

LABORATÓRIO  
**NatalNet**

# ROBOTICA PERCEPTUAL: PRINCIPAIS TENDENCIAS

## Visão Robótica

Luiz Marcos Garcia Gonçalves



# Sumário

- Introdução à visão de baixo nível
- Visão Robótica
- Extração de Features
- Redução de dados (foveamento)
- O modelo de foveamento MR
- Abstração de dados (características)
- Aplicações para a estrutura foveada
- Considerações e sugestões de leituras

# Introdução à visão de baixo nível

- Visão de baixo nível refere-se às etapas iniciais dos sistemas de visão computacional ou visão robótica.
  - Pré-processamento de imagens para fornecer dados mais adequados às etapas posteriores
- *Low-level Vision*

# Introdução à visão de baixo nível

- Requisitos de um sistema de visão robótica devem ser definidos:
  - selecionar uma região de interesse;
  - manter a atenção em uma mesma região, de acordo com a necessidade da tarefa
  - mudar o foco de atenção para outra região à medida que uma ação ou uma tarefa tenha sido terminada.
- Executar várias tarefas (reconhecimento, identificação, percepção) com base nas mesmas características, do buffer visual

# Introdução à visão de baixo nível

- Problema principal da visão de baixo nível:
  - Aplicações de visão robótica, exigem processamento em tempo real e requerem movimentação de câmeras.
- Solução:
  - Reduzir a quantidade de dados providos pelas câmeras e minimizar os movimentos
- Estrutura em multiresolução multifoveada reduz quantidade de dados e possibilita extrair (abstrair) características em tempo real, em vários pontos.

## Introdução à visão de baixo nível

- Possibilitar que experimentos como seguir vários objetos (tracking) e outros usando cálculo de movimento e visão estéreo possam ser implementados visando validar e aprender de melhor forma as características do mundo em volta.

# Introdução à visão de baixo nível

- Construção de módulos de visão de mais baixo nível para serem usados como base em funções de mais alto nível (servidor visual?).
- Permitir a um robô tomar decisões baseadas na informação contida no seu estado perceptual e nos parâmetros da tarefa sendo executada correntemente, selecionando as ações corretas em resposta aos estímulos providos pelo seu entorno.

# Visão Biológica

- A visão é o sentido que provê mais e melhores informações, sendo também o mais complexo permitindo uma interação mais experta com o ambiente
  - Sistemas visuais biológicos são complicados

# Visão Robótica

- Técnicas de visão computacional (visão robótica) visam a obtenção de informações relacionadas com a posição de objetos em ambientes ou cenas
- Modelos computacionais são construídos a partir de dados digitalizados (imagens).
- Imagens são obtidas por câmeras filmadoras, radar, equipamentos de ressonância magnética, ultrassonografia, dispositivos RGB-D, scanners 3D e outros.

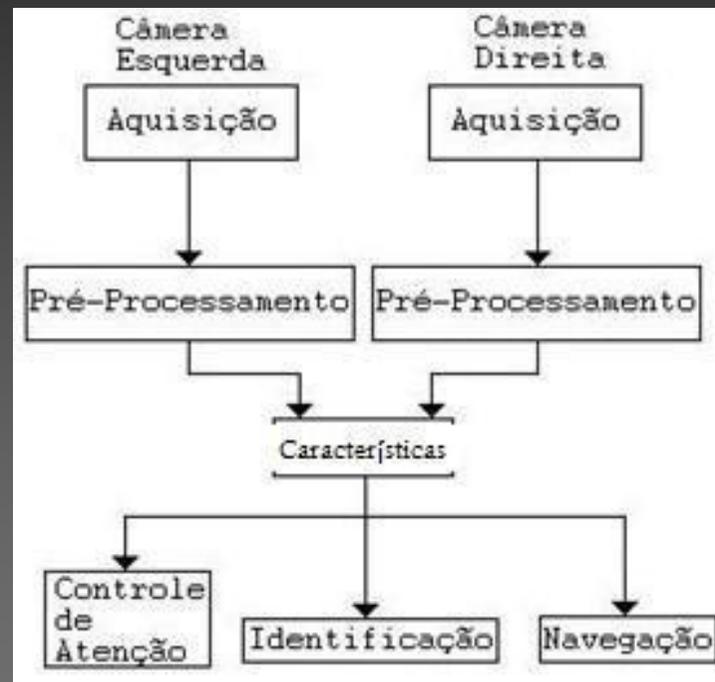
# Visão Robótica

- Propósitos da visão artificial:
  - Controle ou identificação
    - Processamento a posteriori
    - Mapeamento
  - Tomada de decisão em tempo real
    - Processamento imediato
    - Navegação em um ambiente

## Visão robótica

- Visão ativa (reativa):
  - Informação (posição, forma, cor) é fornecida em tempo real a um sistema tomador de decisões
- Coordenar a execução de tarefas em tempo real, permitindo que o agente robótico mude o seu comportamento

# Visão robótica



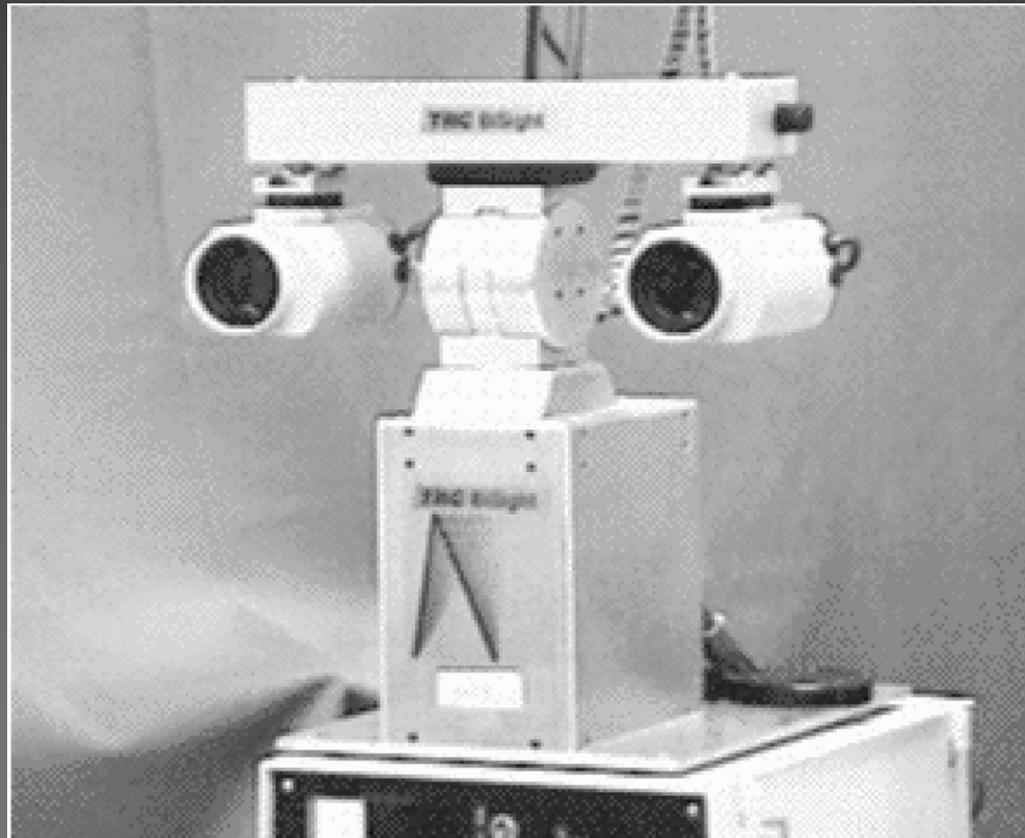
- Fluxo de dados em visão robótica

# Visão robótica

- Aquisição das imagens
  - câmeras e dispositivos RGB-D
- Pré-processamento
  - provê a redução de dados (estrutura foveada em multiresolução)
- Extração de características
  - disparidade estéreo, movimento, intensidade e derivadas gaussianas, SURF, DRINK, ...
- Execução de tarefas
  - atenção, reconhecimento, navegação

# Visão robótica

- Cabeça estéreo da UMASS



# Visão robótica

- Cabeça estéreo da UFRN

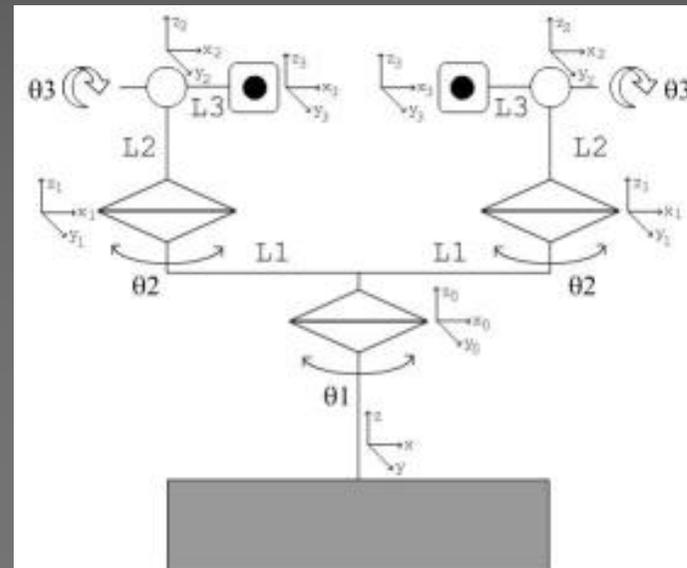


# Visão robótica

- Controle de movimentos obedece restrições bem definidas (teoria de controle)
- Modelo cinemático
  - conjunto de corpos rígidos, chamados elos
  - interligados numa cadeia cinemática aberta através de juntas
  - uma extremidade fixa na base e uma câmera na outra extremidade (móvel)
- Modelo dinâmico
  - forças e massas (aceleração, velocidade, posição)

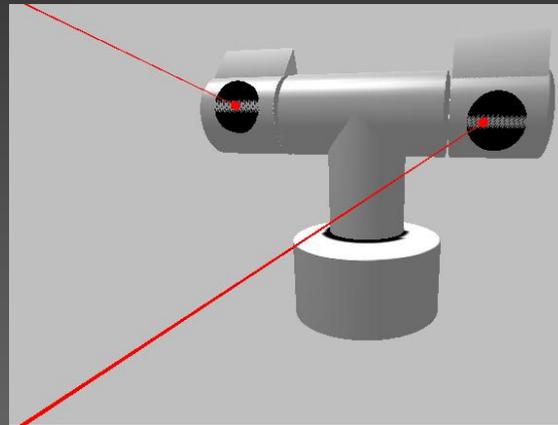
# Visão robótica

- Graus de liberdade (5)
  - em torno do eixo vertical (pescoço ou *pan*)
  - em torno do eixo horizontal de cada câmera (inclinação ou tilt)
  - em torno do eixo vertical de cada câmera (vergência)

