde{y}_i = \frac{1}{K+1} \sum_{k=0}^K z_{i-k}, \quad (21)$$

где параметр K (размер каузального окна) выбирался из условия минимума среднеквадратической ошибки фильтрации. Эти зависимости были получены методом математического моделирования на ЭВМ алгоритмов (18), (21) с использованием модели строчных реализаций (1), (2). Штриховые линии на рисунке соответствуют границе снизу (20).

Ухудшение качества оценки сигнала Y при увеличении в 2 раза переходной вероятности $\pi_x(1/0)$ объясняется тем, что соответственно в 2 раза возрастает частота изменения его уровня. Из приведенных зависимостей следует, что квазиоптимальная нелинейная оценка (18) кусочно-постоянного многоуровневого сигнала Y существенно лучше линейной (21). Вместе с тем она обладает практически потенциально достижимой точностью оценивания.

СПИСОК ЛИТЕРАТУРЫ

1. **Обработка изображений и цифровая фильтрация** /Под ред. Т. Хуанга: Пер. с англ. М.: Мир, 1979.
2. **Прэтт У.** Цифровая обработка изображений: Пер. с англ. /Под ред. Д. С. Лебедева. М.: Мир, 1982. Кн. 2.
3. **Виттих В. А., Сергеев В. В., Сойфер В. А.** Обработка изображений в автоматизированных системах научных исследований. М.: Наука, 1982.
4. **Грузман И. С., Спектор А. А.** Применение свойства условной независимости для симметричного сглаживания марковских процессов // Радиотехн. и электроника. 1997. № 6. С. 701.
5. **Грузман И. С., Микерин В. И., Спектор А. А.** Двумерная фильтрация изображений на основе использования ограниченных данных // Радиотехн. и электроника. 1995. № 5. С. 817.
6. **Грузман И. С.** Применение узкополосных случайных полей в задаче выделения папиллярных линий на изображениях // Автометрия. 1997. № 4. С. 102.
7. **Грузман И. С.** Двухэтапная фильтрация бинарных изображений // Автометрия. 1999. № 3. С. 31.
8. **Тихонов В. И., Миронов М. А.** Марковские процессы. М.: Сов. радио, 1977.
9. **Казаков В. А.** Введение в теорию марковских процессов и некоторые радиотехнические задачи. М.: Сов. радио, 1973.
10. **Сейдж Э., Мелс Дж.** Теория оценивания и ее применение в связи и управлении. М.: Связь, 1976.

*Новосибирский государственный
технический университет,
E-mail: guzman@ktor.ref.nstu.ru*

*Поступила в редакцию
30 марта 2001 г.*