COMPUTER GRAPHICS AND IMAGE PROCESSING 15, 113–129 (1981)

# A Facet Model for Image Data*

ROBERT M. HARALICK AND LAYNE WATSON

*Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061*

Received February 25, 1980

Image processing algorithms implicitly or explicitly assume an idealized form for the image data on which they operate. The degree to which the observed data meets the assumed idealized form is typically not examined or accounted for. This causes processing errors often attributed to noise. In this paper we discuss a facet model for image data which has the potential for fitting the form of the real idealized image, and for describing how the observed image differs from the idealized form. It is also an appropriate form for a variety of image processing algorithms. We give a relaxation procedure, and prove its convergence, for determining an estimate of the ideal image from observed image data.

## 1. INTRODUCTION

Operations on image data are designed to determine or estimate properties about the scene being imaged. A typical problem, for example, might be to determine homogeneous object parts or object edges. Such properties often cannot be determined without error. Sometimes it is because the algorithm works best for image data meeting certain assumptions actually not met by the observed image data.

The statistical uncertainty due to noise is unavoidable. However, the error due to using an algorithm on data that does not meet an assumed form is avoidable by first preprocessing the observed data and generating from it an estimate of its ideal form.

This decomposition of the problem into the two parts of getting an estimate of the ideal image underlying the observed image and then processing the estimate to determine the image properties is suggestive of the way this kind of problem is handled in stochastic control: get the best estimate of system state and then use a deterministic control assuming the best estimate of the actual system state. For the classical linear system with additive Gaussian noise this approach is the optimal one. For nonlinear systems it is not the best approach. But even here it gives good enough answers that it is frequently used nevertheless.

The advantage of the decomposition is its simplicity: it handles the noise when it estimates the ideal image, a process which we can call noise cleaning. Its disadvantage is that errors between the assumed form of the ideal image and the actual form of the ideal image will be exaggerated and propagated by any processing algorithm using the restored image.

In this paper we suggest a facet model for image data. The model specifies how the order and regularity in the world manifests itself in the ideal image and how the real image differs from the ideal image. The model is our working hypothesis. Our exploration here will use the facet model in its simplest form.





0146-664X/81/020113-17$02.00/0

The facet model for ideal image data assumes that the image is everywhere simple. This means that the spatial domain of the image can be partitioned into connected regions called facets each of which satisfies certain gray tone and shape constraints. The gray tones in each facet must be a polynomial function of the row–column coordinates of the pixels in the facet. In this paper we assume that the polynomial function is of degree zero, one, or two. Hence if we consider the gray tones as composing a surface above the resolution cells of the facet, then for the ideal image having a degree-one polynomial function, the surface is a sloped plane. Thus, "sloped facet model" would be an appropriate description of this specialized facet model.

The shape constraint is also simple: Each facet must be sufficiently smooth in shape. We assume that each region in the image can be exactly represented as the union of $K \times K$ blocks of pixels. The value of $K$ associated with an image means that the narrowest part of each of its facets is at least as large as a $K \times K$ block of pixels. Hence, images which can have large values of $K$ have very smoothly shaped regions. In this paper, we will take $K$ less than or equal to 3.

To make these ideas precise, let $Z_r$ and $Z_c$ be the row and column index set for the spatial domain of an image. For any $(r,c) \in Z_r \times Z_c$, let $I(r,c)$ be the gray value of resolution cell $(r,c)$ and let $B(r,c)$ be the $K \times K$ block of resolution cells centered around resolution cell $(r,c)$. Let $\pi = \{\pi_1, \ldots, \pi_N\}$ be a partition of the spatial domain of $Z_r \times Z_c$ into its facets.

In the sloped facet model, for every resolution cell $(r,c) \in \pi_n$, there exists a resolution cell $(i,j) \in Z_r \times Z_c$ such that:

(1) Shape region constraint: $(r,c) \in B(i,j) \subseteq \pi_n$;

(2) Region gray tone constraint: $I(r,c) = \alpha_n r + \beta_n c + \gamma_n$.

An observed image $J$ differs from its corresponding ideal image $I$ by the addition of random stationary noise having zero mean and covariance matrix proportional to a specified one.

$$J(r,c) = I(R,c) + \eta(r,c),$$

where

$$E[\eta(r,c)] = 0,$$

$$E[\eta(r,c)\eta(r',c')] = k\sigma(r - r', c - c').$$

The flat facet model of Tomita and Tsuji [2] and Nagao and Matsuyama [1] differs from the sloped facet model only in that the coefficients $\alpha_n$ and $\beta_n$ are assumed to be zero and Nagao uses a more generalized shape constraint which is also suitable here.

In Section 2 we describe a relaxation procedure which generates images satisfying the facet form. The relaxation procedure is proved to converge in Section 3 and has the important properties suggested by Rosenfeld [3]: in a coordinated and parallel manner, the strong influence the weak in their neighborhoods causing the weak to become consistent with the strong. In Section 4 we show some results.

## 2. NOISE CLEANING UNDER THE FACET MODEL

Noise cleaning is a procedure by which a noisy image is operated on in a manner which produces an image which has less noise and has the form of an ideal image. The facet model suggests the following simple nonlinear relaxation procedure to iteratively operate on the image until the image of ideal form is produced. Each resolution cell is contained in $K^2$ different $K \times K$ blocks. The gray tone distribution in each of these blocks can be fit by either a flat horizontal plane or a sloped plane. One of the $K^2$ blocks has smallest error of fit. Set the output gray value to be that gray value fitted by the block having smallest error. For the flat facet model this amounts to computing the variance for each $K \times K$ block a pixel participates in. The output gray value is then the mean value of the block having smallest variance (Tomita and Tsuji [2], Nagao and Matsuyama [1]).

For the sloped facet model, the procedure amounts to fitting a sloped plane to each of the blocks a given resolution participates in and outputting the fitted gray value of the given resolution cell from the block having the lowest fitting error.

