

Visão Tridimensional

Estereoscopia

→ Estimação de profundidades a partir de imagens binoculares

• Análise Geométrica

- Duas câmeras de distância focal f , e centros em O_1 e O_2

- Distância $O_1O_2 = d$

- Ângulo entre os eixos ópticos: 2θ

- Sistemas de coordenadas:

$$\begin{cases} \text{Camera 1 : } X_1, Y_1, Z_1 \\ \text{Camera 2 : } X_2, Y_2, Z_2 \\ \text{Cena : } X, Y, Z \end{cases}$$

- Relação entre os sistemas:

$$\begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \end{pmatrix} = \begin{pmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{pmatrix} \begin{pmatrix} X + \frac{d}{2} - f \sin \theta \\ Y \\ Z \end{pmatrix}$$

$$\begin{pmatrix} X_2 \\ Y_2 \\ Z_2 \end{pmatrix} = \begin{pmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{pmatrix} \begin{pmatrix} X - \frac{d}{2} + f \sin \theta \\ Y \\ Z \end{pmatrix}$$

- Projeção perspectiva

→ Ponto $P = (X, Y, Z)$ na cena é projetado nas posições (x_1, y_1) e (x_2, y_2) nas duas imagens, onde

$$x_i = \frac{f X_i}{f - Z_i} \quad \text{e} \quad y_i = \frac{f Y_i}{f - Z_i}, \quad i = 1, 2$$

→ A partir dos observáveis (x_1, y_1) e (x_2, y_2) , obtém-se a posição tridimensional do ponto P , invertendo-se as relações acima

$$\begin{cases} x_1 = x_1(X, Y, Z) \\ y_1 = y_1(X, Y, Z) \end{cases} \quad \text{e} \quad \begin{cases} x_2 = x_2(X, Y, Z) \\ y_2 = y_2(X, Y, Z) \end{cases}$$

→ 4 equações a 3 incógnitas, X , Y e Z

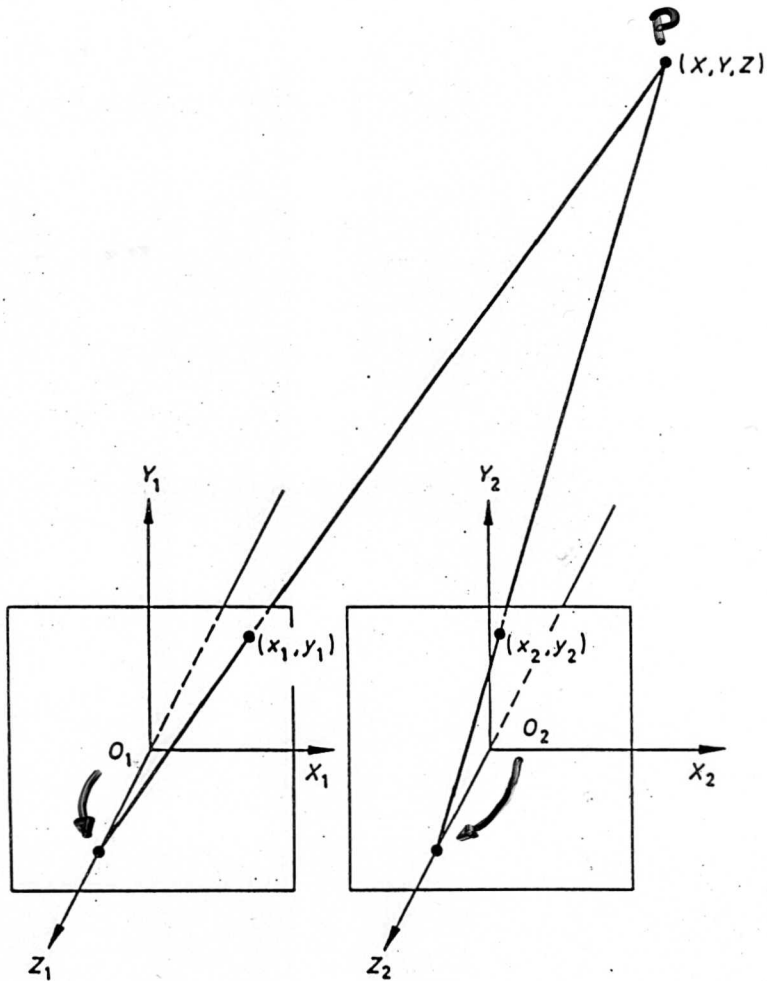
→ Existe uma relação entre as posições de projeção do ponto P nos dois planos de imagem

