In stochastic representation an image is considered to be a sample function of an array of random variables called random fields. This characterization of an ensemble of images is useful in developing image processing techniques that are valid for an entire class and not just for an individual image.

The following figure shows the stochastic models used in the literature.

```
Stochastic Models Used in Image Processing

Covariance Models
1. Separable exponential
2. Nonseparable exponential

1-D Models
1. AR and ARMA
2. State variable
3. Noncausal minimum variance

2-D Models
1. Causal
2. Semicausal
3. Noncausal
```
Covariance Models

In many applications such as image restoration and data compression it is often sufficient to characterize an ensemble of images by its mean and covariance functions. Often one starts with a stationary random field representation where the mean is held constant and the covariance function is represented by the separable or the nonseparable exponential models. For example the separable covariance model

$$r(m,n) = \sigma^2 \rho_1^{|m|} \rho_2^{|n|}, \quad 0 < \rho_1, \rho_2 < 1$$

is very convenient for analysis of image processing algorithms. Also, the nonseparable exponential covariance model

$$r(m,n) = \sigma^2 \exp\left(-\sqrt{\alpha_1^m + \alpha_2^n}\right)$$

is a better model but is not as convenient for analysis.

An alternative to representing random fields by mean and covariance functions is to characterize them as the outputs of linear systems whose inputs are random fields with known statistics (such white noise). Such linear systems are represented by difference equations and are often useful in developing computationally efficient image processing algorithms. The problem of finding a linear stochastic DE model that realizes the covariance of an ensemble is known as the spectral factorization problem.
1-D Models

A simple way to characterize an image is to consider it a 1-D signal that appears at the output of a raster scanner, i.e., a sequence of rows and columns. If the interrow and inter-column dependencies are ignored then 1-D linear systems are useful for modeling such signals.

Let $u(n)$ be a real, stationary random sequence with zero-mean and covariance $r(n)$. If $u(n)$ is considered as the output of a stable, linear shift-invariant system $H(z)$ whose input is a stationary zero-mean random sequence $e(n)$:

$$
\begin{align*}
    e(n) &\overset{g}{\rightarrow} H(z) (1-D) & u(n) \\
    S(e(z)) &\rightarrow S(z) & SDF \{ u(n) \}
\end{align*}
$$

The SDF of $u(n)$ is then given by

$$
S(z) = S_e(z) H(z) H(z^{-1})
$$

where $z = e^{j\omega}$, $-\pi < \omega \leq \pi$. If $H(z)$ must also be causal while remaining stable, then it must have a one-sided Laurent series

$$
H(z) = \sum_{n=0}^{\infty} h(n)z^{-n}
$$

and all its poles must lie inside the unit circle.
We now define an Autoregressive (AR) Model. A zero mean sequence $u(n)$ is called an AR process of order $p$ when it can be generated as the output of the system shown.

\[ u(n) = \sum_{k=1}^{p} a(k) u(n-k) + e(n) \]

where

\[ E[e(n)] = 0, \quad E[e^2(n)] = \beta^2 \quad E[e(n)e(m)] = \beta^2 \delta(n-m) \]

and it is assumed that

\[ E[e(n)u(m)] = 0 \]

for any $n > m$. This means that the output $u(m)$ cannot predict the input $e(n)$ (Causality). Note that this system uses the most recent $p$ outputs and the current input to generate recursively the next output.
Properties of AR Models.
\[ u(n) = \sum_{k=1}^{p} a(k) u(n-k) + e(n), \quad \forall n \] (A)

\( u(n) \) is the best linear mean square predictor of \( u(n) \) based on all its past \( p \) samples.

\[ u(n) = \overline{u(n)} + e(n) \]

also called the innovations sequence.