The relaxation process associated with the facet model is similar in some respects to Diday's dynamic clusters method [4]. The main differences are: (1) the dynamic clusters method is a *finite* process (since there are only finitely many possible partitions and samples), while the sloped facet filtering process produces (in general) an *infinite* sequence of distinct points; (2) the residuals in each partition strictly decrease with the dynamic clusters method, whereas the sloped facet filter residuals (or sum of residuals) are not necessarily monotone; (3) the dynamic clusters method changes assignments of data points but not the points themselves, whereas the facet model relaxation repeatedly changes the data themselves.

Since the relaxation procedure for the sloped facet model is more complicated, we give a derivation here of the required equations. We assume that the block lengths are odd so that one of the block's pixels is its center. Let the block be $(2L + 1) \times (2L + 1)$ with the upper left-hand corner pixel having relative row–column coordinates $(-L, -L)$ and the lower right-hand corner pixel having relative row–column coordinates $(L, L)$. Let $J(r, c)$ be the gray value at row $r$, column $c$. According to the sloped facet model, for any block entirely contained in a facet

$$J(r, c) = \alpha r + \beta c + \gamma + \eta(r, c),$$

where $\eta(r, c)$ is the noise.

A least-squares procedure may be used to determine the estimates for $\alpha$, $\beta$, and $\gamma$. Let

$$f(\alpha, \beta, \gamma) = \sum_{r=-L}^{L} \sum_{c=-L}^{L} (\alpha r + \beta c + \gamma - J(r, c))^2.$$

The least-squares estimates for $\alpha$, $\beta$, and $\gamma$ are those which minimize $f$. To determine these values, we take the partial derivatives of $f$ with respect to $\alpha$, $\beta$, and

$\gamma$, set these to zero and solve the resulting equations for $\alpha$, $\beta$, and $\gamma$.

$$\frac{df}{d\alpha} = 2 \sum_{r=-L}^{L} \sum_{c=-L}^{L} (\alpha r + \beta c + \gamma - J(r,c))r,$$

$$\frac{df}{d\beta} = 2 \sum_{r=-L}^{L} \sum_{c=-L}^{L} (\alpha r + \beta c + \gamma - J(r,c))c,$$

$$\frac{df}{d\gamma} = 2 \sum_{r=-L}^{L} \sum_{c=-L}^{L} (\alpha r + \beta c + \gamma - J(r,c)).$$

Setting the partial derivatives to zero results in

$$\sum_{r=-L}^{L} \sum_{c=-L}^{L} (\alpha r^2 + \beta rc + \gamma r - J(r,c)r) = 0,$$

$$\sum_{r=-L}^{L} \sum_{c=-L}^{L} (\alpha rc + \beta c^2 + \gamma c - J(r,c)c) = 0,$$

$$\sum_{r=-L}^{L} \sum_{c=-L}^{L} (\alpha r + \beta c + \gamma - J(r,c)) = 0.$$

Using the facts that $\sum_{i=-K}^{K} i = 0$ and $\sum_{i=-K}^{K} i^2 = \frac{1}{3}K(K+1)(2K+1)$ we obtain

$$\tfrac{1}{3}L(L+1)(2L+1)^2\alpha - \sum_{r=-L}^{L} r \sum_{c=-L}^{L} J(r,c) = 0,$$

$$\tfrac{1}{3}L(L+1)(2L+1)^2\beta - \sum_{c=-L}^{L} c \sum_{r=-L}^{L} J(r,c) = 0,$$

$$(2L+1)^2\gamma - \sum_{r=-L}^{L} \sum_{c=-L}^{L} J(r,c) = 0.$$

Therefore,

$$\alpha = \frac{3}{L(L+1)(2L+1)^2} \sum_{r=-L}^{L} r \sum_{c=-L}^{L} J(r,c),$$

$$\beta = \frac{3}{L(L+1)(2L+1)^2} \sum_{c=-L}^{L} c \sum_{r=-L}^{L} J(r,c),$$

$$\gamma = \frac{1}{(2L+1)^2} \sum_{r=-L}^{L} \sum_{c=-L}^{L} J(r,c).$$

The meaning of this result can be readily understood for the case when the block

size is $3 \times 3$. Here $L = 1$ and

$$\alpha = \tfrac{1}{6}[\, J(+1, \cdot) - J(-1, \cdot)\,],$$
$$\beta = \tfrac{1}{6}[\, J(\cdot, 1) - J(\cdot, -1)\,],$$
$$\gamma = \tfrac{1}{9}J(\cdot, \cdot),$$

where an argument of $J$ taking the value dot means that $J$ is summed from $-L$ to $L$ in that argument position. Hence, $\alpha$ is proportional to the slope down the row dimension, $\beta$ is proportional to the slope across the column dimension, and $\gamma$ is the simple gray value average over the block. See Beaudet [5] for least-squares estimates of higher-order derivatives.

The fitted gray tone for any resolution cell $(r, c)$ in the block is given by

$$\hat{J}(r, c) = \alpha r + \beta c + \gamma.$$

For the case where $L = 1$, the $3 \times 3$ block,

$$\begin{aligned}\hat{J}(r, c) = {} & \tfrac{1}{6}[\, J(1, \cdot) - J(-1, \cdot)\,]r \\ & + \tfrac{1}{6}[\, J(\cdot, 1) - J(\cdot, -1)\,]c \\ & + \tfrac{1}{9}J(\cdot, \cdot).\end{aligned}$$

Writing this expression out in full:

$$\begin{aligned}\hat{J}(r, c) = \tfrac{1}{18}\{ & J(-1, -1)(-3r - 3c + 2) \\ & + J(-1, 0)(-3r + 2) \\ & + J(-1, 1)(-3r + 3c + 2) \\ & + J(0, -1)(-3c + 2) \\ & + J(0, 0)(2) \\ & + J(0, 1)(3c + 2) \\ & + J(1, -1)(3r - 3c + 2) \\ & + J(1, 0)(3r + 2) \\ & + J(1, 1)(3r + 3c + 2)\}.\end{aligned}$$

This leads to the set of linear filter masks shown in Fig. 1 for fitting each pixel position in the $3 \times 3$ block.

