Computing Joint Distributions of 2D Moving Median Filters With Applications to Detection of Edges

Dawei Huang, *Member, IEEE*, and William T.M. Dunsmuir

Abstract—This paper derives the joint distribution of medians over moving windows in a two dimensional noisy image. The general formulation presented allows derivation of the probability distribution, needed to evaluate the probability of failing to detect an edge when present ("edge miss probability") and the probability of falsely detecting a nonexistent edge.

Index Terms—Image analysis, moving medians, joint distributions, edge detection, edge miss probability.

1 Introduction

MEDIAN filters are now widely used in image processing because they are effective at smoothing out noise while preserving various important root signals which frequently occur in typical two dimensional images. There are many types of median filters which can be considered. In particular, the choice of pixel window over which the filter is defined provides a variety of filters for different applications. In this paper, we will consider pixel windows defined on the square, containing an odd number of pixels, on the *X*-shaped window, defined only on the diagonal elements of the square, and, on the cross-shaped window, which uses pixels aligned vertically and horizontally. In [1], these are defined as SQUARE($2L + 1$), CROSS($2L + 1$), and XSHAPE($2L + 1$). These shapes provide differing abilities for preserving vertically, horizontally, or diagonally aligned edges in an image. Additional flexibility can be introduced to median filters through use of nonuniform weights at different pixels in the window. For example, exponentially decaying weights can be applied as in [3]. In this paper, we will consider exclusively the case of equally weighted medians, in which case the median filter is defined simply as the median of the data observed at the pixels in the window.

When prefiltering a two-dimensional image by moving median filters of particular interest is the probability of failing to detect a real edge in the original, noise-free image ("edge miss probability"). Of equal interest is the probability of falsely concluding that an edge exists (detecting a "pseudoedge"). These probabilities can be obtained from the joint distribution of medians calculated from data in two overlapping windows centered one pixel apart. Derivation of this distribution is complicated by the fact that the data in each of the two windows is not sampled from the same distribution. The derivation of this joint distribution was considered to be "an extremely complex problem" (see [1]). This paper solves this problem by providing a derivation of the required distributions in a series of theorems—see Section 2.

Let (i, j) be the coordinates of a pixel, where i refers to the rows (vertical direction) and j references to the columns (hori-

- *D. Huang is with the School of Mathematics, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia. E-mail: huang@fsc.qut.edu.au.*
- *W.T.M. Dunsmuir is with the School of Mathematics, University of New South Wales, Sydney, NSW 2052, Australia. E-mail: W.Dunsmuir@unsw.edu.au.*

Manuscript received 19 Dec. 1996. Recommended for acceptance by J. Daugman. For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 106154.

zontal direction). We consider the situation in which there is a vertical edge located between pixels $(i, 0)$ and $(i, 1)$ for any i.

In order to derive the required joint distribution, various windows of pixels need to be defined. For brevity, we will do this for the three basic window shapes considered in [1].

For the 3×3 square window, define

$$W_j = \{(i, j);\ \ i = -1, 0, 1\} \quad \text{for} \quad j = -1, 0, 1, 2.$$

Then the 3×3 square centered of $(0, 0)$ is $W_1 \cup W_2 \cup W_3$ and that centered at $(0, 1)$ is $W_2 \cup W_3 \cup W_4$.

For the 5×5 cross-shaped window, define

$$W_1 = \{(0, -2), (-2, 0), (-1, 0), (1, 0), (2, 0)\}$$

$$W_2 = \{(0, -1), (0, 0)\}$$

$$W_3 = \{(0, 1), (0, 2)\}$$

$$W_4 = \{(-2, 1), (-1, 1), (1, 1), (2, 1), (0, 3)\}.$$

The cross-shaped window centered at $(0, 0)$ is $W_1 \cup W_2 \cup W_3$ and that centered at $(0, 1)$ is $W_2 \cup W_3 \cup W_4$.