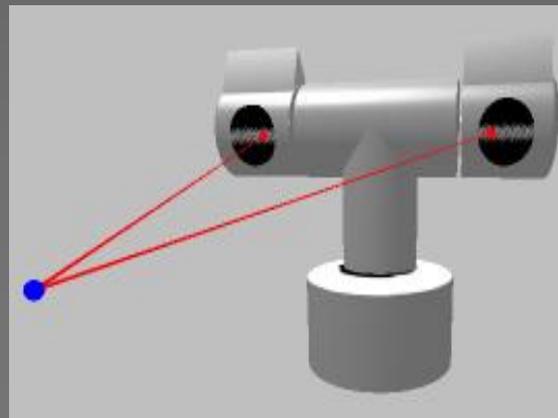


# Visão robótica

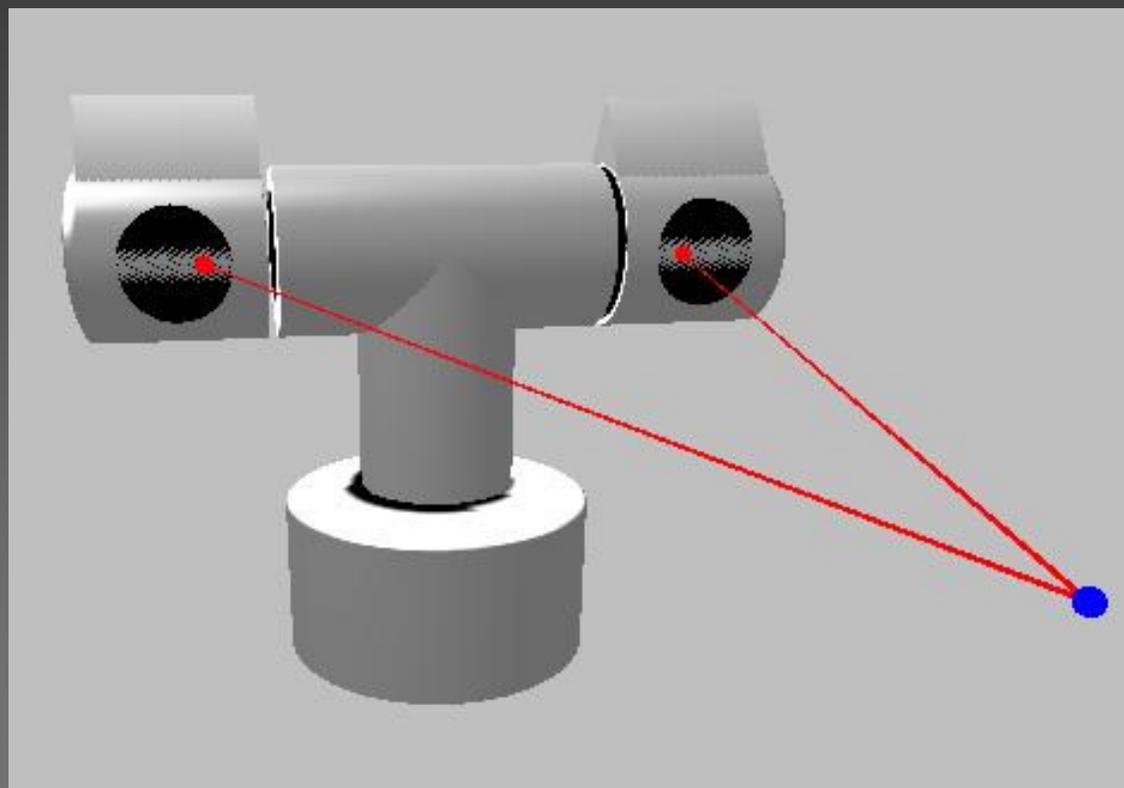
- Movimentos independentes



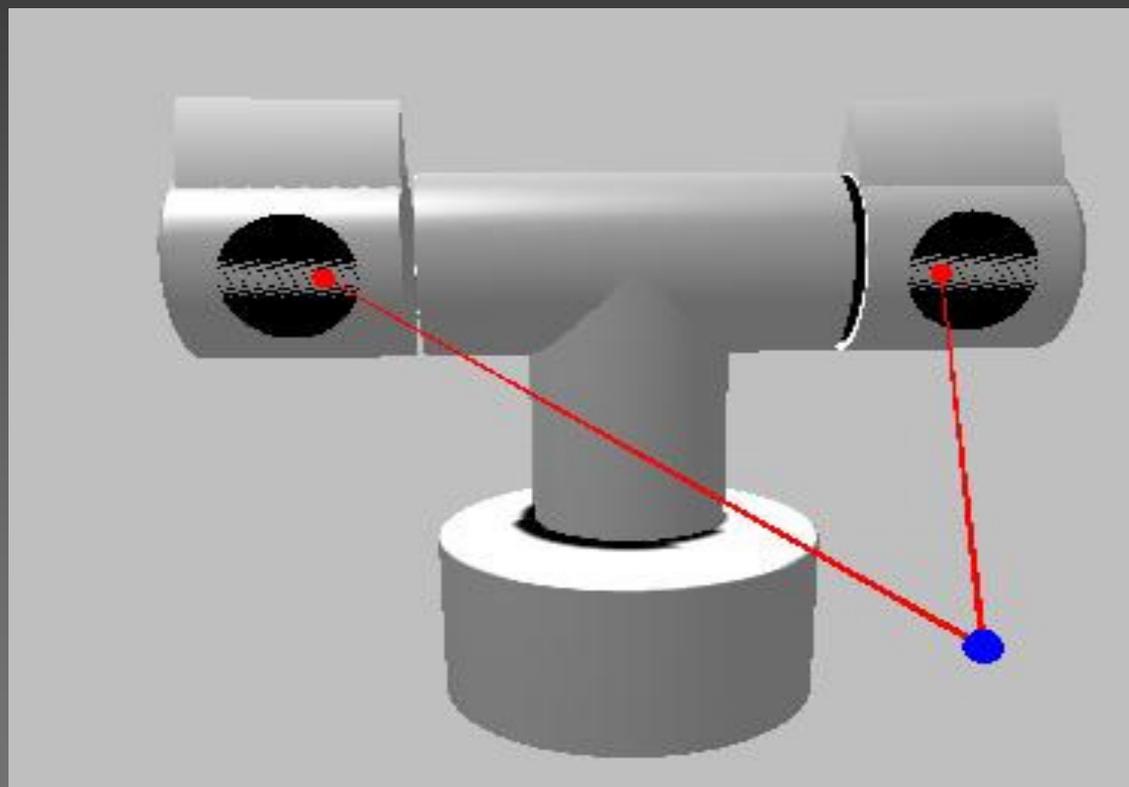
- Movimentos dependentes



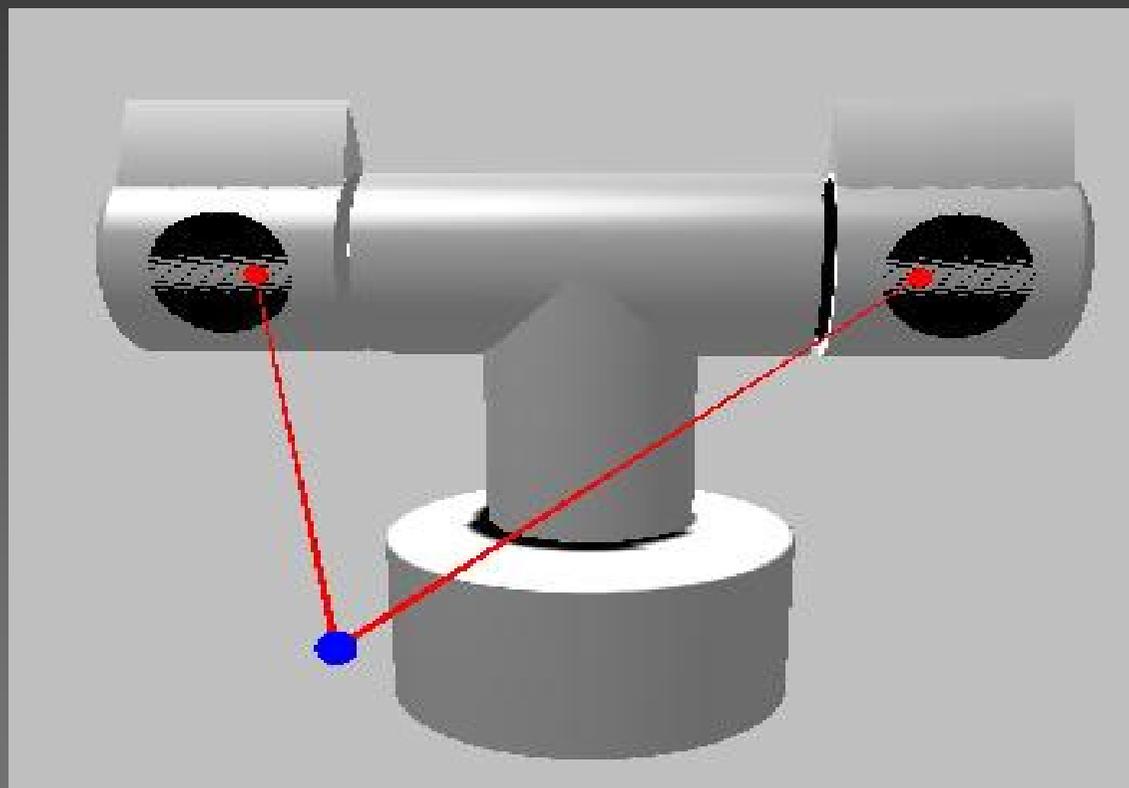
# Visão robótica



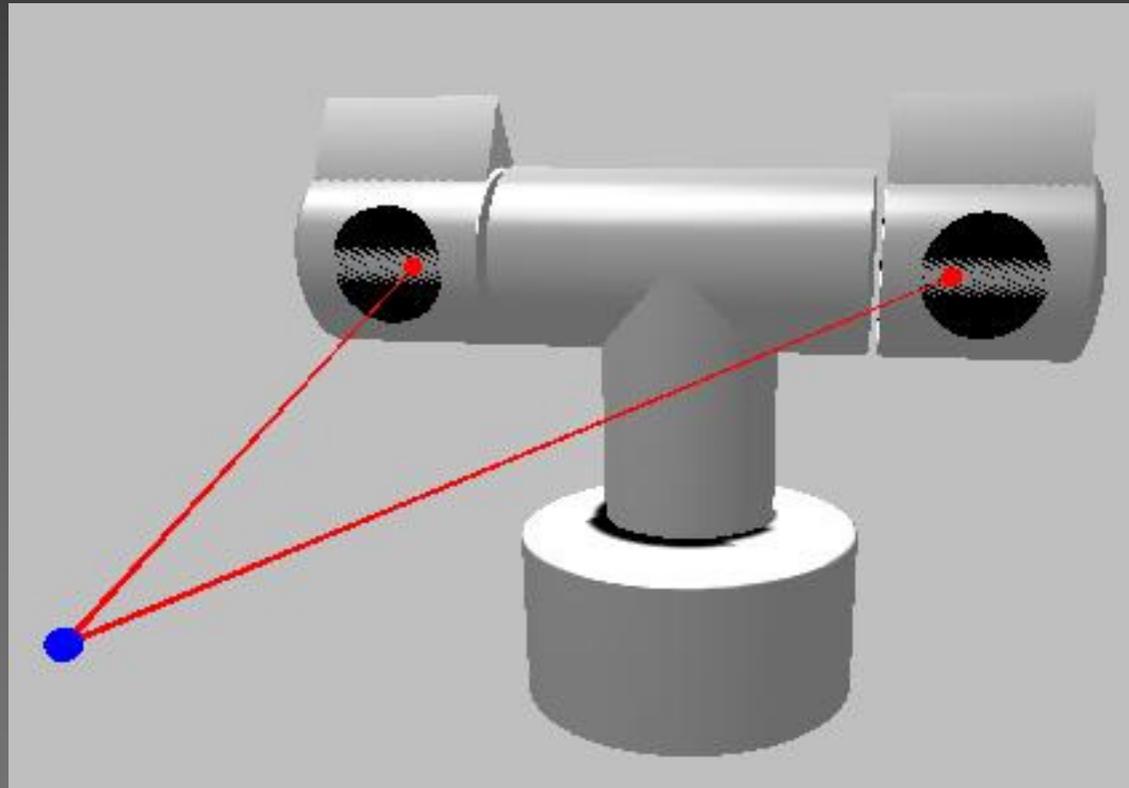
# Visão robótica



# Visão robótica



# Visão robótica



# Features? O que são?

- Conceito de features (características)
  - Detectáveis (posição, orientação)
  - Tem um significado (podem ser medidas)
- Arestas, contornos, curvas, cantos, SIFT, SURF, DRINK, ...
- Features não geométricas

# Definição

- Features (características) de imagem são partes detectáveis, locais da imagem que sejam significativos (meaningful)
- Sequência de operações iniciais dos sistemas de visão (análise de imagens):
  - Detecção (realce) de features
  - Localização de features

# Devem ser significativas

- Associadas a elementos de interesse na cena via o processo de formação de imagens (valor):
  - Variação alta da intensidade causada pelos contornos
  - Regiões com nível de intensidade uniforme
- Algumas vezes procuramos features que não se traduzem a alguma característica na imagem mas refletem arranjos particulares dos pixels com certas propriedades desejadas (invariância, coisas fáceis de detectar).