\[ e(n) \quad H(z) \quad u(n) \]

\[ e(n) \quad + \quad z^{-1} \quad + \]

\[ \sum_{k=1}^{p} a(k) z^{-k} \]

\[ H(z) \]
From \( \mathbf{H} (z) = 0 \)

\[
U(z) = \sum_{k=1}^{P} a(k) U(z) z^{-k} + E(z)
\]

\[
U(z) \left[ 1 - \sum_{k=1}^{P} a(k) z^{-k} \right] = E(z)
\]

\[
H(z) = \frac{U(z)}{E(z)} = \frac{1}{1 - \sum_{k=1}^{P} a(k) z^{-k}} \quad \leftarrow \text{all pole} \quad \text{TF.}
\]

2. Note that if we switch \( e(n) \) and \( u(n) \), we come up with a "whitening filter":

\[
\begin{array}{ccc}
\Downarrow & & \Downarrow \\
U(n) & \rightarrow & Ap(z) \\
& & \rightarrow \\
& & E(n)
\end{array}
\]

Where \( Ap(z) = 1 - \sum_{k=1}^{P} a(k) z^{-k} \).
3. 

\[ s(z) = S_e(z) H(z) H(z^{-1}) \]
\[ = \frac{\beta^2}{A_p(z) A_p(z^{-1})} \]
\[ z = e^{j\omega} \]
\[-\pi < \omega \leq \pi \]

4. So far, we have assumed that \( E[ e(n)^2 ] = 0 \) which results in \( E[ u(n)^2 ] = 0 \). (This is a linear system). If \( u(n) \) is non-zero mean, then the AR model can be modified as

\[ x(n) = \sum_{k=1}^{p} a(k) x(n-k) + e(n) \]

\[ u(n) = x(n) + \mu \]
\[ E[u(n)^2] = \mu \]

where \( x(n) \) is zero mean.
We now use a method to identify the AR Coefs $\beta$.

\[ u(n) = \sum_{k=1}^{p} a(k) u(n-k) + e(n) \]

\[
\begin{align*}
E \left\{ u(n)^2 \right\} &= E \left\{ \sum_{k=1}^{p} a(k) u(n-k)^2 + e(n)^2 \right\} \\
&= \sum_{k=1}^{p} a(k) E \left\{ u(n-k)^2 \right\} + E \left\{ e(n)^2 \right\} \\
&= \sum_{k=1}^{p} a(k) \delta(n-k) + \beta^2 \delta(n-m)
\end{align*}
\]

We have three cases:

I. \( n < m \) \( \Rightarrow \) \( E \left\{ u(n) e(m) \right\} = 0 \) \( \text{by def} \)

II. \( n = m \) \( \Rightarrow \) \( E \left\{ u(n) e(n) \right\} = \sum_{k=1}^{p} a(k) E \left\{ u(n-k) e(n) \right\} + \beta^2 \delta(n-m) \)

\( \Rightarrow \) \( E \left\{ u(n) e(n) \right\} = \beta^2 \)

III. \( n > m \) \( \Rightarrow \) \( E \left\{ u(n) e(m) \right\} \text{ Not needed!} \)
\[ E \{ u(n) e(m) \} = \beta^2 \delta(n-m), \quad n \leq m. \]

Now consider,

\[
E \{ u(0) u(n) \} = E \left\{ u(0) \sum_{k=1}^{p} a(k) u(n-k) \right\} + E \left\{ u(0) e(n) \right\} \\
= \sum_{k=1}^{p} a(k) E \{ u(0) u(n-k) \} + E \{ u(0) e(n) \} \\
= \sum_{k=1}^{p} a(k) r(n-k) + \beta^2 \delta(n) \\
\therefore \quad r(n) = \sum_{k=1}^{p} a(k) r(n-k) + \beta^2 \delta(n), \quad \forall \ n \geq 0.\]
Summary:

We have shown that

\[ r(n) = \sum_{k=1}^{p} a(k) r(n-k) + \beta^2 \delta(n), \quad \forall n \geq 0 \]

or

\[ r(n) - \sum_{k=1}^{p} a(k) r(n-k) = \beta^2 \delta(n) \quad \forall n = 0, 1, 2, \ldots, p. \]