For the 5×5 *X*-shaped window, define

$$W_1 = \{(-2, -2), (-1, -1), (0, 0), (1, -1), (2, -2)\}$$

$$W_2 = \{(-2, -1), (-1, 0), (1, 0), (2, -1)\}$$

$$W_3 = \{(-2, 2), (-1, 1), (1, 1), (2, 2)\}$$

$$W_4 = \{(-2, 3), (-1, 2), (0, 1), (1, 2), (2, 3)\}.$$

Then the *X*-shaped window centered at $(0, 0)$ is $W_1 \cup W_3$ and that centered at $(0, 1)$ is $W_2 \cup W_4$.

Assume that the image has a gray level equal to zero on the left side and level $= h$ on the right side of the edge, and the image is contaminated by noise N_{ij} at pixel location (i, j). Let M_1 be the median on the window positioned to the left of the edge and M_2 be the window positioned to the right of the edge. Then, for the SQUARE and the CROSS pixel windows, we have:

$$M_1 = \text{median}\{z_{ij} : (i, j) \in W_1 \cup W_2 \cup W_3\},$$

$$M_2 = \text{median}\{z_{ij} : (i, j) \in W_2 \cup W_3 \cup W_4\},$$

where $z_{ij} = N_{ij}$ for $(i, j) \in W_1 \cup W_2$, and $z_{ij} = h + N_{ij}$ for $(i, j) \in W_3 \cup W_4$. The situation is slightly simpler for the XSHAPE, because

$$M_1 = \text{median}\{z_{ij} : (i, j) \in W_1 \cup W_3\},$$

$$M_2 = \text{median}\{z_{ij} : (i, j) \in W_2 \cup W_4\},$$

and the regions on which M_1 and M_2 are defined are nonoverlapping and, hence, are based on independent samples.

In the three examples given above, the medians are based on data observed at nine pixels and, therefore, require similar computational effort. However, it is to be expected that there are clear differences in their performance with respect to the probabilities of signaling the existence of an edge when one is truly not present (Type I error in the language of statistical hypothesis testing) and their power at detecting an edge when actually present. Section 3 compares the power in detecting an edge of height h for the three window shapes considered above and for which the Type I error probability is fixed at .05.

2 Joint Distribution on Overlapping Windows

The distribution of the equally weighted median for independent and identically distributed observations $z_{i,j}$ at pixel (i, j) is easily obtained from [2] for example. In the situation where there is a single edge crossing the pixel region, the median of the data equals:

$$M = \text{median}\{z_{ij} : z_{ij} = N_{ij}, (i, j) \in R_1; \qquad z_{ij} = h + N_{ij}, (i, j) \in R_2\} \quad (1)$$

where $R_1 \cup R_2 = R$, and R contains $2m - 1$ pixels. For the square centered at $(0, 0) R_1 = W_1 \cup W_2$ and $R_2 = W_3$. Calculating the distribution of M in this case is no longer straightforward, because the random variables in the set $\{z_{ij}, (i, j) \in R\}$ are not identically distributed. The required result is given in Theorem 1 below.

To study the statistical properties of the various median filters introduced above, the *joint* distribution of two medians, M_1 and M_2, defined on possibly overlapping pixel regions is required. Although the 1D case was studied in detail in [1], the more difficult 2D problem described in Section 1 was not pursued there. Theorem 4 provides the required 2D result.

In [1], it is suggested that the case where the $\{N_{ij}\}$ are dependent should be more relevant in image processing. However, "there has been no work yet advanced for analyzing order statistics arising from overlapping samples of dependent data" [1, p. 189]. Some progress on this problem is made in Theorem 4 below.

Before deriving the joint distribution required, we first derive the distribution of the rth largest value of the z_{ij} when they are drawn from a general multivariate distribution (i.e., z_{ij} which are neither identically nor independently distributed).