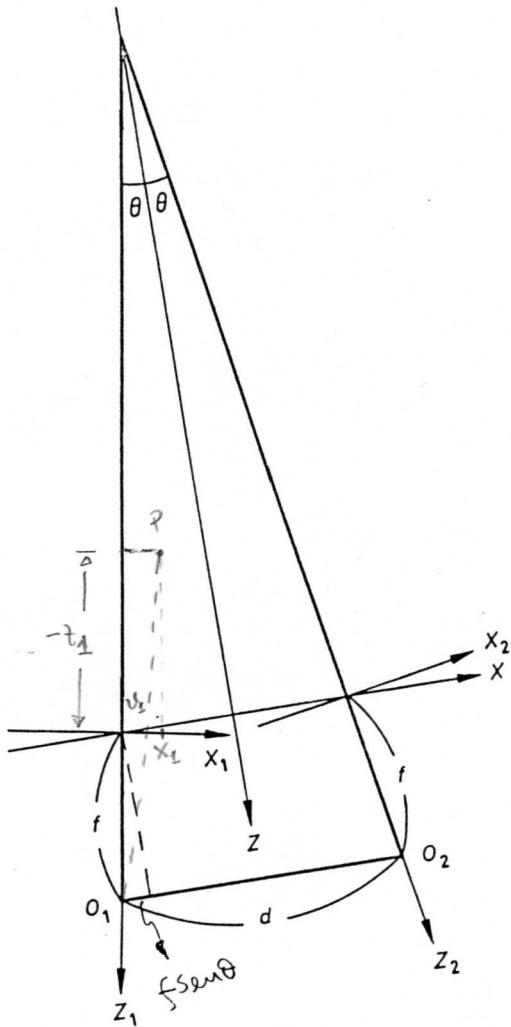
e.g., A partir de x_1, y_1 e x_2 , obtém-se

$$\begin{cases} X = X(x_1, y_1, x_2) \\ Y = Y(x_1, y_1, x_2) \\ Z = Z(x_1, y_1, x_2) \end{cases}$$

→ A quarta equação pode ser escrita como

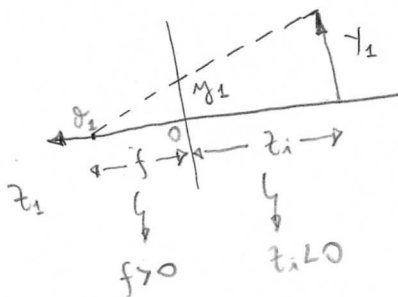
$$y_2 = y_2(x_1, y_1, x_2) \quad (\text{Equação da } \textit{Linha Epipolar})$$





$$\frac{y_i}{f - z_1} = \frac{y_1}{f}$$

$$\therefore y_1 = \frac{f}{f - z_1} y_i$$



- Linha Epipolar

- A um ponto da cena, observado em (x_1, y_1) na primeira imagem, corresponde uma reta passando por (x_1, y_1) e pelo centro de projeção O_1 .

- Esta reta, como vista (i.e., projetada) na segunda imagem, é a *linha epipolar*

- Para $\theta = 0$, a linha epipolar é uma reta horizontal na mesma altura, y_1 , do ponto de projeção na primeira imagem

- Neste caso, a solução do problema de reconstrução estereoscópica é simples:

Relação entre os sistemas

$$X_1 = X + \frac{d}{2}, \quad X_2 = X - \frac{d}{2}, \quad Y_1 = Y_2 = Y, \quad Z_1 = Z_2 = Z$$