This result is very important for identification of the AR model parameters \(a(k)\) and \(\beta^2\) from a given set of covariances \(\{ r(n), -p \leq n \leq p \}\). In fact, a \(p\)th-order AR model can be uniquely determined by solving the last eqn for \(n = 0, 1, \ldots, p\). In matrix notation, this is equivalent to solving the following normal eqns. When \(n = 0\), we have

\[ r(0) - a^T r = \beta^2 \]

and for \(n = 1, 2, \ldots, p\), we get

\[ r - Ra = 0 \]

\[ R \cdot a = r \]

where \(R\) is the \(p \times p\) Toeplitz matrix

\[
R = \begin{bmatrix}
r(0) & r(1) & \cdots & \cdots & r(p-1) \\
r(1) & r(0) & \cdots & \cdots & r(p-2) \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
r(p-1) & \cdots & r(1) & r(0)
\end{bmatrix}
\]

and

\[ a = [a(1) \ a(2) \ \ldots \ a(p)]^T \]

\[ r = [r(1) \ r(2) \ \ldots \ r(p)]^T \]

If \(R\) is positive definite, then the AR model is guaranteed to be stable.
Example \[ r(n) = \sigma^2 / n \]

Let us find an AR model of order 2 \( (P=2) \).

\[
R a = r
\]

\[
\begin{bmatrix}
\sigma^2 & \sigma^2 \\
\sigma^2 & \sigma^2 \\
\end{bmatrix}
\begin{bmatrix}
a(1) \\
a(2)
\end{bmatrix}
=
\begin{bmatrix}
r(1) \\
r(2)
\end{bmatrix}
\]

\[
\begin{bmatrix}
1 & \rho \\
\rho & 1
\end{bmatrix}
\begin{bmatrix}
a(1) \\
a(2)
\end{bmatrix}
= \rho^2
\]

Also,

\[
r(0) - a^T r = \beta^2
\]

\[
\sigma^2 - (a(1) a(1)) \begin{bmatrix} r(1) \\ r(2) \end{bmatrix} = \beta^2
\]

\[
\sigma^2 - \rho^2 \sigma^2 = \beta^2
\]

\[
\beta^2 = \sigma^2 (1 - \rho^2)
\]

\[
\therefore u(m) = \rho u(m-1) + e(m) ; \ \text{Var} \{ e(m) \} = \sigma^2 (1 - \rho^2)
\]
The corresponding representation for a scan line of the image, having pixel mean of μ, is

\[
\begin{align*}
\left\{ \begin{array}{l}
\sum_{k=1}^{2} a(k) x(n-k) + e(n) \\
u(n) = x(n) + \mu
\end{array} \right.
\end{align*}
\]

or

\[
\begin{align*}
x(n) &= \sum_{k=1}^{2} a(k) x(n-k) + e(n) \\
u(n) &= x(n) + \mu
\end{align*}
\]

Also, recall that

\[
H(z) = \frac{1}{1 - \rho z^{-1}} \\
S(z) = \frac{\sigma^2 (1 - \rho^2)}{(1 - \rho^2)(1 - \rho^2)}
\]

\[
S_e(z) = \sigma^2 e^{-z}
\]

\[
\begin{array}{c}
S(z) \\
H(z) \\
S_e(z)
\end{array}
\]
**MOVING AVERAGE REPRESENTATIONS**

A random sequence $u(n)$ is called a moving average (MA) process of order $q$ when it can be written as a weighted running average of uncorrelated random variables

$$u(n) = \sum_{k=0}^{q} b(k) e(n-k)$$

where $e(n)$ is a zero mean white noise process of variance $\beta^2$.