# Propriedades desejáveis

## ■ Invariância

- A feature deve ser localizada, mesmo sob ponto de vista e escala diferentes (invariância afim)

## ■ Repetibilidade

- Localização do mesmo conjunto de *features*, mesmo com mudanças substanciais do ponto de vista da cena (relaciona-se com invariância)

# Devem ser Detectáveis

- Deve existir um algoritmo de localização, caso contrário, a feature não serve para nada
- Features diferentes são associadas a algoritmos de detecção diferentes;
- A saída geralmente são os descritores das features, especificando posição e outras propriedades essenciais da feature.
  - Ex: descritor de linha contém o seu ponto central, seu tamanho e orientação.

# Mais propriedades desejáveis

## ■ Distintibilidade

- Permitir distinção entre elementos (features diferentes)

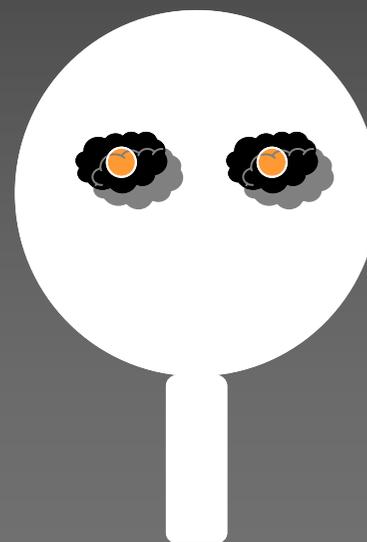
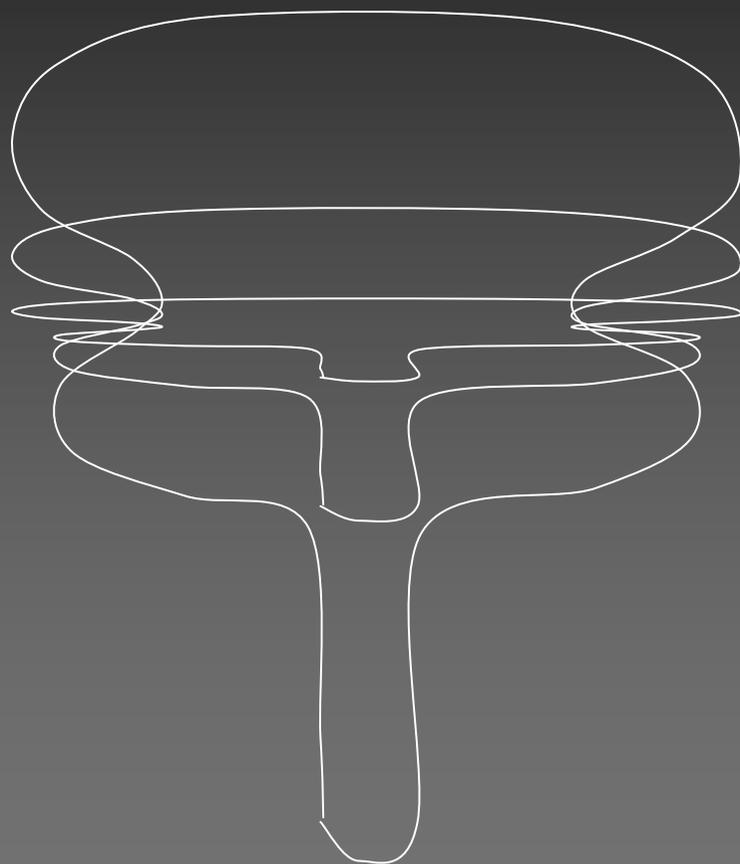
## ■ Robustês

- Imune a ruídos, deslocamento e deformações geométricas e fotométricas

## Arestas (bordas)

- São pixels (ou regiões) da imagem onde o valor da intensidade possui uma variação brusca
- Problema de detecção de arestas:
  - dada uma imagem (com erro), localizar as arestas geradas pelos elementos da cena (não pelo erro)

# Por que aresta importa?

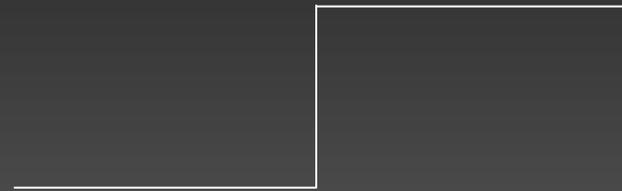


# Algoritmo

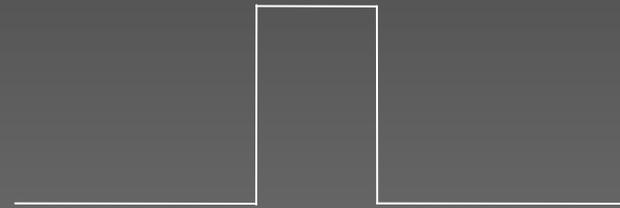
- Suavização do ruído
  - suprimir ruídos sem destruir arestas
  - assumir ruído branco ou gaussiano
- Realce de arestas
  - projetar um filtro que responda a arestas
  - valor alto nas arestas e baixo fora delas
- Localização de arestas
  - decidir que máximos locais no filtro são arestas
    - afinar arestas grossas para um pixel de largura
    - estabelecer um valor mínimo para aresta (treshold)

# Arestas

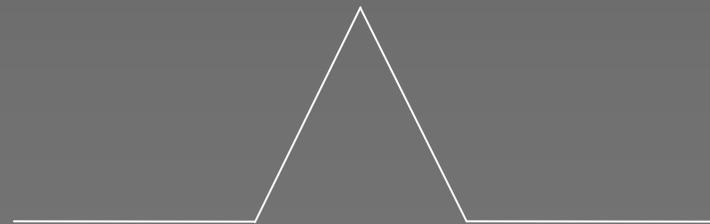
- Step edges



- Ridge edges



- Roof edges



# Convolução

## Early Processing - Convolution

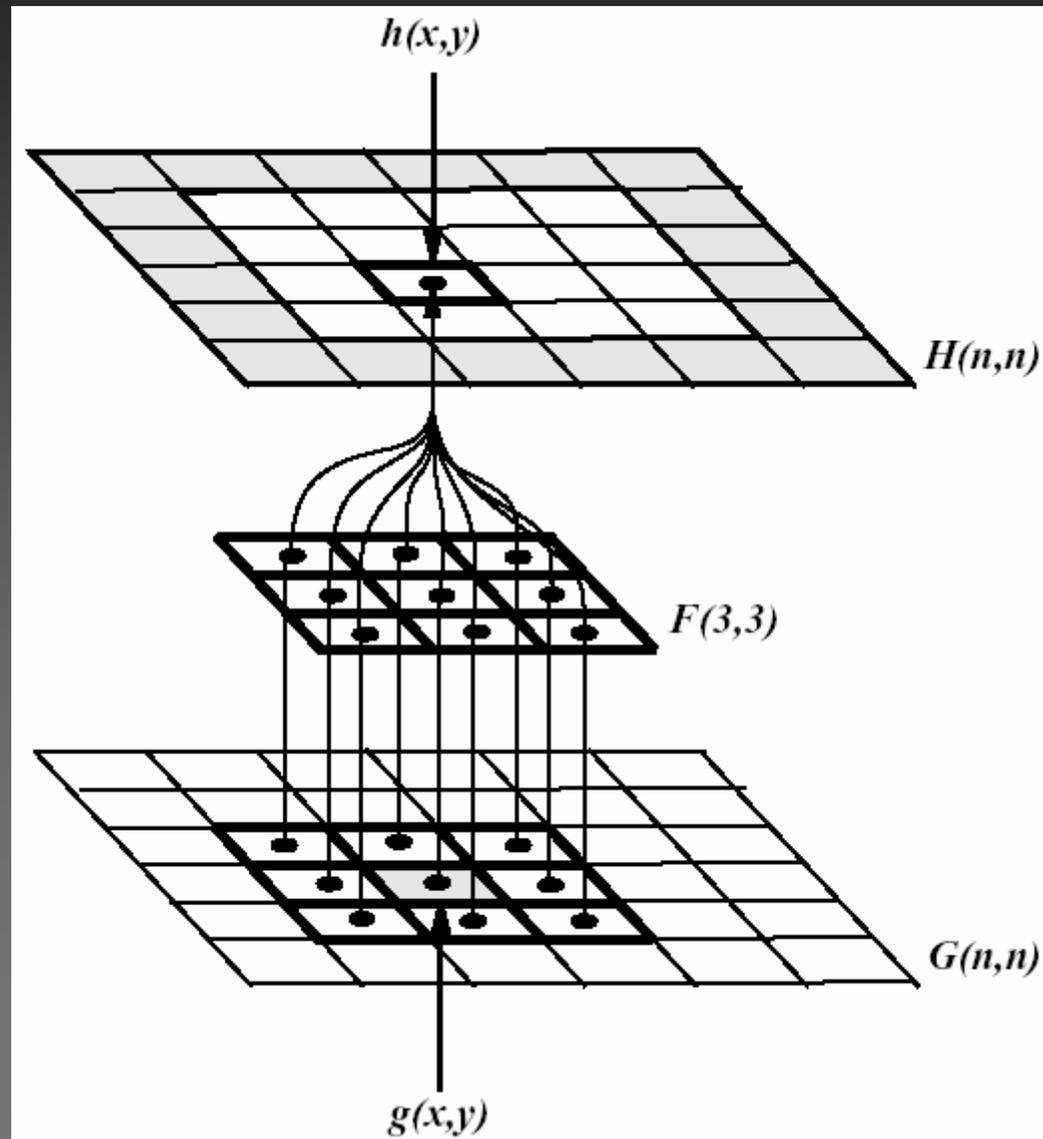
The convolution of two functions  $f(x)$  and  $g(x)$ :

$$f(x, y) * g(x, y) = h(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u, v)g(x - u, y - v)dudv$$

or

$$f(x, y) * g(x, y) = h(x, y) = \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} f(u, v)g(x - u, y - v)$$

# Convolução





# Detecção de arestas

- Magnitude do gradiente

$$|\nabla g| = \left[ \left( \frac{dg}{dx} \right)^2 + \left( \frac{dg}{dy} \right)^2 \right]^{\frac{1}{2}}$$

- Direção de mudança da intensidade

$$\phi = \tan^{-1} \left( \frac{dg/dy}{dg/dx} \right)$$

# Arestas (gradiente)

Image Filtering: Edge Detection



horizontal derivative



vertical derivative

# Detectores de bordas

operator	$\nabla_1$	$\nabla_2$
Roberts	$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$
Prewitt	$\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$
Sobel	$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$

# Descritor de arestas

- Normal
- Direção
- Posição ou centro
- Magnitude (strength) - contraste

## Saída do algoritmo

- Conjunto de listas de arestas, cada uma descrevendo a posição de um contorno conectado na imagem
- Imagem de orientação
- Imagem de magnitude

# Operadores Gaussianos

$$g_{\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}$$

$$g'_{\sigma}(x) = \frac{-x}{\sqrt{2\pi}\sigma^3} e^{-x^2/2\sigma^2}$$

$$g''_{\sigma}(x) = \frac{1}{\sqrt{2\pi}} \left[ \frac{x^2}{\sigma^5} - \frac{1}{\sigma^3} \right] e^{-x^2/2\sigma^2}$$

# Detectando cantos (Harris)

- Considere o gradiente espacial da imagem dado por  $(E_x, E_y)^t$ , com  $E_x = \partial E / \partial x$  e  $E_y = \partial E / \partial y$ ;
- Considere um pixel  $p$  genérico na imagem, uma vizinhança  $Q$  de  $p$  e uma matriz  $C$ :

$$C = \begin{pmatrix} \sum E_x^2 & \sum E_x E_y \\ \sum E_x E_y & \sum E_y^2 \end{pmatrix}$$

onde as somatórias são tomadas sobre a vizinhança  $Q$ .