Projeção perspectiva

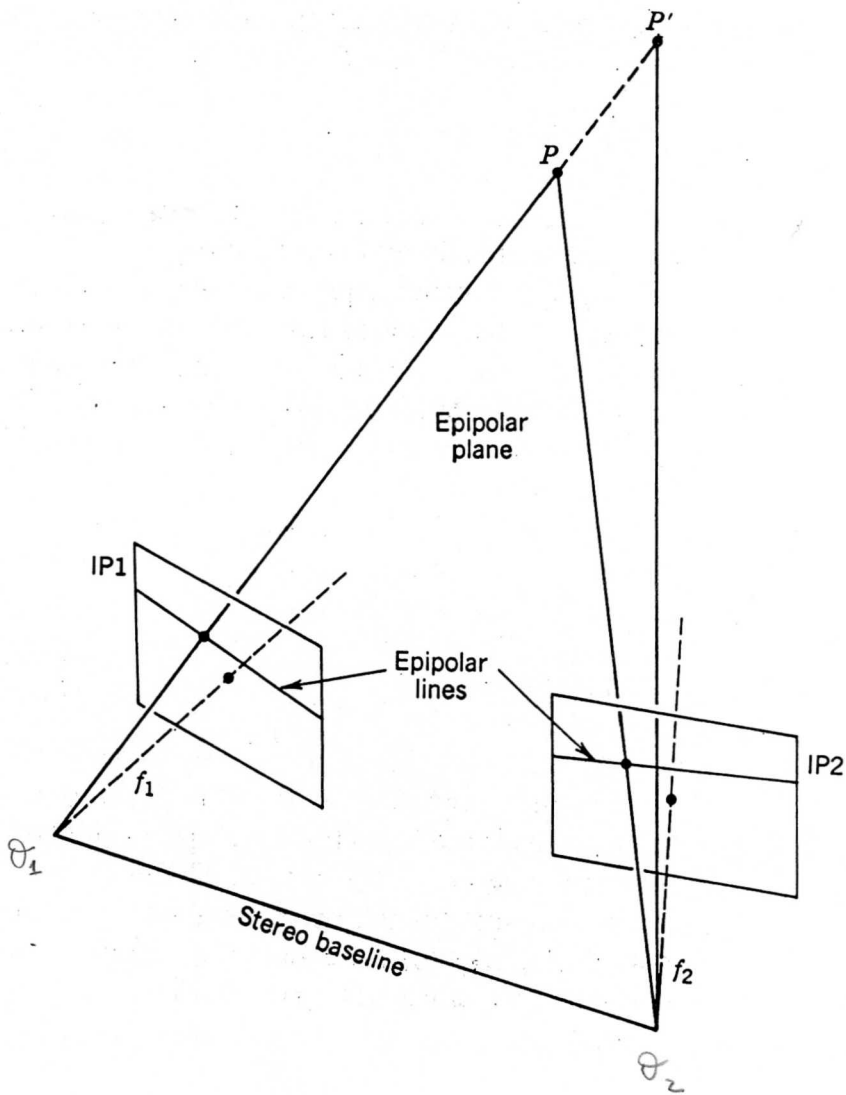
$$\frac{x_1 - x_2}{f} = \frac{X_1 - X_2}{f - Z}$$

De onde se obtém

$$Z = f - \frac{fd}{D}$$

onde $D = x_1 - x_2$ é a *disparidade* entre as projeções nas duas imagens

A solução quando as duas câmeras não são paralelas é mais complicada



Correspondência Estereoscópica

→ Determinar os pontos nas duas imagens que correspondem ao mesmo ponto na cena

• *Correspondência Baseada em Correlação de Áreas*

→ Guiada pela similaridade entre as intensidades em regiões correspondentes das duas imagens

- Extração de pontos característicos:

Escolhem-se pontos em cuja vizinhança os níveis de cinza variem apreciavelmente, de modo que a correspondência não seja ambígua

e.g., Para câmeras horizontais: Aplicação do gradiente horizontal, para selecionar regiões de máxima variação de intensidade

- Avaliação da Similaridade:

Escolhida uma região em torno de um ponto característico na primeira imagem, procura-se a região correspondente na segunda imagem, comparando-se as intensidades dentro de uma certa janela

Medidas de Similaridade:

i) Correlação das Intensidades

$$C = \frac{\sigma_{12}}{(\sigma_1^2 \sigma_2^2)^{1/2}}$$

Onde,

$$\sigma_k^2 = \frac{1}{mn} \sum_{j=1}^n \sum_{i=1}^m (I_k(i, j) - \mu_k)^2$$

é a *variância* das intensidades numa janela de $m \times n$ píxeis centrada no ponto de interesse da imagem k , e

$$\sigma_{kl} = \frac{1}{mn} \sum_{j=1}^n \sum_{i=1}^m (I_k(i, j) - \mu_k)(I_l(i, j) - \mu_l)$$