![Diagram](attachment:image.png)

$$S(z) = S(z) B(z) B_{-1}^{-1}$$

$$= \beta^2 B_{-1}^{0} B_{q}^{-1}$$

Note that $B_{-1}$ is an FIR filter, which means MA representations are **all-zero models**.
Example

Let \( q = 1 \) (1st order MA process)

\[
\begin{align*}
U(n) &= b(0) \epsilon(n) + b(1) \epsilon(n-1) \\
    &= \epsilon(n) - \alpha \epsilon(n-1)
\end{align*}
\]

\[
E\{\epsilon(n)\epsilon(m)\} = \beta^2 \delta(m-n)
\]

Then

\[
B(q) = \sum_{k=0}^{q} b(k) z^{-k} = 1 - \alpha z^{-1}
\]

\[
S(z) = \beta^2 \left[ (1 - \alpha z^{-1})(1 - \alpha (-z)^{-1}) \right]
\]
Autoregressive Moving Average (ARMA) Representations

\[ E(m) \xrightarrow{B_q(z)} X(m) \xrightarrow{\frac{1}{A_p(z)}} U(m) \]

MA output:
\[ q \sum_{k=0}^{q} b(k) E(m-k) = X(m) \]

\[ U(m) = \sum_{k=1}^{p} a(k) U(m-k) + X(m) \]

\[ X(m) = -\sum_{k=1}^{p} a(k) U(m-k) + U(m) \]

\[ = \sum_{k=0}^{p} a(k) U(m-k) \quad , \quad a(0) = 1 \]

\[ \sum_{k=0}^{p} a(k) U(m-k) = \sum_{l=0}^{q} b(l) E(m-l) \]

ARMA model \((p, q)\).
ONE-DIMENSIONAL SPECTRAL FACTORIZATION

Spectral factorization refers to the determination of a white noise driven linear system such that the PSD of its output matches a given PSD. Basically, we have to find a causal and stable system $H(z)$ whose white noise input has the spectral density $K$, a constant, such that

$$S(z) = K H(z) H(z^{-1})$$

Note that the spectral factorization is seen from the eqn above; we need to find $H(z)$ from $S(z)$. Also,

$$S(e^{j\omega}) = K H(e^{j\omega}) H(e^{-j\omega})$$

$$= K |H(\omega)|^2$$

The spectral factorization problem is equivalent to finding a causal, stable, linear filter that realizes a given magnitude frequency response. This is also equivalent to specifying the phase of $H(\omega)$, because its magnitude can be calculated within a constant from $S(\omega)$. 
\[ S(z) = \frac{4z - (z + z^{-1})}{2 - (z + z^{-1})} \]
\[ = \frac{4z - z^2 - 1}{2z - z^2 - 1} \]
\[ = \frac{z^2 - 4z + 1}{z^2 - 0.5z + 1} \]
\[ = \frac{(z-1)(z-0.25)}{(z-1-0.5z^2)} \]
\[ = \frac{(z-1)(0.25z-1)}{(z-1)(1-0.5z^{-1})} \]
\[ = \frac{(1-0.25z)(1-0.25z^{-1})}{(1-0.5z^2)(1-0.5z^{-1})} \]
\[ \therefore H(z) = \frac{U(z)}{E(z)} = \frac{1-0.25z^{-1}}{1-0.5z^{-1}} \]

or
\[ u(n) = 0.5 u(n-1) + e(n) - 0.25 e(n-1) \]
\[ E\{e(n)\} = 0 \]
\[ E\{e(n)e(m)\} = 2 \delta(n-m) \]
LINEAR PREDICTION IN 2-D

The notion of causality does not extend naturally to 2 or higher dimensions. Line-by-line processing techniques that utilize the simple 1-D algorithms do not exploit the 2-D structure and the interline dependence. There are three types of data structures to characterize 2-D models.

1. Causal
2. Semicausal
3. Noncausal

We shall consider these in the framework of linear prediction.