- A matriz  $C$  caracteriza a estrutura dos níveis de cinza da imagem, em cada pixel.

# Auto-valores e auto-vetores

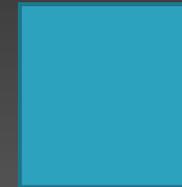
- A matriz  $C$  é simétrica, portanto pode ser diagonalizada por uma rotação do eixo de coordenadas, sem perda de generalidades:

$$C \stackrel{Q}{=} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

Os dois auto-valores  $\lambda_1$  e  $\lambda_2$  são não negativos (assumamos  $\lambda_1 > \lambda_2$ ).

# Interpret. Geomét. de $\lambda_1$ e $\lambda_2$

- 1) Considere uma vizinhança  $Q$  uniforme: o gradiente desaparece em toda a vizinhança,  $C$  vira uma matriz nula e  $\lambda_1 = \lambda_2 = 0$

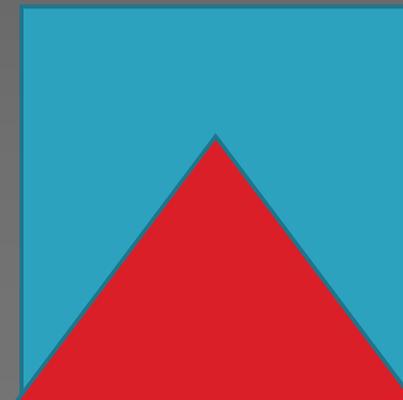


- 2) Assuma que  $Q$  contém um degrau ideal preto para branco; temos então  $\lambda_2 = 0$ ,  $\lambda_1 > 0$  e o auto-vetor associado com  $\lambda_1$  é paralelo ao gradiente da imagem
- O posto (rank) de  $C$  é deficiente nos dois casos (0 e 1, respectivamente).



## Interpret. Geomét. de $\lambda_1$ e $\lambda_2$

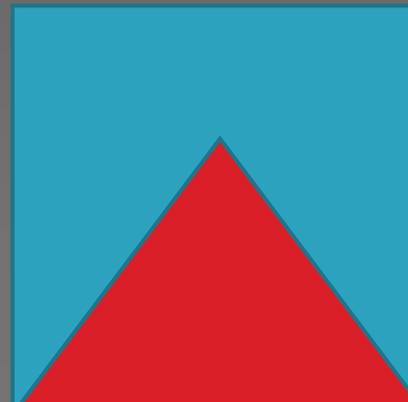
- 3) Considere que  $Q$  contém o canto de um quadrado vermelho contra um fundo verde: como tem duas direções principais em  $Q$ , espera-se que  $\lambda_1 > \lambda_2 > 0$  e que quanto maior forem, mais forte (maior contraste) serão as linhas da imagem correspondente
- Posto de  $C$  é 2



## Interpret. Geomét. de $\lambda_1$ e $\lambda_2$

- Auto-vetores codificam direções das arestas e auto-valores codificam sua magnitude.

Um canto é dado por duas arestas fortes; dados  $\lambda_1$  e  $\lambda_2$ , um canto é uma localização onde o menor deles ainda é grande



# Operadores de Frei e Chen

- Considere o conjunto de operadores:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} -1 & -\sqrt{2} & -1 \\ 1 & \sqrt{2} & 1 \end{bmatrix}$$

$$\begin{bmatrix} -1 & \sqrt{2} \\ 1 & -1 \\ -\sqrt{2} & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & & \\ -1 & & 1 \\ & -1 & \end{bmatrix}$$

$$\begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}$$

$$\begin{bmatrix} -1 & 1 \\ -\sqrt{2} & \sqrt{2} \\ -1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} \sqrt{2} & -1 \\ -1 & 1 \\ & 1 & -\sqrt{2} \end{bmatrix}$$

$$\begin{bmatrix} -1 & 1 \\ & & \\ 1 & -1 \end{bmatrix}$$

$$\begin{bmatrix} -2 & 1 & -2 \\ 1 & 4 & 1 \\ -2 & 1 & -2 \end{bmatrix}$$

k=0

k=1,2

k=3,4

k=5,6

k=7,8

# Operadores de Frei e Chen

- Seja  $g_k = f * h_k$ , então:

$$\text{Edge Energy}(x) = \sum_{k=1}^2 g_k^2(x)$$

$$\text{Total Energy}(x) = \sum_{k=0}^8 g_k^2(x)$$

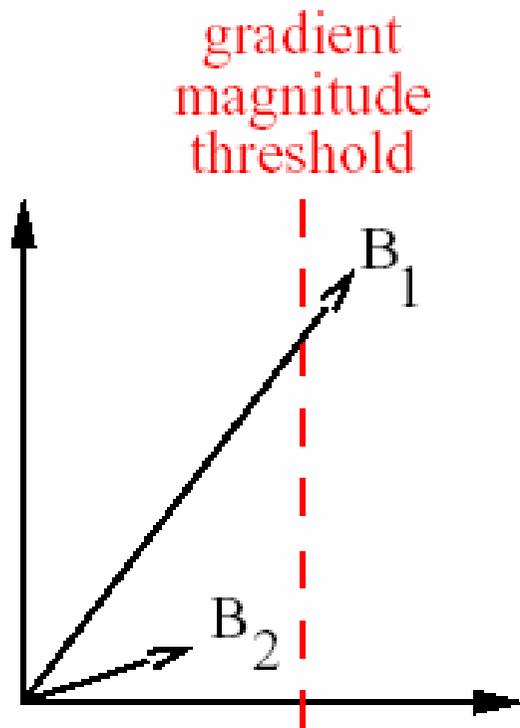
- Definindo

$$\cos(\theta) = \left( \frac{EE}{TE} \right)^{1/2}$$

- Se  $\theta > \text{threshold}$ , reporte/rotule aresta

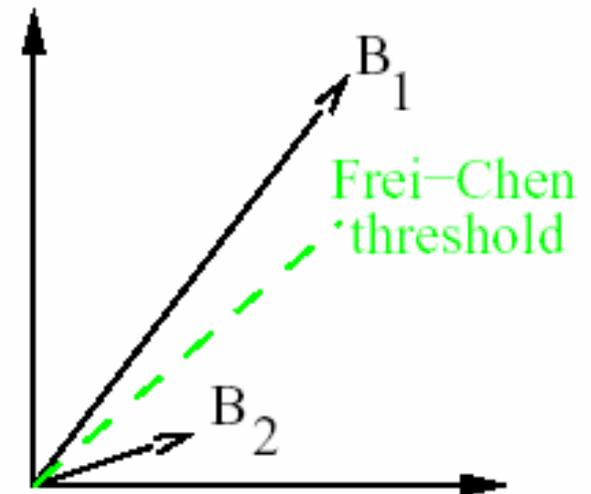
# Gradiente x Frei-Chen

Non-Edge  
Space  
Projection



Edge Space  
Projection

Non-Edge  
Space  
Projection



Edge Space  
Projection

# Características SIFT e SURF

- Idéia:
  - usar algo similar ao operador de Harris em resoluções diferentes
- Invariante à escala e rotação
- SIFT: Lowe
- PCA-SURF: Yan Ke
- SURF: Bay/Luc Van Gool

# Scale Invariant Feature Transform (SIFT)

- Histograma do gradiente local (Harris) orientado, ao redor do ponto de interesse
- Armazenamento em um vetor 128-dimensional
  - ▣ 8 orientações para cada 4 x 4 bins locais

# Speeded Up Robust Features

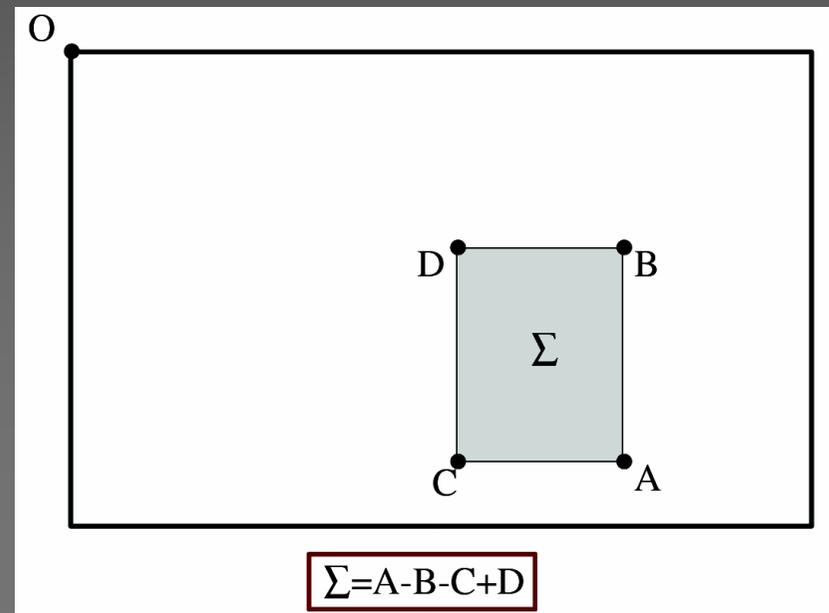
- Dado um ponto  $p = (x, y)$  numa imagem  $I$ , a matriz Hessiana  $H(p, \sigma)$  em  $p$ , na escala  $\sigma$  é definida como:

$$\mathcal{H}(\mathbf{x}, \sigma) = \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{xy}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{bmatrix}$$

- sendo  $L_{xx}(p, \sigma)$  resultado da convolução com o filtro gaussiano de segunda ordem (LoG); o mesmo para  $L_{xy}(p, \sigma)$  e  $L_{yy}(p, \sigma)$