é a *covariância* das imagens k e l na janela considerada

μ_k é a intensidade média da imagem k na janela considerada

$\rightarrow C$ cresce com a similaridade (valor máximo 1)

ii) Diferença das Intensidades

$$D = \sum_{j=1}^n \sum_{i=1}^m |I_1(i, j) - I_2(i, j)|$$

ou

$$D = \sum_{j=1}^n \sum_{i=1}^m (I_1(i, j) - I_2(i, j))^2$$

$\rightarrow D$ decresce com a similaridade

Compensação do contraste

$$D = \sum_{j=1}^n \sum_{i=1}^m (I_1(i, j) - I_2(i, j))^2 / \sigma_1 \sigma_2$$

ou

$$D = \sum_{j=1}^n \sum_{i=1}^m (I_1(i, j) - I_2(i, j))^2 / \sigma_1^2$$

Compensação do brilho

$$\begin{aligned} D &= \sum_{j=1}^n \sum_{i=1}^m (\{I_1(i, j) - \mu_1\} - \{I_2(i, j) - \mu_2\})^2 / \sigma_1^2 = \\ &= 2 - 2C \end{aligned}$$

Questão: Tamanho da Janela

- Janelas Maiores

→ Profundidade global

→ Menor sensibilidade a ruído

- Janelas Menores

→ Maior precisão

→ Menor carga computacional

● *Correspondência Multiescala*

→ Emprega imagens de menor resolução para uma correspondência grosseira, e imagens de maior resolução para o ajuste mais fino

→ Imagem de baixa resolução é obtida tomando-se a média das intensidades em blocos da imagem original

→ O custo computacional é reduzido: A busca para a correspondência de maior resolução fica restrita pelo casamento grosseiro

Exemplo:

Processo de dois estágios para correspondência horizontal

- Imagem $N \times N$

- Em um único estágio:

→ N^3 cálculos: N^2 pontos a casar, N casamentos possíveis por ponto

- Em dois estágios:

Assumindo que M^2 pontos tornam-se um único ponto na imagem de baixa resolução

i) Primeiro estágio:

→ N^3/M^3 cálculos: N^2/M^2 pontos, N/M casamentos possíveis

ii) Segundo estágio:

→ N^2M cálculos: N^2 pontos, M casamentos possíveis

(A busca é restrita ao intervalo de comprimento M em torno do ponto obtido no estágio anterior)

→ Redução da carga computacional:

$$\frac{N^3/M^3 + N^2M}{N^3} = \frac{1}{M^3} + \frac{M}{N} \ll 1$$

- *Correspondência Baseada em Características*

→ Casa feições extraídas das duas imagens

→ Mais rápida e menos sensível a variações fotométricas

→ Requer interpolação

Exemplo: *Algoritmo de Marr-Grimson-Poggio*

→ Baseado em *zero-crossings*: Casa os contornos de diferentes resoluções obtidos nas duas imagens (Resolução: $w = 2\sqrt{2}\sigma$)

Zero-crossings podem ser caracterizados pelo sinal da variação de intensidade, e por sua orientação espacial aproximada

Casamento:

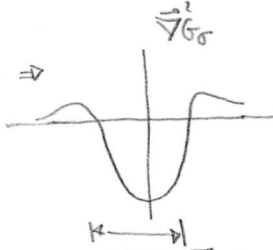
- Contornos para w grande dão correspondência grosseira, e contornos para w pequeno dão ajuste fino

- Cada casamento começa com a disparidade obtida na resolução anterior

- Região de busca para as disparidades: $-w \leq D \leq w$

- Casam-se contornos cujas orientações não diferem em mais de 30°

- Se há ambigüidades, escolhe-se o casamento na categoria dominante da vizinhança: $D > 1$, $-1 \leq D \leq 1$, ou $D < -1$

$$\vec{\nabla}^2 [G_\sigma * f] \equiv \vec{\nabla}^2 G_\sigma * f \quad \text{ou} \quad \vec{\nabla}^2 G_\sigma \Rightarrow$$