These three types of stochastic models have applications in many image processing problems. For example, causal models yield recursive algorithms in data compression of images by differential pulse code modulation (DPCM) technique and in recursive filtering of images.

Semicausal models are causal in one direction and noncausal in the other, and lead themselves naturally to hybrid algorithms, which are recursive in one dimension and nonrecursive in the other. Noncausal models give rise to transform-based algorithms.
We now define the linear prediction models. Let \( u(m,n) \) be a stationary random field with zero mean and covariance \( r(k,\ell) \). Let \( \bar{u}(m,n) \) denote a linear prediction estimate of \( u(m,n) \), defined as

\[
\bar{u}(m,n) = \sum_{(k,\ell) \in \hat{S}_x} a(k,\ell)u(m-k,n-\ell)
\]

where \( a(k,\ell) \) are called the predictor coefficients and \( \hat{S}_x \), a subset of the 2-D lattice, is called the prediction region.

The samples included in \( \hat{S}_x \) depend on the type of the prediction considered, namely, causal \( (x = 1) \), semicausal \( (x = 2) \), or noncausal \( (x = 3) \). With a hypothetical scanning mechanism that scans sequentially from top to bottom and left to right, the three prediction regions are defined as follows.
Causal Prediction

This predictor is a function of only the elements that arrive before it. Thus,

\[ S = \{ l \geq 1, \forall k \in U \{ l = 0, k \geq 1 \} \} \]

This is also called the non-symmetric half-plane (NSHP) model.

\( (*) \) also includes the special case of single-quadrant causal predictors:

\[ u(m, n) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} a(k, l) u(m-k, n-l) \]

This is called a strongly causal predictor.
Semi-causal prediction

Here,

\[ \hat{S}_2 = \frac{1}{l \geq 1, \forall k^2 \cup \{ l = 0, k \neq 0 \} } \]

Note that \( \hat{S}_2 \) is causal in \( n \) direction and noncausal in \( m \) direction.
Non Causal Prediction

A non causal predictor $\tilde{u}(m,n)$ is a function of possibly all the variables in the random field except $u(m,n)$ itself. The region is defined as:

$$\hat{S}_3 = \{ \forall (k,l) \neq (0,0) \}$$
Example

(a) $U(m,n) = a_1 u(m-1,n) + a_2 u(m, n-1) + a_3 u(m-1, n-1)$

Causal

(b) $U(m,n) = a_1 u(m-1,n) + a_2 u(m+1, n) + a_3 u(m, n-1)$

Semi causal
\( \tilde{u}(m,n) = a_1 u(m-1,n) + a_2 u(m+1,n) + a_3 u(m,n-1) \)

\( k=0, l=-1 \)

\( + a_4 u(m,n+1) \)

Non Causal
Remarks

In practice, only a finite neighborhood, called a prediction window

\[ \hat{W}_x \subset S_x \]

can be used in the prediction process, so that

\[ \bar{u}(m,n) = \sum \sum_{(k,l) \in \hat{W}_x} a(k,l)u(m-k,n-l) \]

Some commonly used \( \hat{W}_x \) are:

1. Causal:

\[ \hat{W}_1 = \{ -p \leq k \leq p, 1 \leq l \leq q \} \cup \{ 1 \leq k \leq p, l = 0 \} \]

2. Semicausal:

\[ \hat{W}_2 = \{ -p \leq k \leq p, 0 \leq l \leq q, (k,l) \neq (0,0) \} \]

3. Noncausal:

\[ \hat{W}_3 = \{ -p \leq k \leq p, -q \leq l \leq q, (k,l) \neq (0,0) \} \]

We also define

\[ W_x = \hat{W}_x \cup (0,0) \]
Recall that

\[ \bar{u}(m,n) = \sum_{(k,l) \in \hat{W}_x} a(k,l) u(m-k, n-l) \]

\[ (m,n) \in S_x \quad , \quad x = 1, 2, 3. \]