THEOREM 1. *Suppose that there are m_1 pixels in W_1 and m_2 pixels in W_2. Let x_i, y_i and T_i be the ith smallest element of $\{z_{ij}, (i, j) \in W_1\}$, $\{z_{ij}, (i, j) \in W_2\}$, and $\{z_{ij}, (i, j) \in W_1 \cup W_2\}$, respectively. Let $x_i = y_i = T_i = -\infty$, $i \le 0$, and $x_i = \infty$, $i > m_1$, $y_i = \infty$, $i > m_2$. Then, no matter what joint distribution the $\{z_{ij}\}$ obey,*

$$P\{T_r \le a\} = \sum_{j=0}^{m_1} P\{y_{r-j} \le a \mid x_j \le a < x_{j+1}\} P\{x_j \le a < x_{j+1}\} \quad (2)$$

PROOF. The proof follows easily from the definition of the rth order statistics and the following disjoint partition:

$$\{T_r \le a\} = \bigcup_{j=0}^{m_1} \{y_{r-j} \le a\} \cap \{x_j \le a < x_{j+1}\} \quad (3)$$

□

This general result is now applied to the case where the noise terms are independent and identically distributed at each pixel.

COROLLARY 2. *If, in Theorem 1, the noise $\{N_{ij}\}$ is independent, we have*

$$P\{M \le a\} = \sum_{j=0}^{m_1} P\{y_{m-j} \le a\} P\{x_j \le a < x_{j+1}\} \quad (4)$$

where $P\{y_k \le a\} = 1$ if $k < 1$, $P\{y_k \le a\} = 0$ if $k > m_2$,

$$P\{y_k \le a\} = \sum_{j=k}^{m_2} \binom{m_2}{j} F(a-h)^j \left[1 - F(a-h)\right]^{m_2 - j}, \quad (5)$$

if $1 \le k < m_2$, and

$$P\{x_j \le a < x_{j+1}\} = \binom{m_1}{j} F(a)^j \left[1 - F(a)\right]^{m_1 - j}, \quad (6)$$

if $0 \le j \le m_1$.

PROOF. $\{x_i\}$ and $\{y_i\}$ are independent, so the conditional probabilities in (2) are the same as unconditional probabilities from which we obtain (4). The probabilities given in (5) and (6) follow directly from the distribution of order statistics of i.i.d. samples, as in [2] for example. □

The next theorem concerns the joint distribution of order statistics and is required to calculate the joint distribution of the moving medians M_1 and M_2.

THEOREM 3. *For any real numbers $a_1 \le a_2$ and integers $r \le s$, we have, when the noise $\{N_{i,j}\}$ is independent,*

$$
\begin{aligned}
&P\{T_r \le a_1, T_s \le a_2\} \\
&= \sum_{k=0}^{m_1-1}\Bigg(P\{y_{r-k} \le a_1,\ y_{s-k} \le a_2\}P\{x_k \le a_1 \le a_2 < x_{k+1}\} \\
&+ \sum_{j=k+1}^{m_1} P\{y_{r-k} \le a_1, y_{s-j} \le a_2\}P\{x_k \le a_1 \le x_{k+1}, x_j \le a_2 < x_{j+1}\}\Bigg) \\
&+ P\{y_{r-m_1} \le a_1, y_{s-m_1} \le a_2\}P\{x_{m_1} \le a_1\},
\end{aligned} \tag{7}
$$

where

$$P\{x_k \le a_1 \le a_2 < x_{k+1}\} = \binom{m_1}{k}F(a_1)^k\left[1 - F(a_2)\right]^{m_1-k}, \tag{8}$$

$$
\begin{aligned}
&P\{x_k \le a_1 < x_{k+1}, x_j \le a_2 < x_{j+1}\} \\
&= \binom{m_1-k}{j-k}\left[F(a_2) - F(a_1)\right]^{j-k}\left[1 - F(a_2)\right]^{m_1-j}\binom{m_1}{k}F(a_1)^k
\end{aligned} \tag{9}
$$