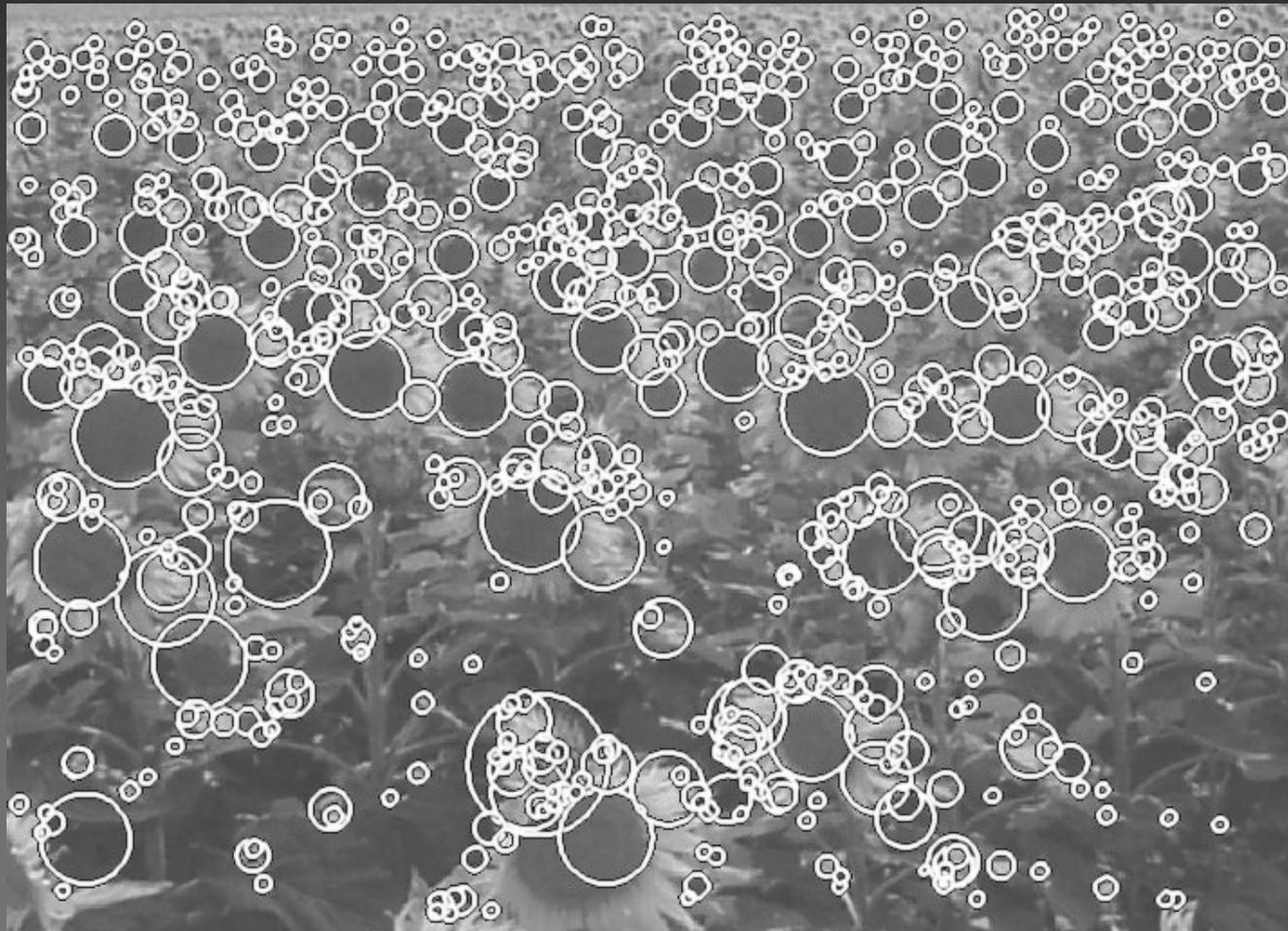
# Speeded Up Robust Features

- Imagem Integral
  - ▣ 3 adições e 4 acessos à memória para calcular a soma das intensidades em uma vizinhança retangular.





# Speeded Up Robust Features



## Redução de dados e abstração de características

- Problemas intrínsecos da extração de características
  - definir o conjunto de características a serem extraídas das imagens
  - definir quais das características serão usadas em uma determinada tarefa
- Cenário ideal:
  - operadores para todas as características a todo ciclo de processamento dentro do tempo desejado (real-time)

## Redução de dados e abstração de características

- Primeiro passo (definir características):
  - disparidade estéreo, movimento, derivadas gaussianas parciais de ordem 0 a 2, momentos estatísticos de ordem 0 a 2, intensidade, SURF, DRINK, ...
- Features computadas para cada ponto nas imagens (muito processamento☺):
  - possível com uso de processadores dedicados, em paralelo
  - não é possível obter todas elas em tempo real se forem usadas as imagens em resolução completa

## Redução de dados e abstração de características

- Uso de alguma técnica para reduzir dados ou facilitar a extração de features parece ser uma boa idéia 😊

## Redução de dados e abstração de características

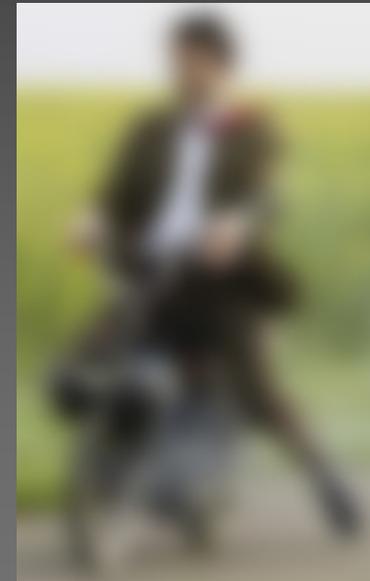
- Uso de alguma técnica para reduzir dados ou facilitar a extração de features é uma boa idéia 😊

## Redução de dados e abstração de características

- Espaço de escalas
  - geração de várias imagens em resoluções piores a cada nível
  - Para todo pixel da imagem (imagens com mesmo tamanho da imagem original)
  - Leonard Uhr (1972), Andrew Witkin, (1983), e Tony Lindeberg (1991)
- Espaço de escala (Witkin)
  - $L(x, y, t) = g(x, y, t) \otimes f(x, y); t \geq 0$
  - $g(x, y, t) = 1/(2\pi t) \exp(-(x^2+y^2)/2t)$

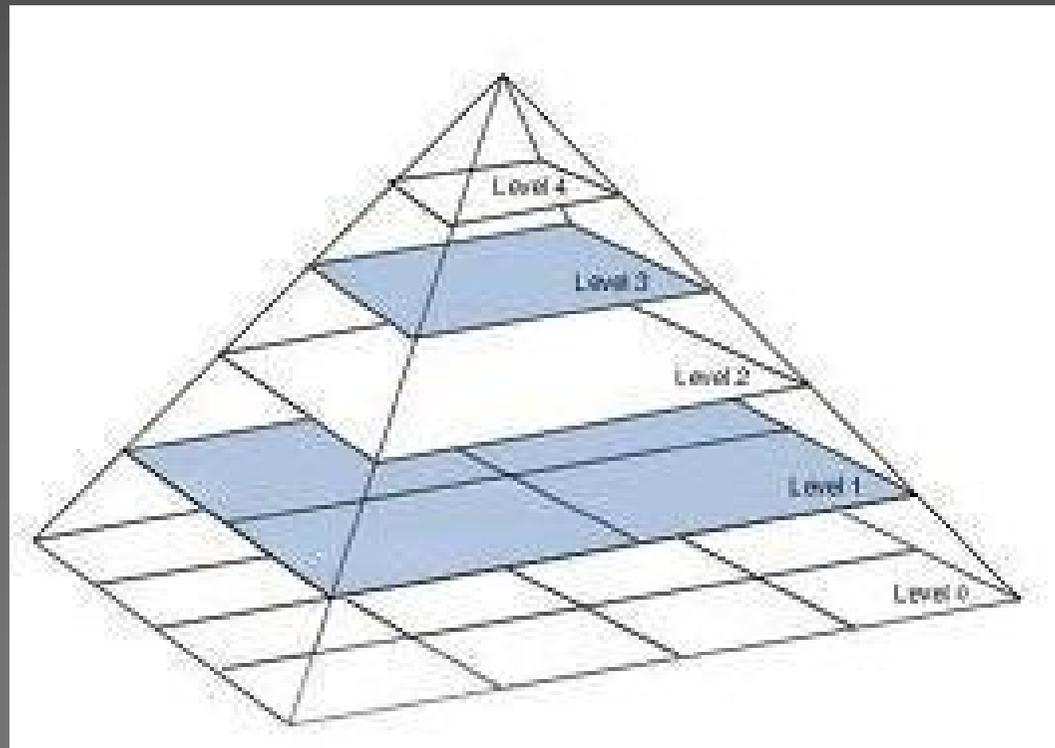
## Redução de dados e abstração de características

- Espaço de escalas para  $t = \{0; 2; 4; 20\}$



## Redução de dados e abstração de características

- Pirâmide em multiresolução
  - pirâmide formada por várias camadas, em resoluções cada vez mais grosseiras



## Redução de dados e abstração de características

- Pirâmide em multiresolução
  - amostra pixels espaçados horizontal e verticalmente, conforme a resolução



## Redução de dados e abstração de características

- Cálculo dos valores dos pixels da pirâmide:
  - reamostragem por filtragem dos pixels vizinhos
  - filtragem, aumentando o suporte do filtro a cada camada
- Filtros mais comuns :
  - média, mediana, gaussiano, laplaciano do gaussiano, e até mesmo wavelets (vistos adiante)

## Redução de dados e abstração de características

- Pirâmide gaussiana:



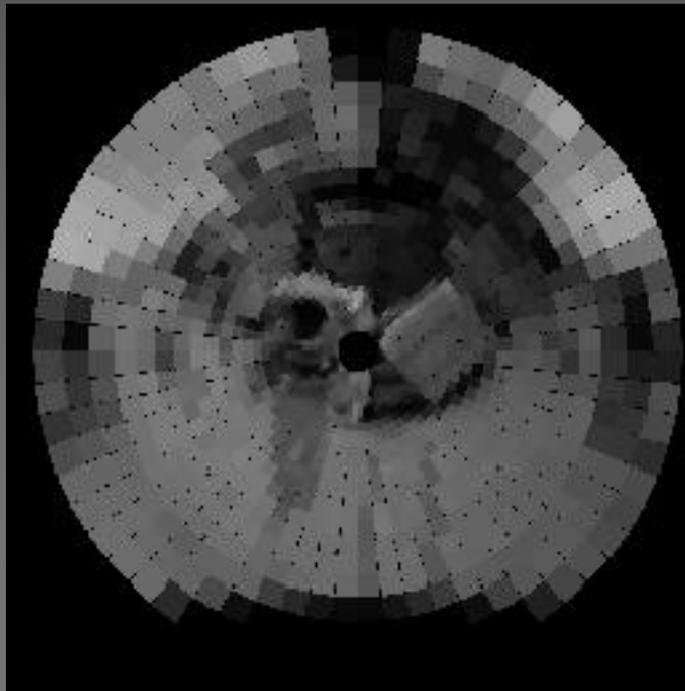
## Redução de dados e abstração de características

- Pirâmide laplaciana:
  - Peter Burt (1988) usa diferenças entre cada nível da pirâmide gaussiana e o seguinte, reamostrado, para determinação de cada nível da pirâmide laplaciana.



## Redução de dados e abstração de características

- Espaços variantes (baseado em log-polar):
  - Sensores com distribuição não uniforme, no hardware de aquisição, concentrando mais sensores na região da fóvea



## Redução de dados e abstração de características

- Onduletas ou *Wavelets*:
  - Funções capazes de decompor e descrever outras funções no domínio da frequência, permitindo analisar estas funções em diferentes escalas de frequência e de tempo.
- Decomposição com wavelets é conhecida como transformada wavelet e tem suas variantes contínua e discreta.