The sloped facet model relaxation procedure examines each of the $K^2$ $K \times K$ blocks a pixel $(r, c)$ belongs to. For each block, a block error can be computed by

$$\epsilon^2 = \sum_{r=-L}^{L} \sum_{c=-L}^{L} \left(\hat{J}(r, c) - J(r, c)\right)^2.$$

$\hat{J}(-,1)$

| | | |
|---|---|---|
| 8 | 5 | 2 |
| 5 | 2 | −1 |
| 2 | −1 | −4 |

$\hat{J}(-1,0)$

| | | |
|---|---|---|
| 5 | 5 | 5 |
| 2 | 2 | 2 |
| −1 | −1 | −1 |

$\hat{J}(-1,1)$

| | | |
|---|---|---|
| 2 | 5 | 8 |
| −1 | 2 | 5 |
| −4 | −1 | 2 |

$\hat{J}(0,-1)$

| | | |
|---|---|---|
| 5 | 2 | −1 |
| 5 | 2 | −1 |
| 5 | 2 | −1 |

$\hat{J}(0,0)$

| | | |
|---|---|---|
| 2 | 2 | 2 |
| 2 | 2 | 2 |
| 2 | 2 | 2 |

$\hat{J}(0,1)$

| | | |
|---|---|---|
| −1 | 2 | 5 |
| −1 | 2 | 5 |
| −1 | 2 | 5 |

$\hat{J}(1,-1)$

| | | |
|---|---|---|
| 2 | −1 | −4 |
| 5 | 2 | −1 |
| 8 | 5 | 2 |

$\hat{J}(1,0)$

| | | |
|---|---|---|
| −1 | −1 | −1 |
| 2 | 2 | 2 |
| 5 | 5 | 5 |

$\hat{J}(1,1)$

| | | |
|---|---|---|
| −4 | −1 | 2 |
| −1 | 2 | 5 |
| 2 | 5 | 8 |

FIG. 1. $3 \times 3$ linear estimators of a pixel's grey tone for the nine different $3 \times 3$ neighborhoods the pixel participates in. If the pixel's position is $(i,j)$ in the neighborhood, the estimate is $\hat{J}(i,j)$. Each mask must be normalized by dividing by 18.

One of the $K \times K$ blocks will have lowest error. Let $(r^*, c^*)$ be the coordinates of the pixel $(r, c)$ in terms of the coordinate system of the block having smallest error. The output gray value at pixel $(r, c)$ is then given by $\hat{J}(r^*, c^*)$, where $\hat{J}$ is the linear estimate of gray values for the block having smallest error of fit.

## 3. CONVERGENCE

We now prove that the sequence of images produced by relaxation using either the flat facet or sloped facet model converges. Only the one-dimensional simplest versions of the facet models are considered because these proofs capture the essence of the more general cases where technical details obscure the proof. The proofs are readily generalized to two dimensions and larger neighborhoods.

We think of the gray tones at the $m$th iteration as a finite sequence of numbers

$$X_1^{(m)}, \ldots, X_N^{(m)}$$

which becomes the sequence

$$X_1^{(m+1)}, \ldots, X_N^{(m+1)}$$

after one application of the relaxation procedure for the flat facet or sloped facet model.

*3.1. The Flat Facet Model*

We discuss the convergence of the flat facet model using a one-dimensional sequence of gray tones and a neighborhood size of 2. The flat facet relaxation has the interesting property that the algebraic order of a pair of gray tones in a block is unchanged by the relaxation. Hence if $X_k^{(m)} \geq X_{k+1}^{(m)}$, then $X_k^{(m+1)} \geq X_{k+1}^{(m+1)}$. It is this property which drives the convergence monotonically.

PROPOSITION 1. *Let $a, b, c$ be numbers and*

$$b' = \frac{a+b}{2} \quad \text{if } |a-b| \leq |b-c|$$
$$= \frac{b+c}{2} \quad \text{if } |a-b| > |b-c|.$$

*Then* (1) *$b \geq c$ implies $(b+c)/2 \leq b' \leq b + (b-c)/2$,*

(2) *$b \leq c$ implies $b + (b-c)/2 \leq b' \leq (b+c)/2$.*

*Proof.* From the definition of $b'$, $|b' - b| \leq |(b-c)/2|$.

(1) Suppose $b \geq c$. Then $|b' - b| \leq (b-c)/2$. Hence, $-(b-c)/2 \leq b' - b \leq (b-c)/2$ so that $b - (b-c)/2 \leq b' \leq b + (b-c)/2$.

(2) Suppose $b \leq c$. Then $|b' - b| \leq (c-b)/2$. Hence, $-(c-b)/2 \leq b' - b \leq (c-b)/2$ so that $b - (c-b)/2 \leq b' \leq b + (c-b)/2$.

PROPOSITION 2. *Let $a$, $b$, $c$, $d$, and $e$ be numbers. Define*

$$f(x,y,x) = (x+y)/2 \quad \text{if } |y-x| \leq |y-z|$$
$$= (y+z)/2 \quad \text{if } |y-x| > |y-z|.$$

*Let $b' = f(a,b,c)$, $c' = f(b,c,d)$, $d' = f(c,d,e)$, and $c'' = f(b',c',d')$. Suppose $b < c < d$. Then*

(1) *$c \leq c'$ implies $c' \leq c''$,*

(2) *$c \geq c'$ implies $c' \geq c''$.*

*Proof.* By Proposition 1, $b < c < d$ implies $b + (b-c)/2 \leq b' \leq (b+c)/2 \leq c' \leq (c+d)/2 \leq d' \leq d + (d-c)/2$.