$x = (u' + v')^2/2$

$w = 2\sqrt{2}\sigma$

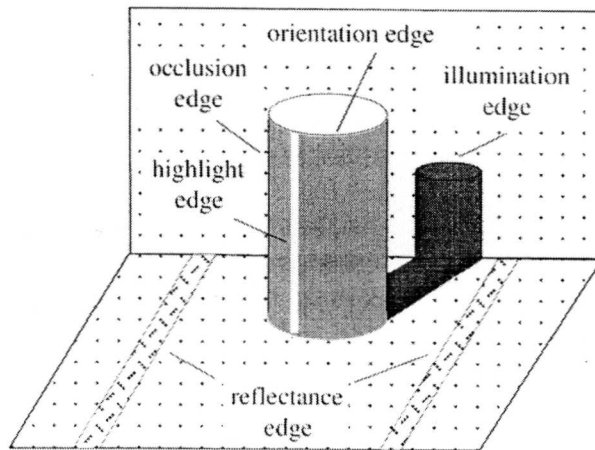
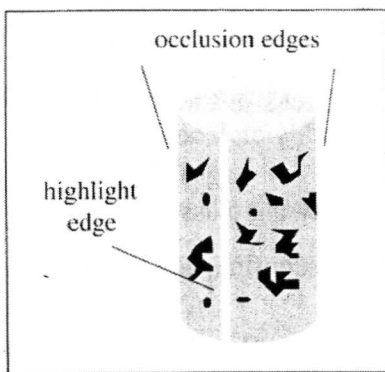
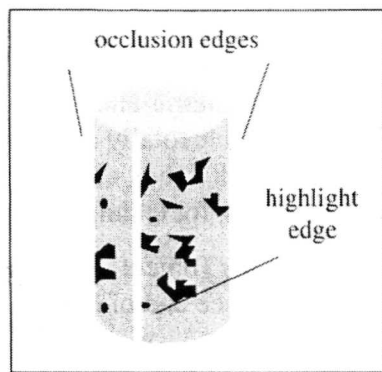


Figure 4.3: Classification of edges in an image of a scene.

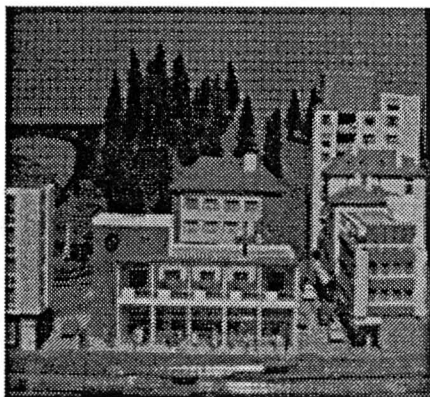
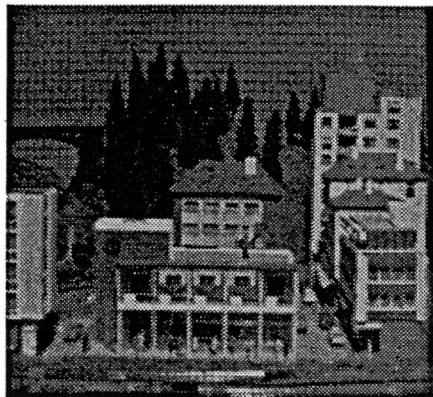


left stereo image

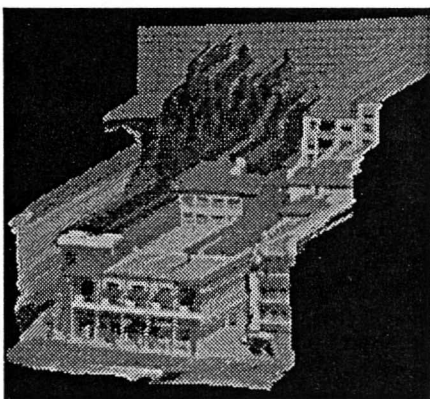
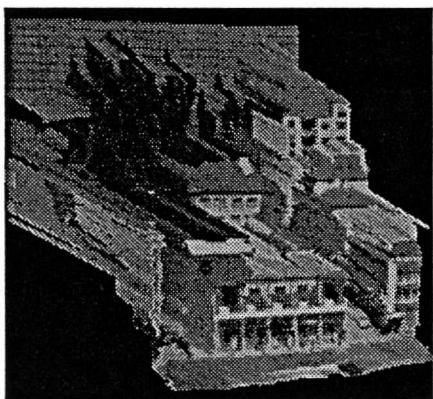