and

$$
\begin{aligned}
&P\{y_k \le a_1, y_j \le a_2\} \\
&= \sum_{i=k}^{j-1}\sum_{n=i}^{m_2}\binom{m_2-i}{n-i}\left[F(a_2-h) - F(a_1-h)\right]^{j}\left[1 - F(a_2-h)\right]^{m_2-n}\binom{m_2}{i}F(a_1-h)^i \\
&+ \sum_{n=j}^{m_2}\binom{m_2}{n}F(a_1-h)^n\left[1 - F(a_1-h)\right]^{m_2-n}.
\end{aligned} \tag{10}
$$

PROOF. It follows from (3) that

$$
\begin{aligned}
&[T_r \le a_1, T_s \le a_2] \\
&= \bigcup_{k=0}^{m_1}\bigcup_{j=0}^{m_1}\left[y_{r-k} \le a_1, x_k \le a_1 < x_{k+1}, y_{s-j} \le a_2, x_j \le a_2 < x_{j+1}\right].
\end{aligned} \tag{11}
$$

Since $a_1 \le a_2$, when $j < k$, $[y_{r-k} \le a_1, x_k \le a_1 < x_{k+1}, y_{s-j} \le a_2, x_j \le a_2 < x_{j+1}]$ must be empty; and when $j = k$,

$$
\begin{aligned}
&\left[y_{r-k} \le a_1, x_k \le a_1 < x_{k+1}, y_{s-j} \le a_2, x_j \le a_2 < x_{j+1}\right] \\
&= \left[y_{r-k} \le a_1, y_{s-k} \le a_2, x_k \le a_1 \le a_2 < x_{k+1}\right].
\end{aligned}
$$

Also

$$
\begin{aligned}
&\left[y_{r-m_1} \le a_1, x_{m_1} \le a_1, y_{s-m_1} \le a_2, x_{m_1} \le a_2\right] \\
&= \left[y_{r-m_1} \le a_1, x_{m_1} \le a_1, y_{s-m_1} \le a_2\right].
\end{aligned}
$$

So, by the independence between $\{x_i\}$ and $\{y_i\}$, (7) follows. Probability formulae (8) and (9) follow in a straightforward way from standard results about order statistics—see [2], for example. Finally

$$
\begin{aligned}
&\left[y_k \le a_1, y_j \le a_2\right] \\
&= \bigcup_{i=k}^{j-1}\left[y_i \le a_1 < y_{i+1}, y_j \le a_2\right] \cup \left[y_j \le a_1, y_j \le a_2\right] \\
&= \bigcup_{i=k}^{j-1}\bigcup_{n=j}^{m_2}\left[y_i \le a_1 < y_{i+1}, y_n \le a_2 < y_{n+1}\right] \cup \bigcup_{n=j}^{m_2}\left[y_n \le a_1 < y_{n+1}\right],
\end{aligned}
$$

so that (10) follows using (9). □

Theorem 3 can be used to calculate the joint distribution of two medians in adjacent windows as follows:

THEOREM 4. *Let $a_1 \le a_2$. Under the model* (1), *let $\{x_{ij}, x_{ij} \le x_{i,j+1}\}$, $i = 1, 2, 3$, be the sorted sets of W_1, $W_2 \cup W_3$, and W_4, respectively. Suppose that there are m_1 pixels in both W_1 and W_4, m_2 pixels in $W_2 \cup W_3$, and $2m - 1$ pixels in both $W_1 \cup W_2 \cup W_3$ and $W_2 \cup W_3 \cup W_4$. Then*

$$
\begin{aligned}
P\{M_1 \le a_1, M_2 \le a_2\} &= \sum_{k=0}^{m_1}\sum_{j=0}^{m_1} P\{x_{1j} \le a_1 < x_{1,j+1}, \\
&\quad x_{3k} \le a_2 < x_{3,k+1} \mid x_{2k} \le a_1, x_{2,m-k} \le a_2\} \\
&\quad \times P\{x_{2,m-j} \le a_1, x_{2,m-k} \le a_2\}.
\end{aligned} \tag{12}
$$