(1) Suppose $c \leq c'$. Then since $b < c < d$ it follows from the definition of $f$ that $c' = (c+d)/2$ and $d - c \leq c - b$. Now note that

$$d' - c' \leq [d + (d-c)/2] - (c+d)/2 = d - c$$

and

$$c' - b' \geq (c+d)/2 - (b+c)/2 = (d-c)/2 + (c-b)/2$$
$$\geq (d-c)/2 + (d-c)/2.$$

Hence, $d' - c' \leq d - c \leq c' - b'$. Since $b' \leq c' \leq d'$ we must have by definition of $f$, $c'' = (c' + d')/2 \geq (c' + e')/2 = c'$.

(2) Suppose $c \geq c'$. Then since $b < c < d$ it follows from the definition of $f$ that $c' = (b + c)/2$ and $c - b \leq d - c$. Now note that

$$\begin{aligned} d' - c' &\geq (c + d)/2 - (b + d)/2 = (c - b)/2 + (d - c)/2 \\ &\geq (c - b)/2 + (c - b)/2 \end{aligned}$$

and

$$c' - b' \leq (b + c)/2 - [b + (b - c)/2] = c - b.$$

Hence, $c' - b' \leq c - b \leq d'c'$. Since $b' \leq c' \leq d'$ we must have by definition of $f$, $c'' = (b' + c')/2 \leq (c' + c')/2 = c'$.

THEOREM 1. *Let $X_1^{(0)}, \ldots, X_N^{(0)}$ be a given sequence of numbers, and define sequences $X_1^{(m)}, \ldots, X_N^{(m)}$, $m = 1, 2 \ldots$, by best adjacent least-squares averaging*:

$$\begin{aligned} X_k^{(m+)} &= \frac{X_{k-1}^{(m)} + X_k^{(m)}}{2}, |X_{k-1}^{(m)} - X_k^{(m)}| \leq |X_k^{(m)} - X_{k+1}^{(m)}| \quad \text{or} \quad k = N \\ &= \frac{X_k^{(m)} + X_{k+1}^{(m)}}{2}, \quad \text{otherwise.} \end{aligned}$$

*Then* (1) *the algebraic order of $X_1^{(0)}, \ldots, X_N^{(0)}$ is preserved by each sequence $X_1^{(m)}, \ldots, X_N^{(m)}$;*

(2) *If $X_i^{(0)}$ is a local* min(max), *then so is $X_i^{(m)}$;*

(3) *For all $k$ and $m$, $\min_i X_i^{(0)} \leq X_k^{(m)} \leq \max_i X_i^{(0)}$;*

(4) *For fixed $k$, $X_k^{(m)}$ is either monotone increasing or decreasing;*

(5) *$\lim_{m \to \infty} [X_1^{(m)}, \ldots, X_N^{(m)}] = [X_1^{\infty}, \ldots, X_N^{\infty}]$ exists, and each $X_k^{\infty}$ is a local non-strict extremum.*

*Proof.* (1) The proof is by induction on $m$. Suppose $X_k^m \leq X_{k+1}^m$. Then by Proposition 1, part (1), $X_{k+1}^{m+1} \geq (X_k^m + X_{k+1}^m)/2$ and by Proposition 1, part (2), $X_k^{m+1} \leq (X_k^m + X_{k+1}^m)/2$. Hence, $X_k^{m+1} \leq X_{k+1}^{m+1}$. Similarly, if $X_k^m \geq X_{k+1}^m$ then by Proposition 1, part (1), $X_k^{m+1} \geq (X_k^m + X_{k+1}^m)/2$ and by Proposition 1, part (2), $X_{k+1}^{m+1} \leq (X_k^m + X_{k+1}^m)/2$. Hence $X_{k+1}^{m+1} \leq X_k^{m+1}$.

(2) Since order relationships between neighboring points are preserved, as proved in (1), local extrema must remain extrema.

(3) If $X_k^0$ is a local minimum, then by (2) $X_k^m$ remains the local minimum for all $m$. The averaging process can do nothing but increase $X_k^m$. Hence, $X_k^m$ must be monotonically increasing. It now follows from (1) that $\min_i X_i^0 \leq X_j^m$ for all $j$ and $m$. Similarly, every local maximum is preserved and if $X_k^0$ is a local maximum, then $X_k^m$ is monotonically decreasing. Also, because of (1), $\max_i X_i^0 \geq X_j^m$ for all $j$ and $m$.

(4) The proof is by induction on $m$. If $X_k^m$ is a local extremum, the result follows from (2) and the observation made in the proof of (3). If $X_k^m$ is not a local extremum then we have two cases depending on whether $X_{k-1}^{m-1} < X_k^{m-1} < X_{k+1}^{m-1}$ or $X_{k-1}^{m-1} > X_k^{m-1} > X_{k+1}^{m-1}$. But these cases are really identical since reordering the

indices in one case will produce the other. Thus without loss of generality we suppose $X_{k-1}^{m-1} < X_k^{m-1} < X_{k+1}^{m-1}$. Now by Proposition 2, $X_k^{m-1} \leq X_k^m$ implies $X_k^m \leq X_k^{m+1}$ and $X_k^{m-1} \geq X_k^m$ implies $X_k^m \geq X_k^{m+1}$.

(5) By (3) and (4), for each fixed $k$, $X_k^m$ is a bounded monotone sequence. Therefore, by the Bolzano–Weierstrass theorem, the sequence converges

$$\lim_{m\to\infty} X_k^m = X_k^\infty.$$

Taking limits in the definition of $X_k^m$ yields that $X_k^\infty$ equals either $(X_k^\infty + X_{k+1}^\infty)/2$ or $(X_k^\infty + X_{k-1}^\infty)/2$, which implies $X_k^\infty$ equals either $X_{k-1}^\infty$ or $X_{k+1}^\infty$. Hence $X_k$ is a nonstrict local extremum for every $k$, $1 \leq k \leq N$.