right stereo image

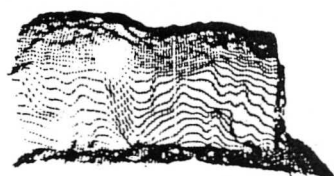
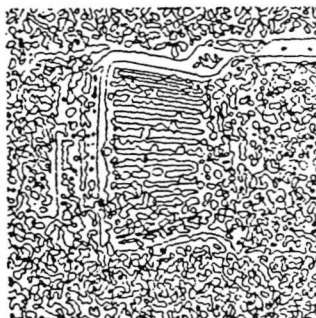
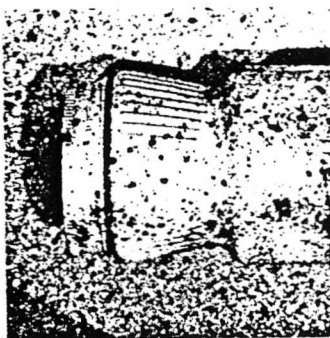
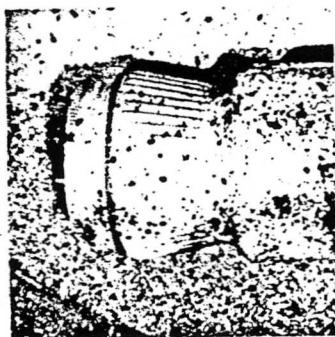
Figure 4.4: Example of occlusion edges and highlight edges in a pair of stereo images that depend on the position of the camera.



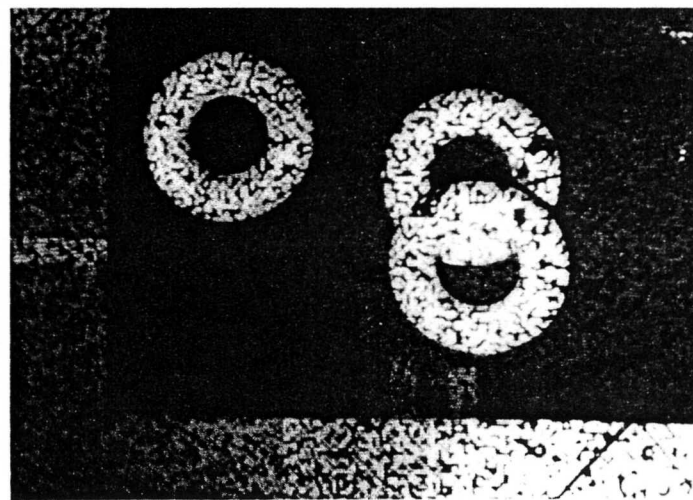
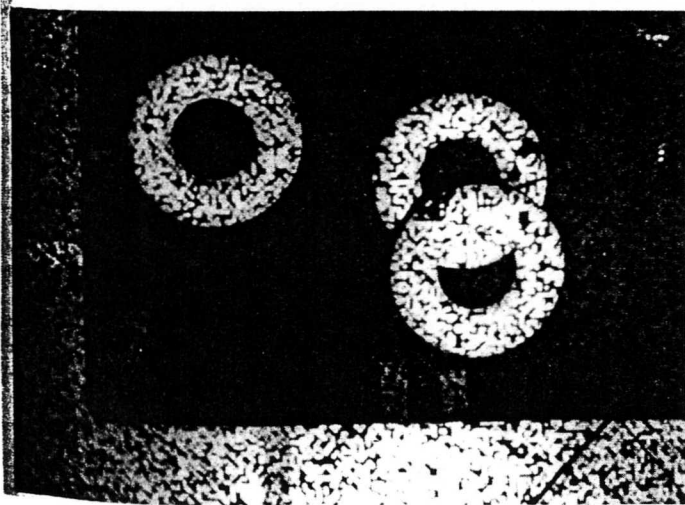
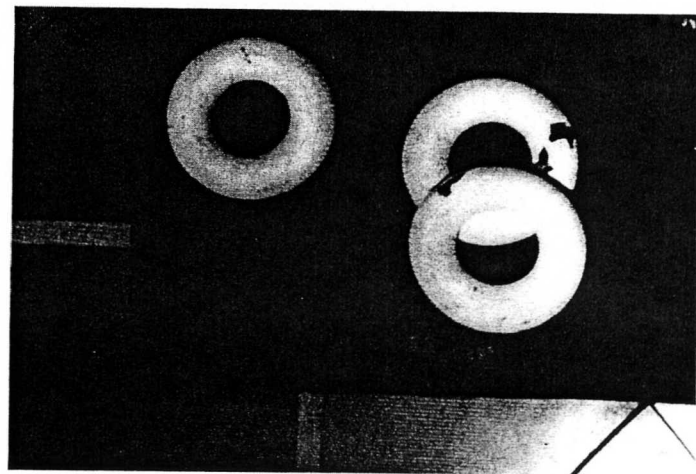
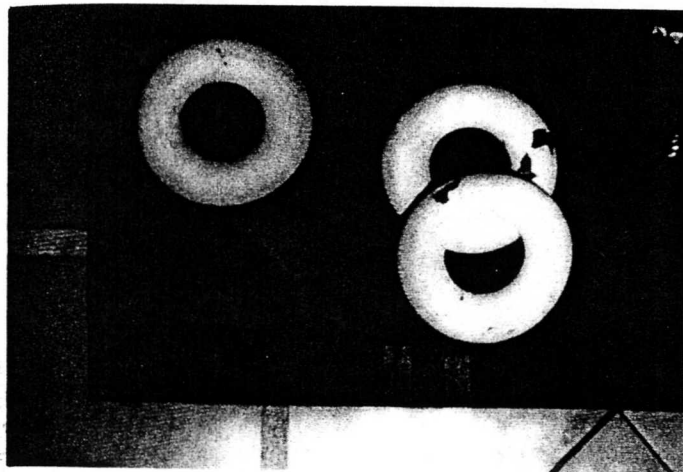
(a)