PROOF. By (3) and the definition of $\{x_{ij}\}$

$$[M_1 \le a_1] = \bigcup_{j=0}^{m_1}\left[x_{2,m-j} \le a_1, x_{1j} \le a_1 < x_{1,j+1}\right],$$

$$[M_2 \le a_2] = \bigcup_{k=0}^{m_1}\left[x_{2,m-k} \le a_2, x_{3k} \le a_2 < x_{3,k+1}\right],$$

from which (12) follows. □

COROLLARY 5. *Under the notation in Theorem 4, when the noise $\{N_{ij}\}$ in model* (1) *is independent*

$$
\begin{aligned}
P\{M_1 \le a_1, M_2 \le a_2\} &= \sum_{j=0}^{m_1}\sum_{k=0}^{m_1} P\{x_{1j} \le a_1 < x_{1,j+1}\} \\
&\quad P\{x_{3k} \le a_2 < x_{3,k+1}\}P\{x_{2,m-j} \le a_1, x_{2,m-k} \le a_2\},
\end{aligned} \tag{13}
$$

where the probabilities $P\{x_{2,m-j} \le a_1, x_{2,m-k} \le a_2\}$ can be calculated by (7), *with T_r and T_s replaced by $x_{2,m-j}$ and $x_{2,m-k}$, respectively, and $P\{x_{1j} \le a_1 < x_{1,j+1}\}$, $P\{x_{3k} \le a_2 < x_{3,k+1}\}$ can be calculated by* (6).

Corollary 5 can now be applied to derive the required probabilities of missing edges or for false detection of nonexistent edges.

THEOREM 6. *In the model* (1), *the edge missing probability is calculated as follows:*

$$P\{M_2 - M_1 \le \tau\} = \int_{a_2 - a_1 < \tau} dP\{M_1 \le a_1, M_2 \le a_2\}. \tag{14}$$

In a picture with N gray levels, the distribution of noise is discrete, and we have

$$
\begin{aligned}
&P\{M_2 - M_1 \le \tau\} \\
&= \sum_{k=1}^{N}\left[P\{M_2 \le \tau + k, M_1 \le k + 1\} - P\{M_2 \le \tau + k, M_1 \le k\}\right].
\end{aligned} \tag{15}
$$

3 APPLICATION TO DETECTING AN EDGE IN TWO DIMENSIONS

In this section, the distributional results of Section 2 will be applied to determining the probabilities of "pseudoedges" (Type I error probability, denoted α) and the probabilities of "edge missing" (Type II error probability, denoted β) when the SQUARE(3), CROSS(5), and the XSHAPE(5) pixel window shapes are used. Some preliminary notation will help the discussion.

Consider testing the null hypothesis $H_0 : h = 0$ (i.e., there is no edge) against the alternative $H_1 : h > 0$. The power functions for a given decision rule τ and an edge to be detected of height h are given by the probabilities:

$$p(\tau, h) = P\{M_2 - M_1 > \tau \mid h\}$$

from which we get the Type I error probabilities for different decision rules, τ, of

$$\alpha(\tau) = p(\tau, 0).$$

These are also referred to as the probability of detecting a "pseudoedge" in [1].

The Type II error probabilities for testing the null hypothesis that the edge is not present (i.e., has zero height) against the alternative that the edge has height = h are given by

$$\beta(\tau) = 1 - p(\tau, h).$$

These are referred to as the probability of an "edge miss" in [1].

Let τ_α denote the decision rule which attains a fixed Type I error of size α. Then the power function of the α-level test is given by $p(\tau_\alpha, h)$. When comparing different median filters (i.e., those based on different pixel shapes or pixel numbers), it is relevant to consider their power characteristics *after* their Type I error rates have been calibrated.