We wish to determine the prediction coefficient \( a(m,n) \) using a minimum variance criterion. This requires that the variance of the prediction error be minimized; i.e.,

\[
\min_{a(m,n)} E \left\{ \frac{1}{E(m,n)} \right\} \left[ u(m,n) - \bar{u}(m,n) \right]^2 \leq \beta^2
\]

Using the orthogonality condition, we have

\[
E \left\{ E(m,n) \bar{u}(m,n) \right\} = 0 \quad , \quad \forall (m,n)
\]

\[ \sum_{(k,l) \in \hat{W}_x} a(k,l) u(m-k, n-l) \]
\[
\sum \sum_{(k,l) \in \hat{W}_x} a(k,l) E \{ E(m,n) u(m-k, n-l) \} = 0, \quad \forall (m,n) \].

or
\[
\sum \sum_{(i,j) \in \hat{W}_x} a(i,j) u(m-i, n-j) = 0, \quad (i,j) \in \hat{W}_x, \quad \forall (m,n) \].

or
\[
E \{ [u(m,n) - \sum \sum_{(i,j) \in \hat{W}_x} a(i,j) u(m-i, n-j)] u(m-k, n-l) \} = 0.
\]  
\[
(k,l) \in \hat{W}_x, \quad \forall (m,n) .
\]

or
\[
E \{ u(m,n) u(m-k, n-l) \} - \sum \sum_{(i,j) \in \hat{W}_x} a(i,j) E \{ u(m-i, n-j) u(m-k, n-l) \} = 0.
\]  
\[
(k,l) \in \hat{W}_x, \quad \forall (m,n) .
\]

or
\[
E \{ u(m,n) u(m-k, n-l) \} - \sum \sum a(i,j) r(k-i, l-j) = 0.
\]  
\[
(k,l) \in \hat{W}_x, \quad \forall (m,n) .
\]  
\[
(r(k,l) = r(m-i - m+k, n-j - n+l)\]
Note that \( \hat{x} \) does not include \((0,0)\). Recall that the orthogonality condition is

\[
E \{ E(m,n) u(m-k, n-l) \} = 0. \tag{1}
\]

For \((k,l) = (0,0)\), however, we get a non-zero value for \((1)\). This is shown as follows.

\[
\beta^2 \leq \min E \left\{ E(m,n)^2 \right\} = \min E \left\{ E(m,n) \left[ u(m,n) - \bar{u}(m,n) \right]^2 \right\}
\]

\[
= \min \left[ E \left\{ E(m,n) u(m,n)^2 \right\} - E \left\{ E(m,n) \bar{u}(m,n) \right\} \right]
\]

or orthogonality principle

\[
= E \left\{ E(m,n) u(m,n)^2 \right\} \tag{2}
\]

(1) \[ \text{and } (2) \[ = 1 \]

\[
E \left\{ E(m,n) u(m-k, n-l) \right\} = \beta^2 \tag{3}
\]

\((k,l) = (0,0)\).
From (4) and (3) \( D \)

\[
\begin{align*}
    r(k, l) - \sum_{(i, j) \in \hat{W}_x} a(i, j) r(k-i, l-j) &= \beta^2 \delta(k, l), \\
    (k, l) \in W_x, & \forall (m, n)
\end{align*}
\]

where

\[
W_x = \hat{W}_x \cup (0, 0), \quad x = 1, 2, 3.
\]

the solution to (4) gives the predictor \( a(k, l) \) and the prediction error variance \( \beta^2 \).
Stochastic Representation of Random Fields

In general,

\[ U(m,n) = \bar{u}(m,n) + E(m,n) \]

here are the types of representations:

1. Minimum Variance Representations (MVR)

2. White noise driven Representations (WNDR)

3. ARMA

in this case \( E(m,n) \) is a 2-D moving average

here \( E(m,n) \) is chosen to be a white noise field

here \( \bar{U}(m,n) \) is chosen to be a white noise predictor