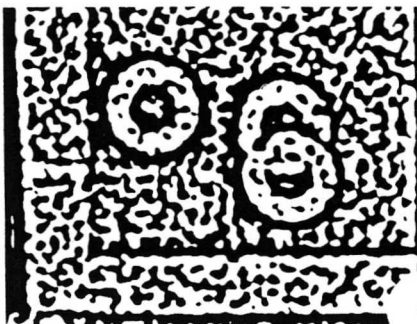


(b)

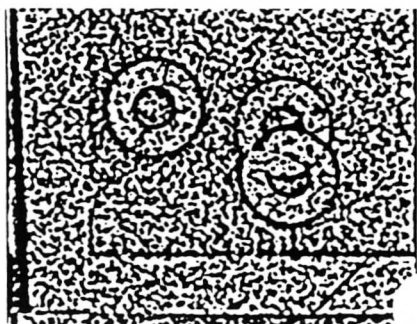
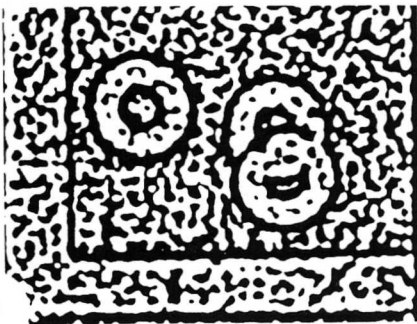


c

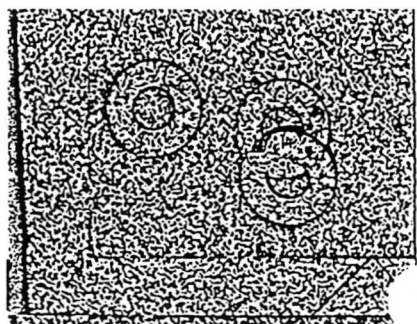
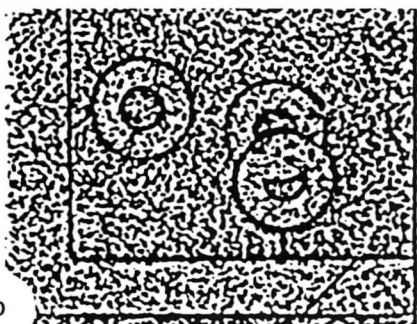




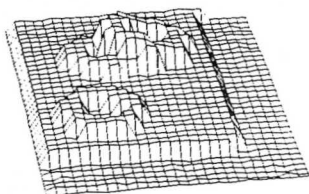
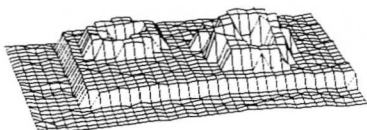
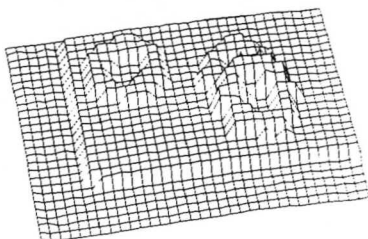
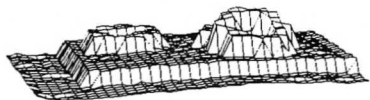
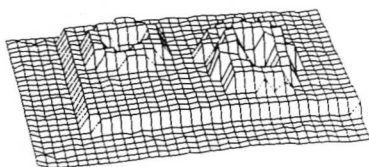
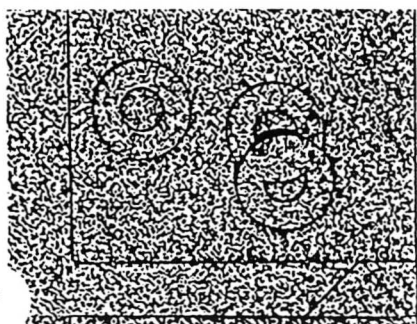
a