We turn now to application of the results of Section 2 to edge detection using the three windows mentioned above. In all cases, the surface has a mean level of zero to the left of the vertical edge and mean level of h to the right of the edge. The noise terms, N_{ij}, added to this surface at pixels (i, j) are independent from pixel to pixel and are normally distributed with mean = 0 and standard deviation = 1.

To compute the integral in (14), it would be necessary to obtain the densities of the distributions needed for the formulae of Section 2. This is a straightforward, but somewhat tedious process. Consequently, the following approximation to the integral (14) is used to calculate the values of $p(\tau, h)$:

$$P\{M_2 - M_1 \le \tau\} \cong \sum_{k=k_o}^{k_1}\Big[P\{M_2 \le (2k-1+\tau)/2N, M_1 \le k/N\} - P\{M_2 \le (2k-1+\tau)/2N, M_1 \le (k-1)/N\}\Big], \tag{16}$$

where N is a suitably large integer, say $N = 30$, and k_0 and k_1 are the smallest integers such that

$$P\{M_1 \le k_0/N\} > .0001, P\{M_1 \le k_1/N\} < .9999,$$

respectively.

This approximate procedure appears to give probabilities accurate to two decimal places. This is likely to be adequate in practice. Simulation results using 10,000 replications showed a very close match obtained between the simulated version of $p(\tau, h)$ and the approximation using (16). This confirms that, from a practical point of view, the summation approximation is adequate. Accuracy can be improved by using an even finer discrete scale than that used here (i.e., using a larger N).

Fig. 1 gives the power curves for the three windows XSHAPE(5), SQUARE(3), and CROSS(5), respectively. These confirm the behavior that might be expected from the three methods as discussed in Section 1. For instance, if vertical edges were being sought, and it was decided to use a nine-pixel smoother with a probability of detecting a "pseudoedge" of 5 percent, then the CROSS(5) gives uniformly highest power and, hence, lower probabilities of missing edges of any height. It is therefore recommended.

Of course, in practice, filters are not always designed to detect vertical (or horizontal) edges. However, with the results of this paper, the performance characteristics of various filter designs for detecting various edge shapes can be determined and used as the basis for selection of filters which are optimal for any given shape, or which show good general-purpose attributes.

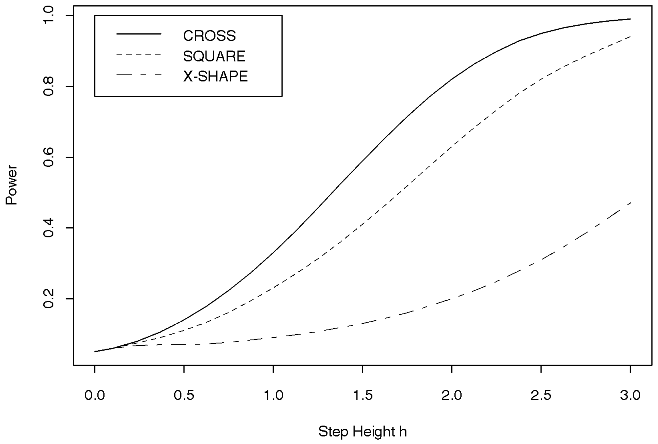

Fig. 1. Power curves $p(\tau_{.05}, h)$ for three window shapes when $\alpha = .05$ in a test of a vertical edge of height = h, calculated using approximation (16) to (14).

REFERENCES

[1] A.C. Bovik, T.S. Huang, and D. C. Munson, "The Effect of Median Filtering on Edge Estimation and Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 9, no. 2, pp. 181-194, Mar. 1987.
[2] H.A. David, *Order Statistics*. New York: John Wiley & Sons, 1970.
[3] D. Huang, W. Dunsmuir, and F. Liu, "R2D Exponential Weighted Moving Median Filter," *Proc. Second Int'l Conf. Image Processing*, pp. 621-623, World Scientific Publishing, Singapore, 1992.