## Redução de dados e abstração de características

- A transformada wavelet discreta mais simples (Haar, 1909) consiste em identificar os parâmetros  $c_k$  e  $d_{k,j}$ ,  $k \in \mathbb{N}$ ,  $j \in \mathbb{N}$ , da equação :

$$f(t) = \sum_{k=-\infty}^{\infty} c_k \phi(t - k) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} d_{k,j} \psi(2^j t - k)$$

$\phi(t)$  e  $\psi(t)$  são as funções wavelet *pai* e *mãe*

## Redução de dados e abstração de características

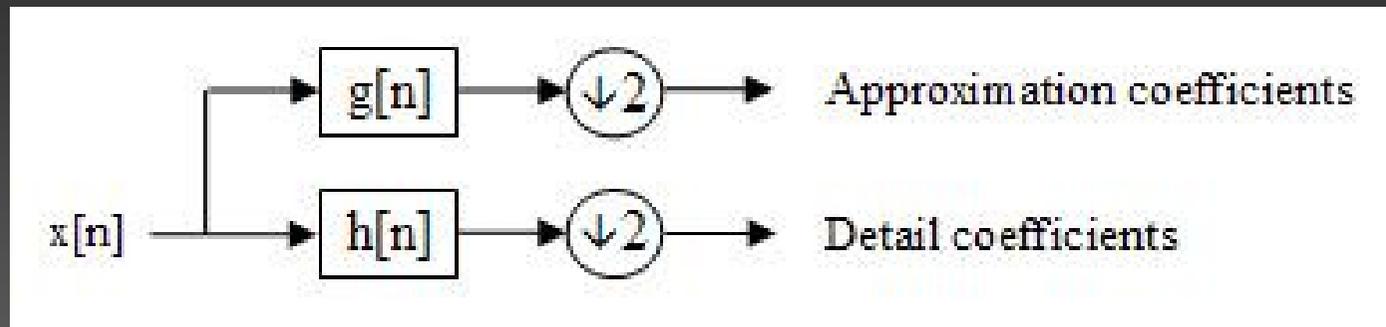
- A *wavelet* pai é uma função de escala que depende da *wavelet* mãe. Das funções  $\phi(t)$  e  $\psi(t)$ , podemos calcular as seqüências  $h = \{h_n\}, n \in \mathbb{Z}$ , e  $g = \{g_n\}, n \in \mathbb{Z}$ , como:

$$h_n = \langle \psi_{0,0}, \psi_{1,n} \rangle e^{\psi(t)} = \sqrt{2} \sum_{n \in \mathbb{Z}} h_n \psi(2t - n)$$

$$g_n = \langle \phi_{0,0}, \phi_{1,n} \rangle e^{\phi(t)} = \sqrt{2} \sum_{n \in \mathbb{Z}} g_n \phi(2t - n)$$

## Redução de dados e abstração de características

- Simplificando (banco de filtros):



- Passa alta:  $h = \{h_n\}, n \in \mathbf{Z}$
  - Passa baixa:  $g = \{g_n\}, n \in \mathbf{Z}$
- Filtros  $h$  e  $g$  são lineares (convolução):
  - $c(n) = \sum_k g(k) x(n-k) = g \otimes x = \{c_n\}, n \in \mathbf{Z}$  (aprox.)
  - $d(n) = \sum_k h(k) x(n-k) = h \otimes x = \{d_n\}, n \in \mathbf{Z}$  (difer.)

## Redução de dados e abstração de características

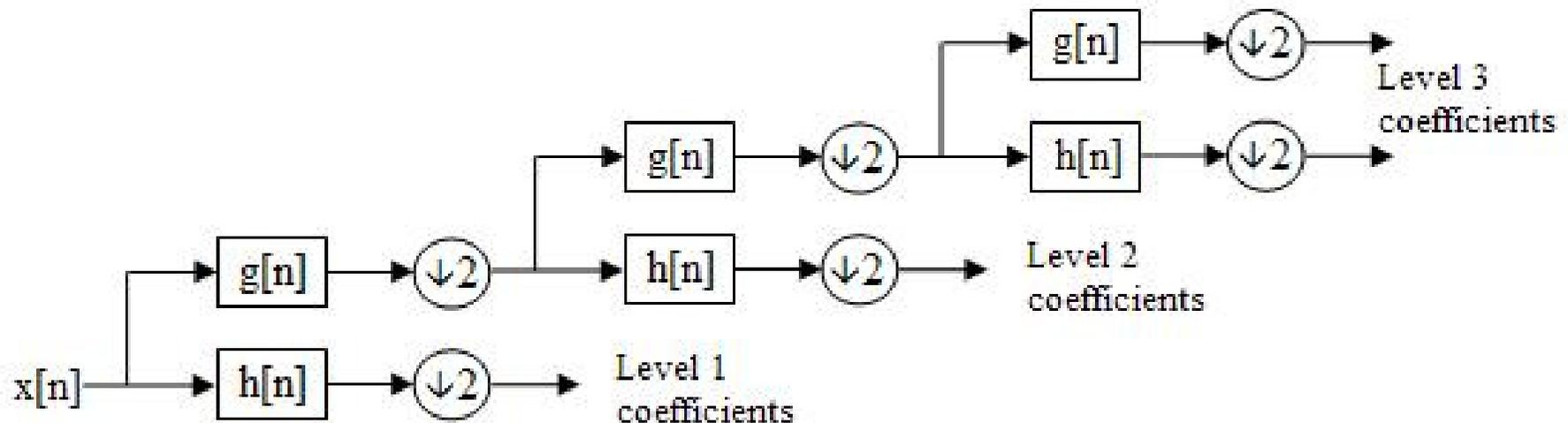
- Na prática:
  - seja o operador de subamostragem  $\downarrow 2$ , tal que

$$\downarrow 2 x(n) = x_0, x_2, x_4, \dots$$

- Filtros ortogonais e bi-ortogonais
  - reconstrução perfeita do sinal, apesar da perda devido à sub-amostragem.
  - decompõe o sinal em apenas duas faixas de frequência.

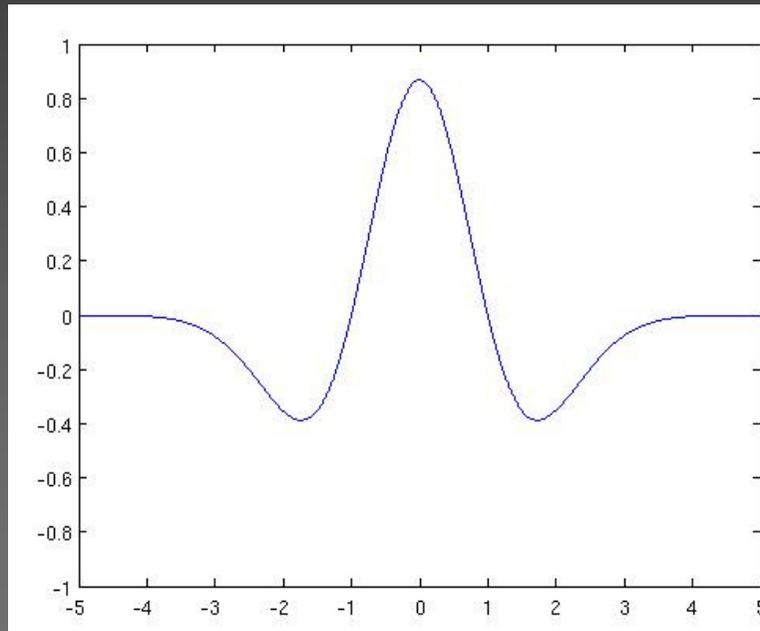
## Redução de dados e abstração de características

- Encadear uma série de bancos de filtros, usando operações de sub-amostragem (divisão da frequência de amostragem por 2):

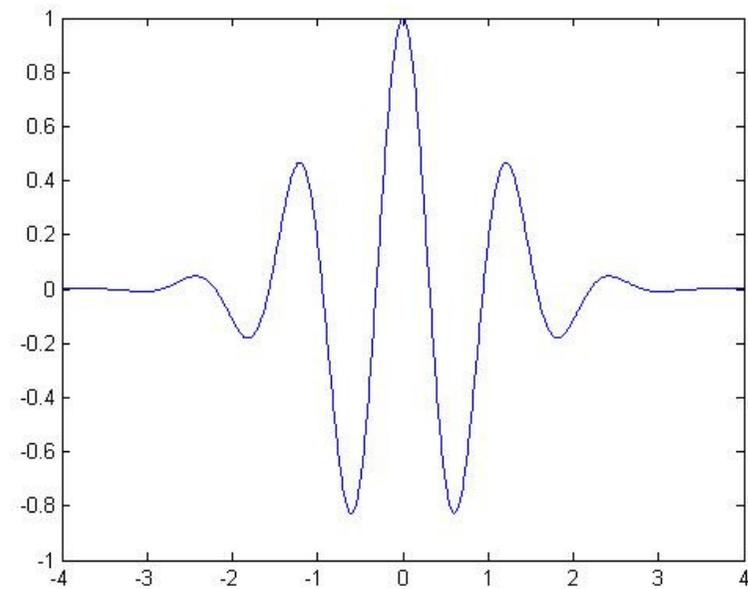


# Redução de dados e abstração de características

- Outras bases (restrições de unicidade e ortogonalidade, espaço de Hilbert):

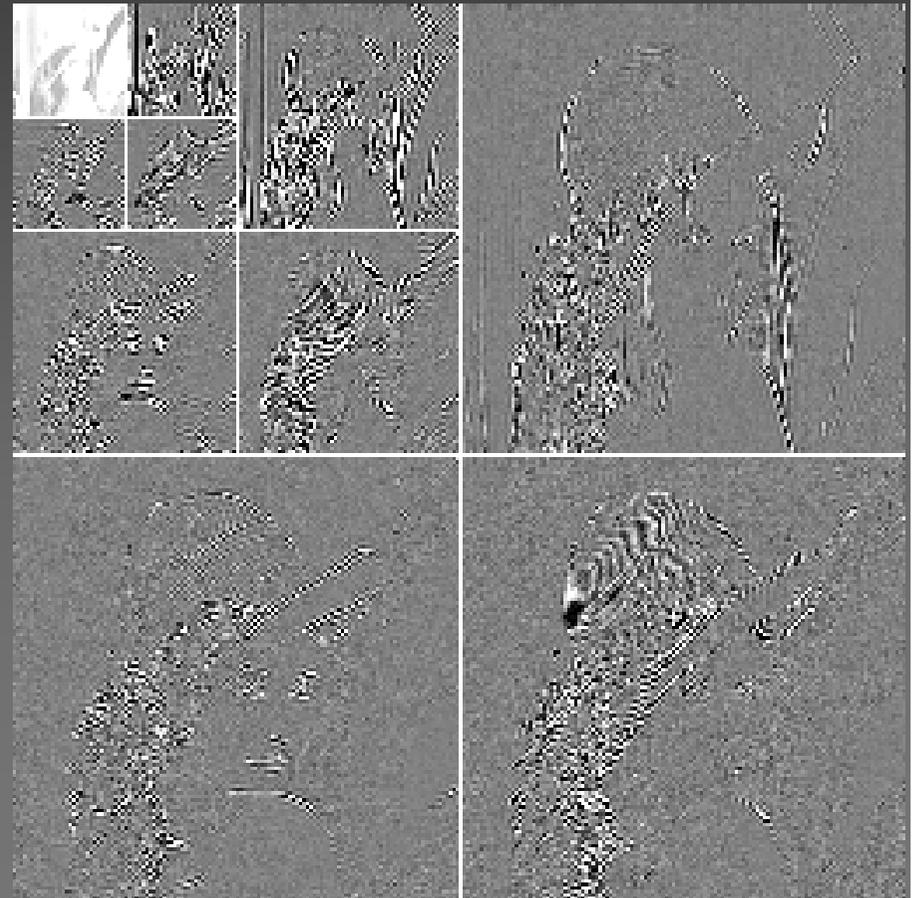
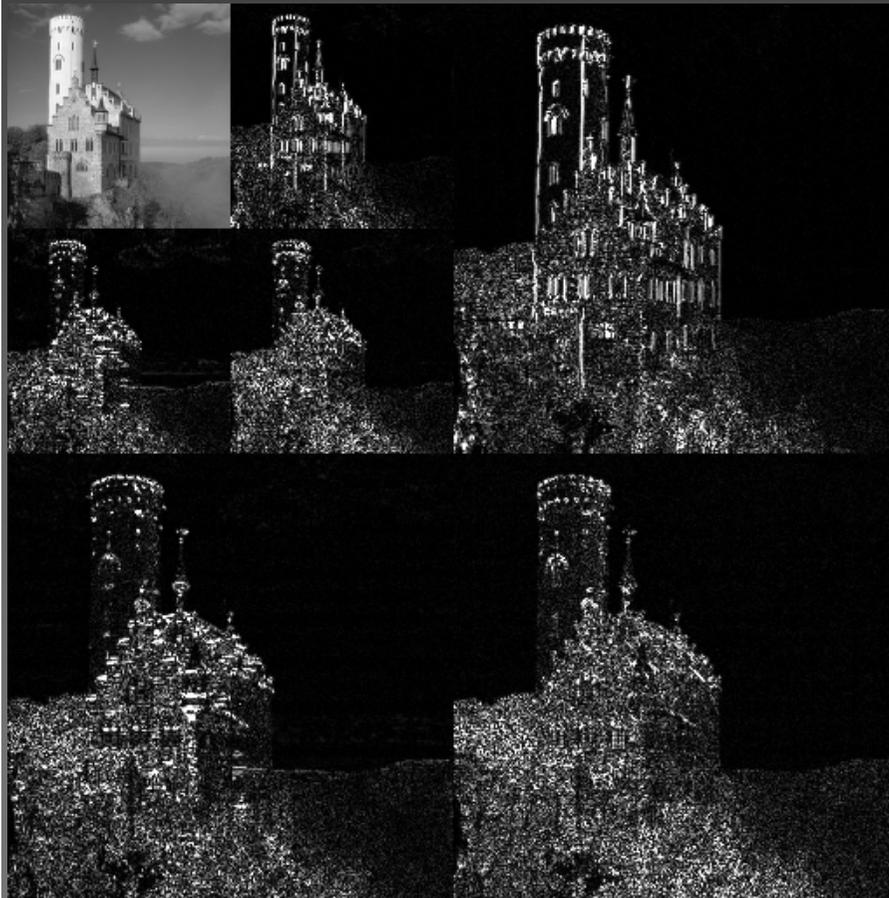