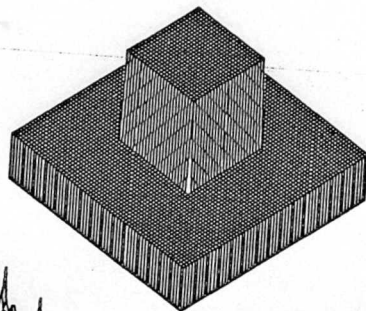


b

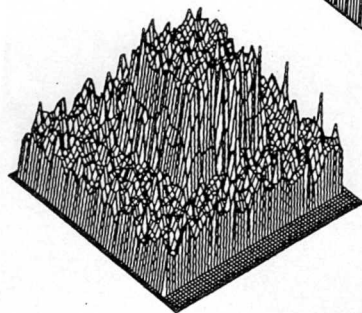


c

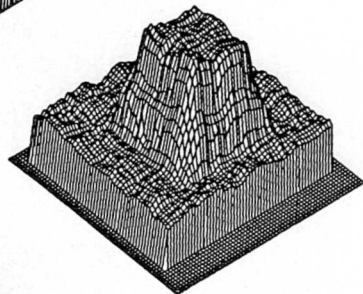




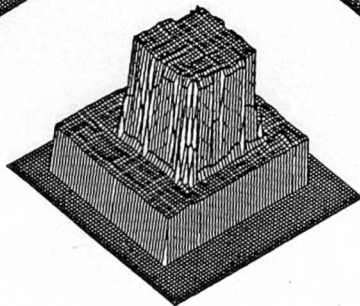
(a)



(b)



(c)



(d)

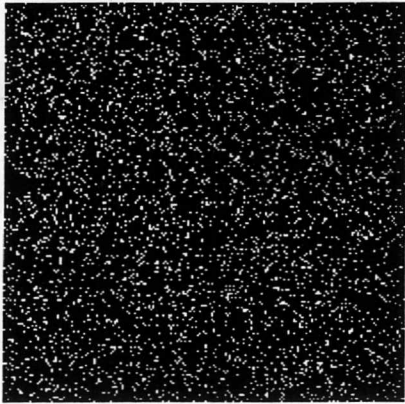


Figure 3a

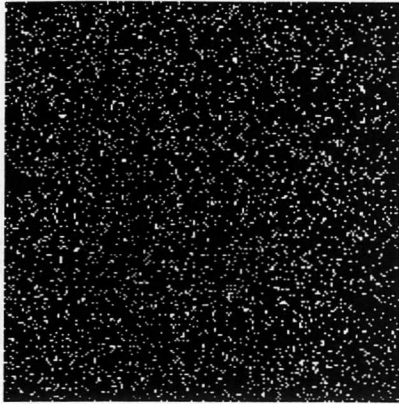


Figure 3b

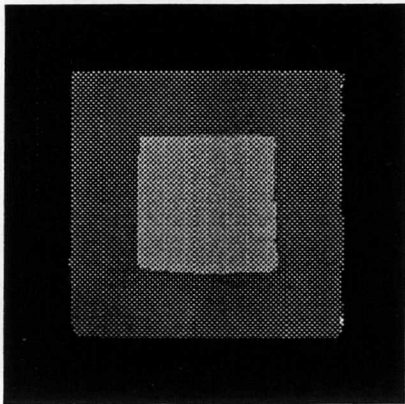


Figure 3c

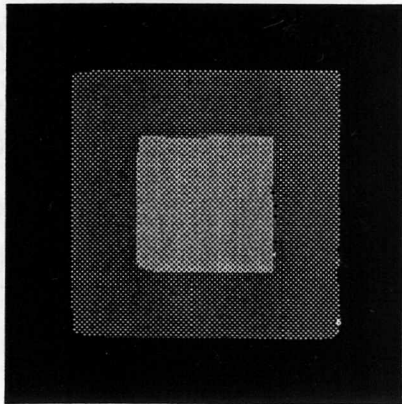


Figure 3d