### *3.2. The Slope Facet Model*

We discuss the convergence of the slope facet model using a one-dimensional sequence and a neighborhood size of 3. The slope facet relaxation has the interesting convergence property that those neighborhoods of points most collinear converge first. Hence, the strongly consistent neighborhoods do not change much and force the weakly consistent neighborhoods to be consistent with the strongly consistent neighborhoods which have already converged. What happens is very similar to the property of relaxation procedures desired by Rosenfeld [3].

LEMMA 1. *Let $y = ax + b$ be the best polynomial least-squares approximation of degree $\leq 1$ to the data points $(x_1,y_1)$, $(x_2,y_2)$, $(x_3,y_3)$, where $x_1 < x_2 < x_3$ are equally spaced. Then a residual $y_i - ax_i - b = 0$ for some $i$ if and only if the three points are collinear.*

*Proof.* By scaling and translating the points if necessary, it may be assumed without loss of generality that $x_1 = -1$, $x_2 = 0$, $x_3 = 1$. Solving the normal equations gives the least-squares fit

$$y = \frac{-y_1 + y_3}{2}x + \frac{y_1 + y_2 + y_3}{3}.$$

For each $i$, $ax_i + b - y_i = 0$ implies $(y_1 + y_3)/2 = y_2$, which implies $(x_2,y_2)$ is the midpoint of the line segment between $(x_1,y_1)$ and $(x_3,y_3)$. Hence the points are collinear. The converse is trivial.

LEMMA 2. *Let $x_1 < x_2 < x_3$ be equally spaced points and $P(x)$ the best polynomial discrete least-squares approximation of degree $\leq 1$ to the points $(x_1,y_1)$, $(x_2,y_2)$, $(x_3,y_3)$. Then*

$$p(x_1) = \frac{5y_1 + 2y_2 - y_3}{6}, \qquad y_1 - p(x_1) = \frac{y_1 + y_3 - 2y_2}{6} = s,$$

$$p(x_2) = \frac{y_1 + y_2 + y_3}{3}, \qquad y_2 - p(x_2) = -2s,$$

$$p(x_3) = \frac{5y_3 + 2y_2 - y_1}{6}, \qquad y_3 - p(x_3) = s.$$

*Proof.* As in the proof of Lemma 1 it may be assumed that $x_1 = -1$, $x_2 = 0$, $x_3 = 1$. The results then follow from the explicit formula for $p(x)$ given in the proof of Lemma 1.

THEOREM 2. *Let $y_1^0, \ldots, y_N^0$ be a given sequence of numbers, and define sequences $y_1^m, \ldots, y_N^m$, $m = 1, 2, \ldots$, by*

$$y_i^{m+1} = p(i),$$

*where $p(x)$ is the polynomial of degree of $\leq 1$ producing the best least-squares approximation among all polynomials of degrees $\leq 1$ of best discrete least-squares approximation to any three consecutive points from*

$$(i-2, y_{i-2}^m), (i-1, y_{i-1}^m), (i, y_i^m), (i+1, y_{i+1}^m), (i+2, y_{i+2}^m).$$

*Then $\lim_m y_i^m$ exists for all $i$, $1 \leq i \leq N$.*

*Proof.* The proof is by induction on the number $K$ of consecutive points which are converging. Let $(k-1, y_{k-1}^0), (k, y_k^0), (k+1, y_{k+1}^0)$ be the three points which have the best least-squares fit $p(x)$ over all the points $(i, y_i^0)$. Then by definition, $y_j^1 = p(j)$, $j = k-1, k, k+1$, and these three new points are collinear. By the definition of the sequences, collinear points remain collinear, and therefore these three points converge. Hence, $K \geq 3$. Applying this same argument to the remaining points produces, after a finite number of iterations, blocks of three or more consecutive points which remain fixed.

For the inductive step, suppose that $y_k, y_{k+1}, \ldots, y_l$ are converging and the next block of converging points starts with $y_r$. (The following argument also applies if there is no $y_r$.) Take $m$ large enough that $y_k^m, \ldots, y_l^m, y_r^m, y_{r+1}^m, \ldots$, may be considered fixed. Consider the minimum least-squares error at the points $y_{l+1}^m, y_{l+2}^m, \ldots, y_{r-1}^m$. There are some three points, say

$$y_{t-1}^m, y_t^m, y_{t+1}^m,$$

whose least-squares error is the smallest. If $l < t-1 < t+1 < r$, then $y_{t-1}^{m+1}, y_t^{m+1}, y_{t+1}^{m+1}$ are collinear, hence converging, and the induction is complete. Otherwise the minimum residual occurs at one of the ends. There are two cases:

*Case* 1. $l = t$ (the case $r = t$ is similar). By Lemma 2, $y_{l+1}^{m+1} = \frac{1}{6}(5y_{l+1}^m + 2y_l^m - y_{l-1}^m)$, $y_l^{m+1} = y_l^m$, $y_{l-1}^{m+1} = y_{l-1}^m$, and a short calculation shows that the residual at $y_{l-1}^{m+1}, y_l^{m+1}, y_{l+1}^{m+1} \leq \frac{5}{6}$ (residual at $y_{l-1}^m, y_l^m, y_{l+1}^m$).

*Case* 2. $l = t-1$ (the case $r = t+1$ is similar). Using Lemma 2 again, a short calculation shows that the residual at $y_l^{m+1}, y_{l+1}^{m+1}, y_{l+2}^{m+2} \leq \frac{1}{6}$ (residual at $y_l^m, y_{l+1}^m, y_{l+2}^m$).

As $m \to \infty$, either the minimum residual occurs in the middle, in which case the induction is complete, or at one of the ends, in which case the minimum residual decreases by a factor of $\frac{5}{6}$ or $\frac{1}{6}$ at each iteration. Therefore, the minimum residual converges to zero, and either $y_{l+1}^m$ or $y_{r-1}^m$ converges, completing the induction step.