Chapéu Mexicano



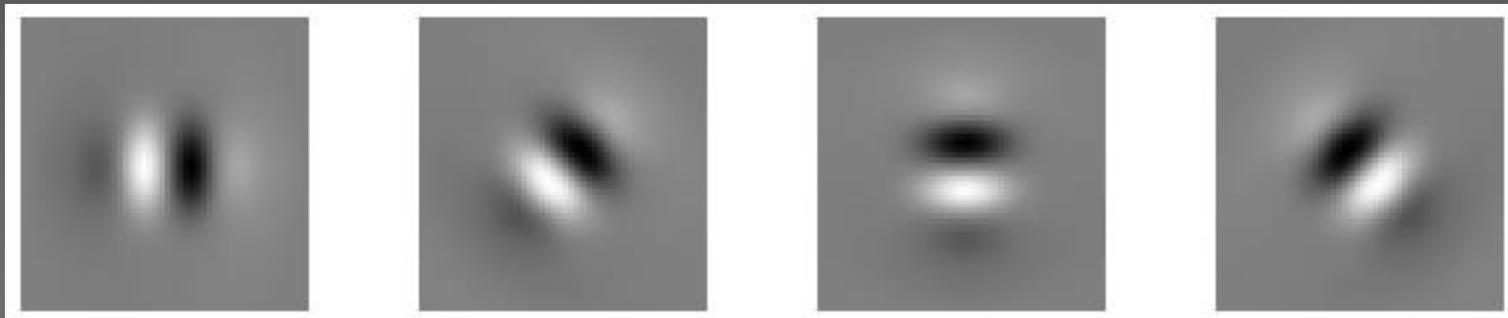
Morlet

# Imagem decomposta por Wavelet



## Redução de dados e abstração de características

- Pirâmide com filtros direcionais (Willian Freeman):
  - imagens sucessivamente menores usando filtros direcionais



0

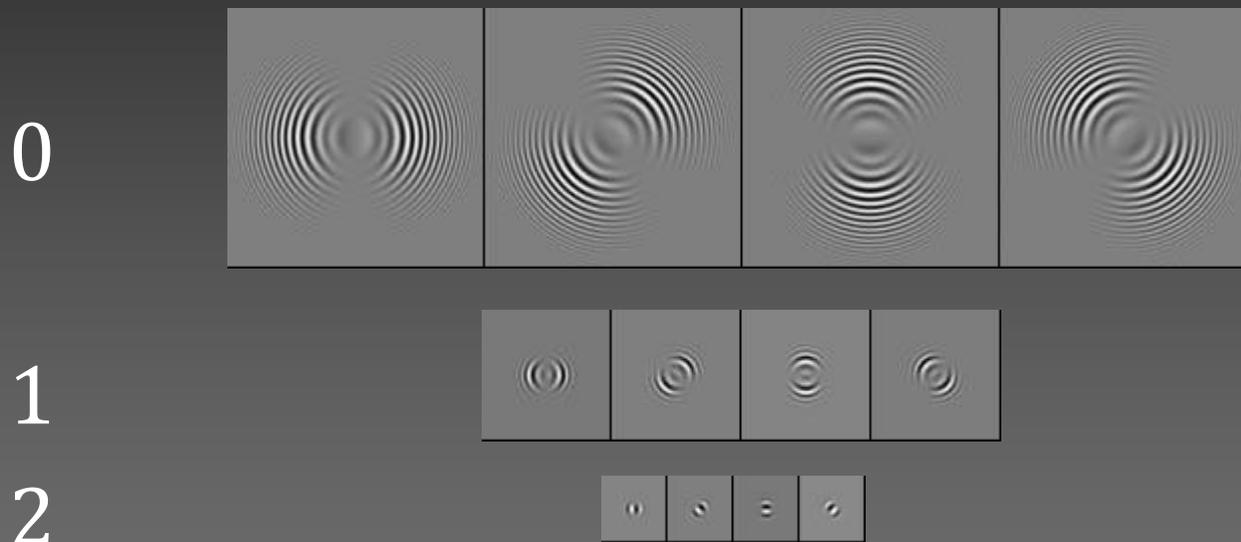
45

90

135

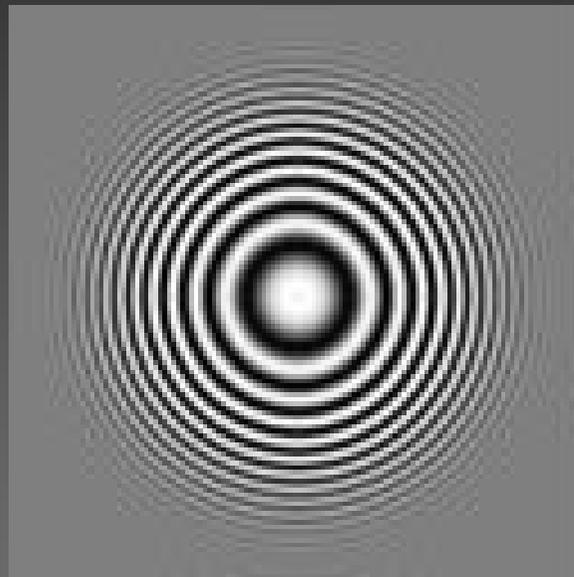
## Redução de dados e abstração de características

- Pirâmide direcional (steerable):

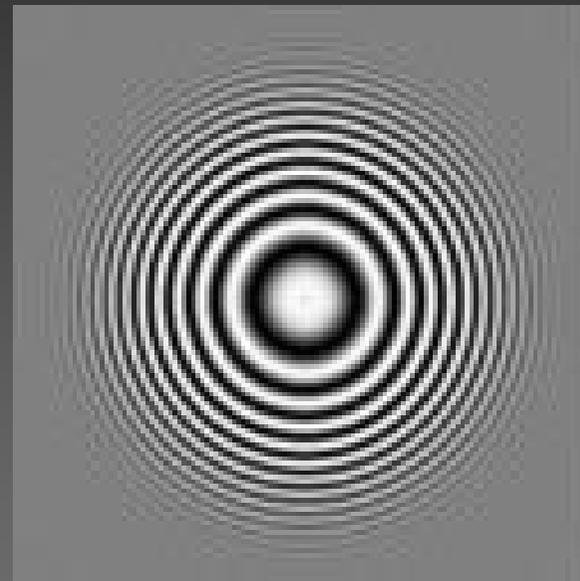


## Redução de dados e abstração de características

- Pirâmide direcional (steerable):



Original



Reconstruída

## Redução de dados e abstração de características

- Foveamento
- Ideia de usar estrutura em forma de fóvea para diminuir o processamento
- Pré-processamento de imagens
- Foveated rendering (nVidia, Google)

# Anatomia do olho humano

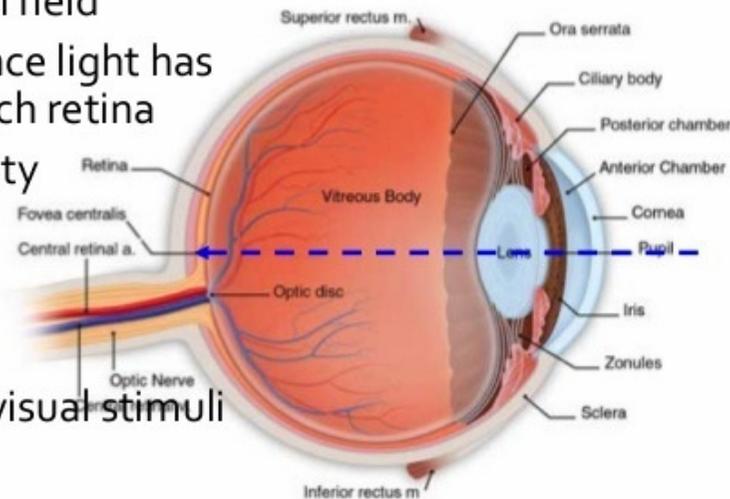
## Anatomy of the Eye

### ■ Fovea Centralis or Fovea

- Centre of visual field
- Shortest distance light has to travel to reach retina
- best visual acuity