*3.3. Remarks*

Theorems 1 and 2 are true in a number of extensions:

(1) The neighborhood size can be increased to an arbitrary $K$ points.

(2) The domain of the points can be extended from one dimension to two dimensions, thereby making the results true for image data.

(3) The neighborhood shape does not have to be square or even rectangular in the case of the two-dimensional data. This takes care of the variety of neighborhood shapes employed by Nagao and Matsuyama [1] in their flat facet iterations.

(4) The theorems are true even if norms other than the $L_2$ norm are used to determine the best approximation. For example, they are true for the $L_1$ norm and $L_\infty$ norm.

The theorems were stated and proved in their present form only in order to provide as much insight as possible. The proofs of the extensions just mentioned are conceptually the same as the proofs given here, but they present considerably more technical and notational difficulties.

Although the relaxation has been proved to converge, the meaning of the limit in relation to the starting point has not been established. In other words, this relaxation procedure suffers the same fault of many of the probabilistic relaxation labeling procedures used in image processing: the results are interesting and

FIG. 2. Original 256 × 256 test image.

FIG. 3. Result of processing the original image of Fig. 2 with five iterations of the slope facet procedure.

perhaps useful, but the problem being solved has not been stated or understood. We are actively working to try to remedy this weakness.

## 4. RESULTS

Figure 2 illustrates a 256 × 256 outdoor scene which the VISIONS group at the University of Massachusetts is working with and has supplied to us. Figure 3 illustrates the image after five iterations using the slope facet relaxation technique with a 3 × 3 window. Contrast between different regions has improved. Any region depicting features smaller or thinner than the 3 × 3 window is degraded. Most noticeable is the light above the garage door, the linear gutter feature, and the leaves on the trees. In general, textured areas become smoother and coarser textured. Edge boundaries become sharper and homogeneous areas have less noise.

To help discover what changed between the original image of Fig. 2 and the slope facet image, the absolute value of the difference between the original and the first slope facet iteration is shown in Fig. 4. Most of the changes have occurred along edges and in textured regions. This suggests that the slope facet model is a good model for the interior of any region in the image. But at the edges of region or for regions which are smaller than the $K \times K$ window size used, the model does not fit well.

FIG. 4. The absolute value of the difference between the original and the result of the first iteration of the slope facet procedure.

Just for the sake of comparison, Fig. 5 illustrates the fifth iteration of the flat facet procedure. Note the blocky, flat appearance the image has. This image clearly cannot be the ideal underlying image for the original of Fig. 2. Figure 6 shows the absolute value of the difference between the original and the first iteration of the flat facet procedure. The changes certainly occur everywhere they did with the slope facet procedure except that they are larger in magnitude and larger in area.

The quadratic facet model allows a much better fit to occur near the edges. Figure 7 shows a blow-up of rows 80 to 143 and columns 60 to 123 of the left-hand corner of the house. Figure 8 shows the corresponding blow-up of the fifth iteration quadratic facet. Aside from the difference due to the photographic developing process, the images look the same at the pixel level. Confirming this, Fig. 9 illustrates the absolute value of the differences between the original in Fig. 7 and the quadratic facet result of Fig. 8. There is relatively little spatial structure in the difference image.

For the sake of comparison, Fig. 10 shows the same section after processing with five iterations of the slope facet procedure and Fig. 11 the same for five iterations of the flat facet procedure. These images confirm our earlier comments about the flat facet model being really incorrect and the slope facet model being correct in region interiors but not at edges.

FIG. 5. The result of processing the original image of Fig. 2 with five iterations of the flat facet procedure.

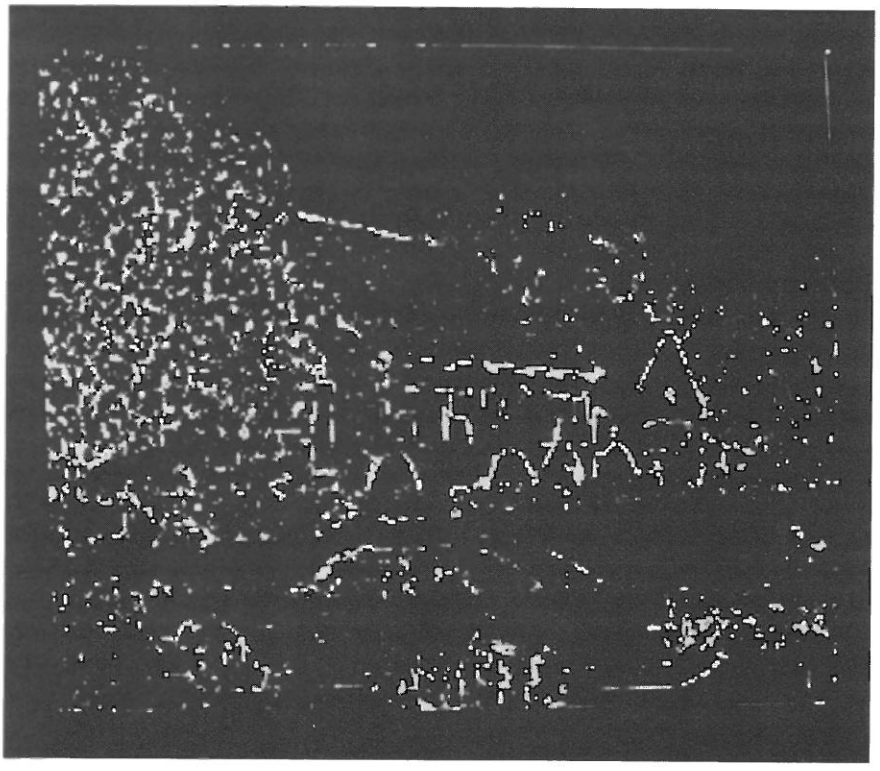

FIG. 6. The absolute value of the difference between the original and the first iteration of the flat facet procedure.

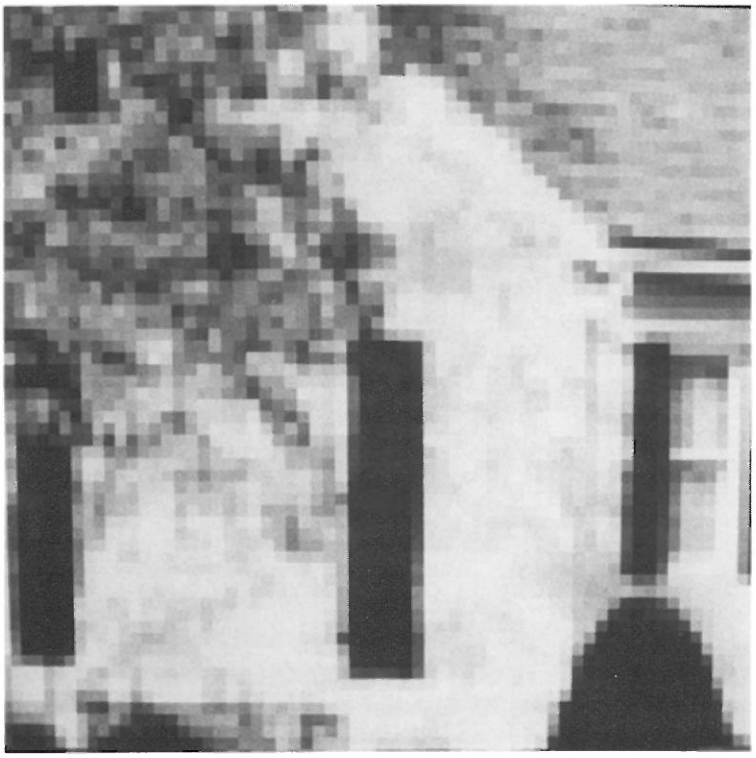

FIG. 7. Blow-up of the left corner of the house of the image of Fig. 2.

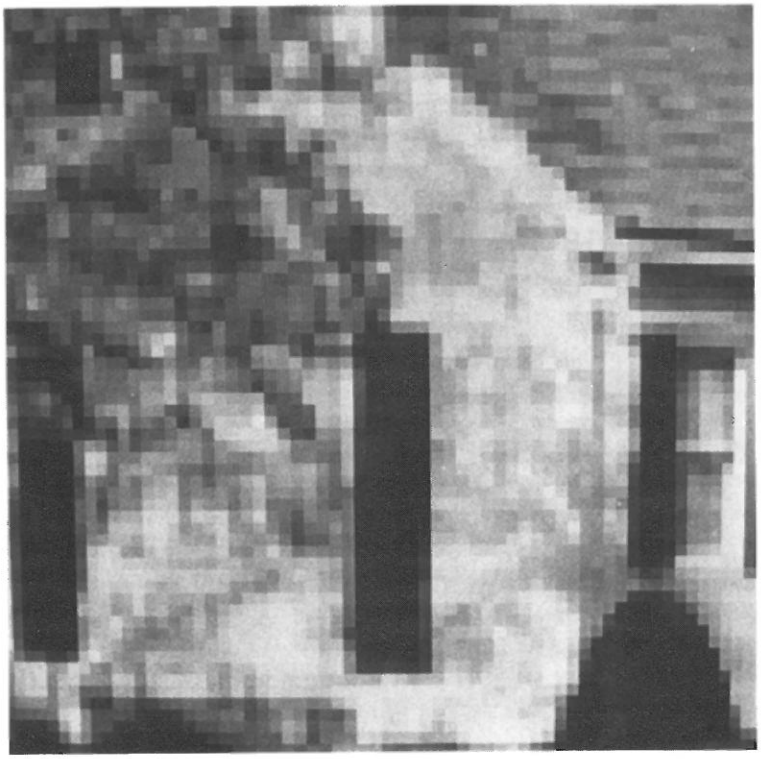

FIG. 8. The result of processing the image of Fig. 7 with the quadratic facet model procedure.

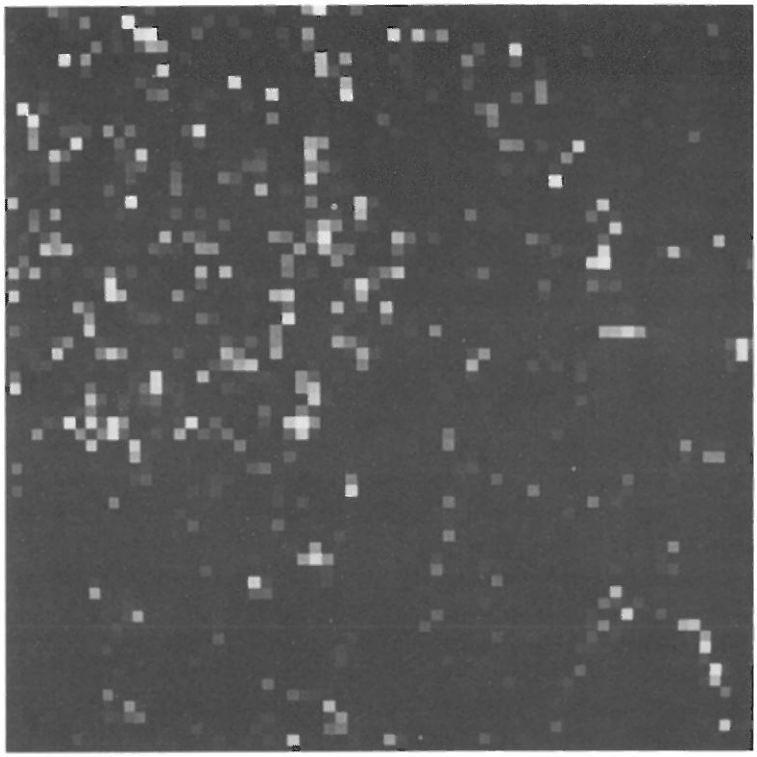

FIG. 9. The absolute value of the difference between the original of Fig. 7 and the quadratic facet image of Fig. 8.

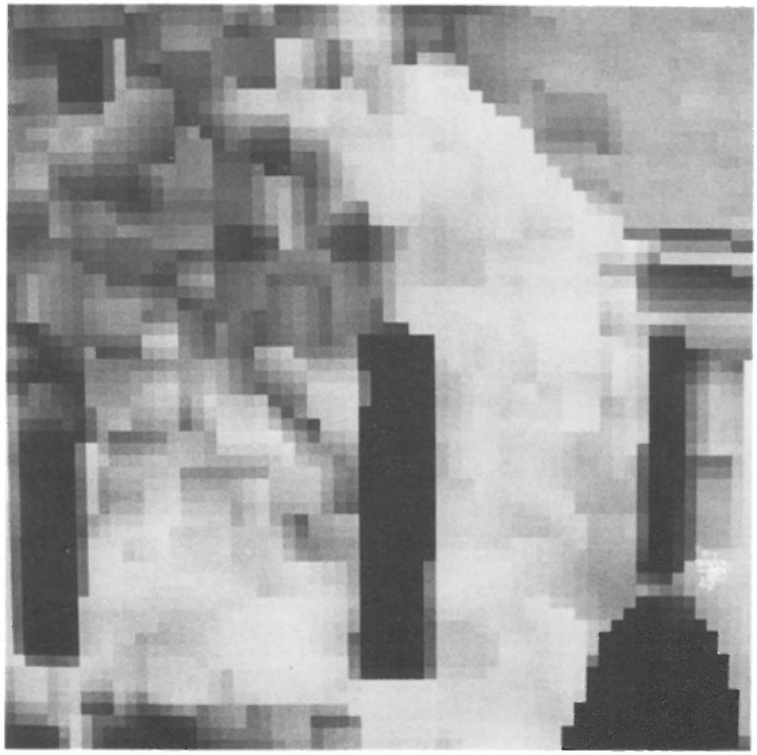

FIG. 10. The result of processing the blow-up of Fig. 7 with five iterations of the slope facet procedure.

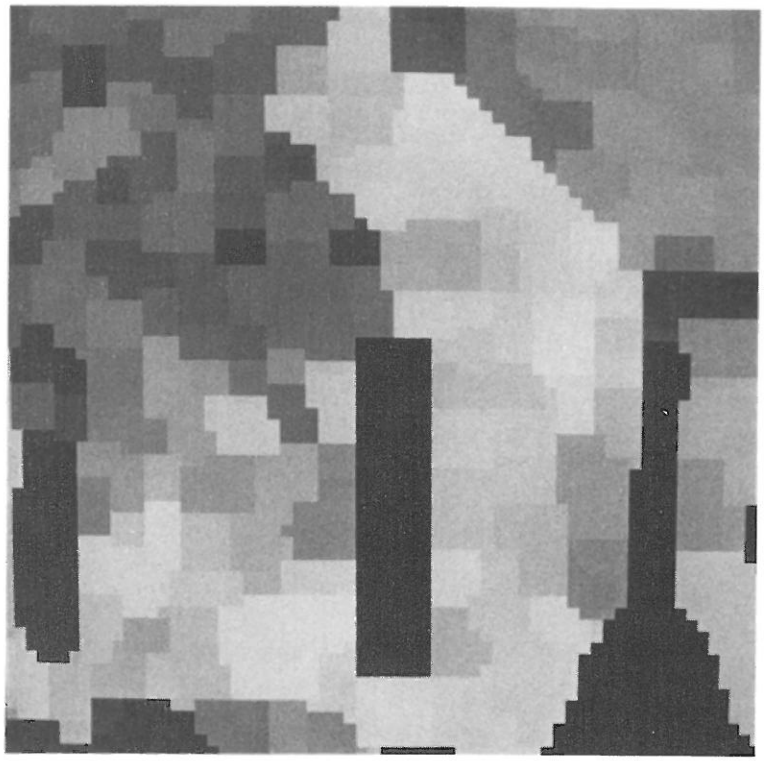

FIG. 11. The result of processing the blow-up of Fig. 7 with five iterations of the flat facet procedure.

## 5. CONCLUSION

We have discussed a facet model for image data which represents the underlying ideal image as a piecewise polynomial function, each piece being called a facet. We have suggested that those image processing operations that require an image to be in idealized form first restore the image to that ideal form by the facet iteration algorithm which we described and proved convergent. There is much work that remains to be done concerning the choice of the slope facet or quadratic facet model for any particular image and the size of window that is most appropriate. Experiments using the resulting facet image as the input image to various edge operators and region segmentation operators need to be tried and we will report on them in future papers.

## REFERENCES

1. M. Nagao and T. Matsuyama, Edge preserving smoothing, in *Proceedings, Fourth International Joint Conference on pattern Recognition, Kyoto, Japan, November 1978.*
2. F. Tomita and S. Tsuji, Extraction of multiple regions by smoothing in selected neighborhoods, *IEEE Trans. Syst. Man Cybernet.* **SMC-7**, No. 2, 1977, 107–109.
3. A. Rosenfeld, Iterative methods in image analysis, *Pattern Recognition* **10**, 1978, 181–187.
4. E. Diday, The dynamic clusters method in non-hierarchial clustering, *Int. J. Comput. Inform. Sci.* **2**, No. 1, 1973, 61–88.
5. P. R. Beaudet, Rotationally invariant image operators, in *Proceedings, Fourth International Joint Conference on Pattern Recognition, Kyoto, Japan, November 1978*, pp. 579–583.