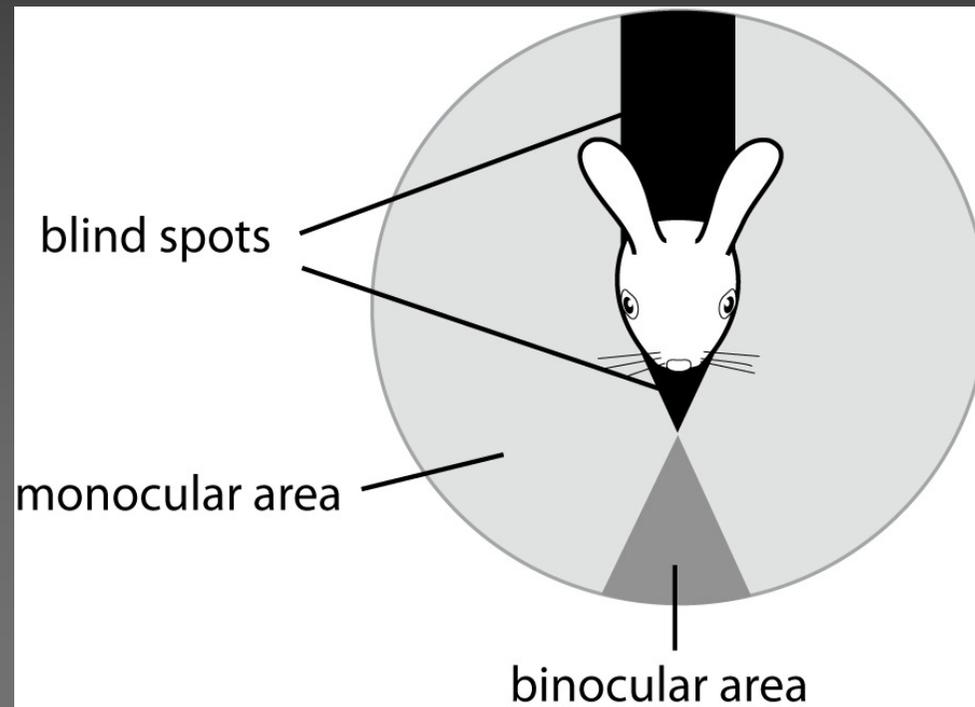
### ■ Optic Nerve

- Neurons send visual stimuli to the brain



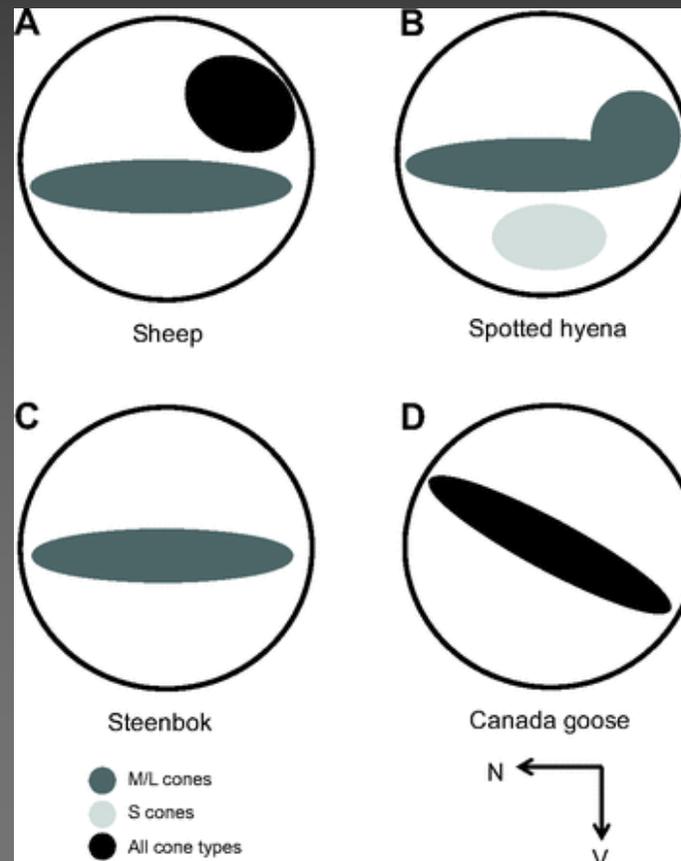
# Rabbits

Minimal not indented fovea



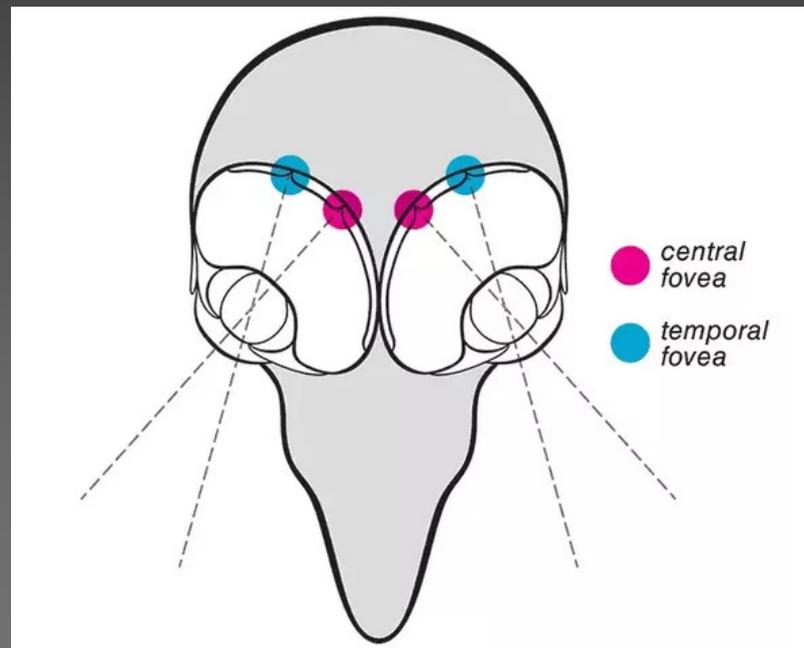
# Different animals foveae

Alguns tipos de fóvea (ovelha, hiena, antílope, ganso)



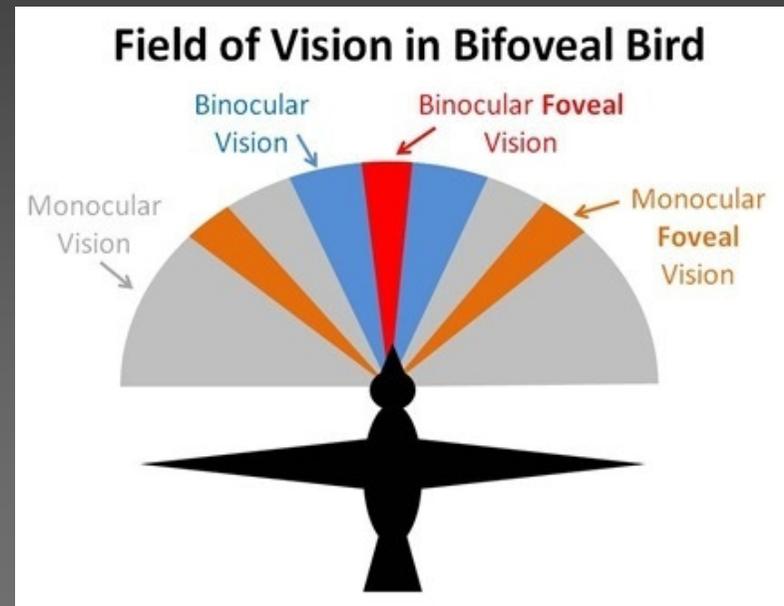
# Bird eyes

A metade dos pássaros tem duas fóveas



# Bird eyes

## Campo de visão bifoveal

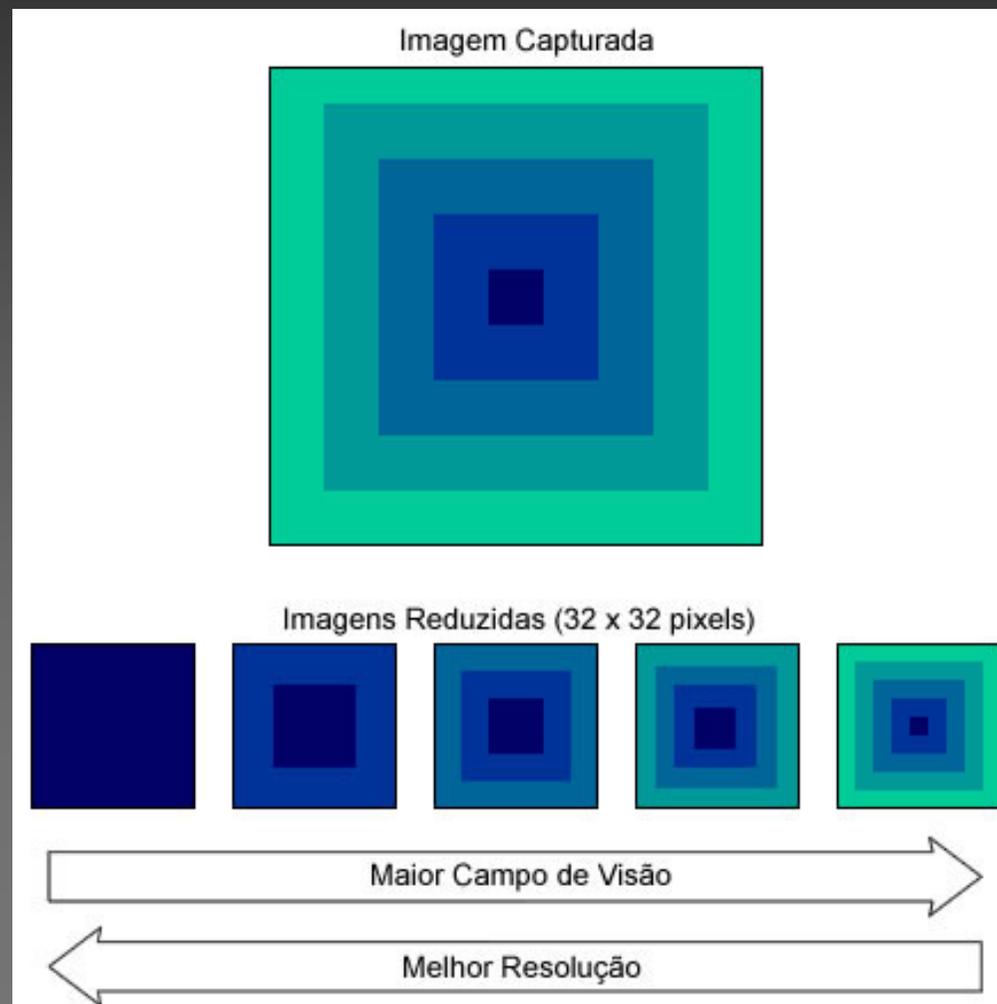


## Redução de dados e abstração de características

- Fóvea em multiresolução
  - imagens com maior resolução incluem mais detalhes dos objetos da cena, porém cobrindo uma região menor da cena.
  - imagens com menor resolução incluem menos detalhes dos objetos, mas cobrem regiões maiores da cena.

## Redução de dados e abstração de características

- Fóvea em multiresolução (clássico)



# Fóvea em Multirresolução

- Algoritmo fóvea MR
  - Entrada: Imagem Im, Nível de resolução N
  - Saída: SubImagem SubIm;
  - Calcular h;
  - Calcular subh;
  - para i = 0 até i < 32 faça
    - para j = 0 até j < 32 faça
      - $$\text{SubIm}(i, j) = ( \text{Im}(i*h, j*h) +$$
      - $$+ \text{Im}(i*h + \text{subh}, j*h) +$$
      - $$+ \text{Im}(i*h - \text{subh}, j*h) +$$
      - $$+ \text{Im}(i*h, j*h + \text{subh}) +$$
      - $$+ \text{Im}(i*h, j*h - \text{subh})) / 5;$$
    - fim para
  - fim para

## Redução de dados e abstração de características

- Imagem original e em forma de fóvea MR

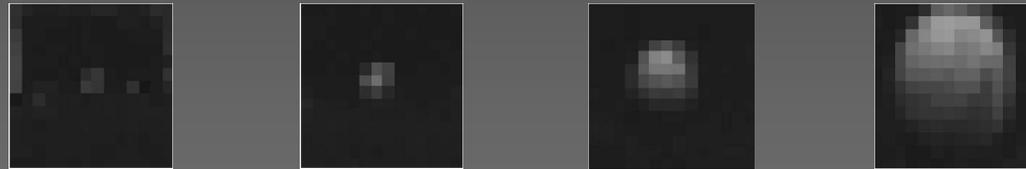


## Redução de dados e abstração de características

- Primeiro nível cobre toda a imagem original (resolução mais grosseira)
- Último nível cobre uma parte pequena da imagem (resolução mais fina), a fóvea
- Níveis intermediários cobrem regiões de diferentes tamanhos na imagem
- $(256 \times 256) \Rightarrow 5 \times (32 \times 32)$ :
  - redução de 92%

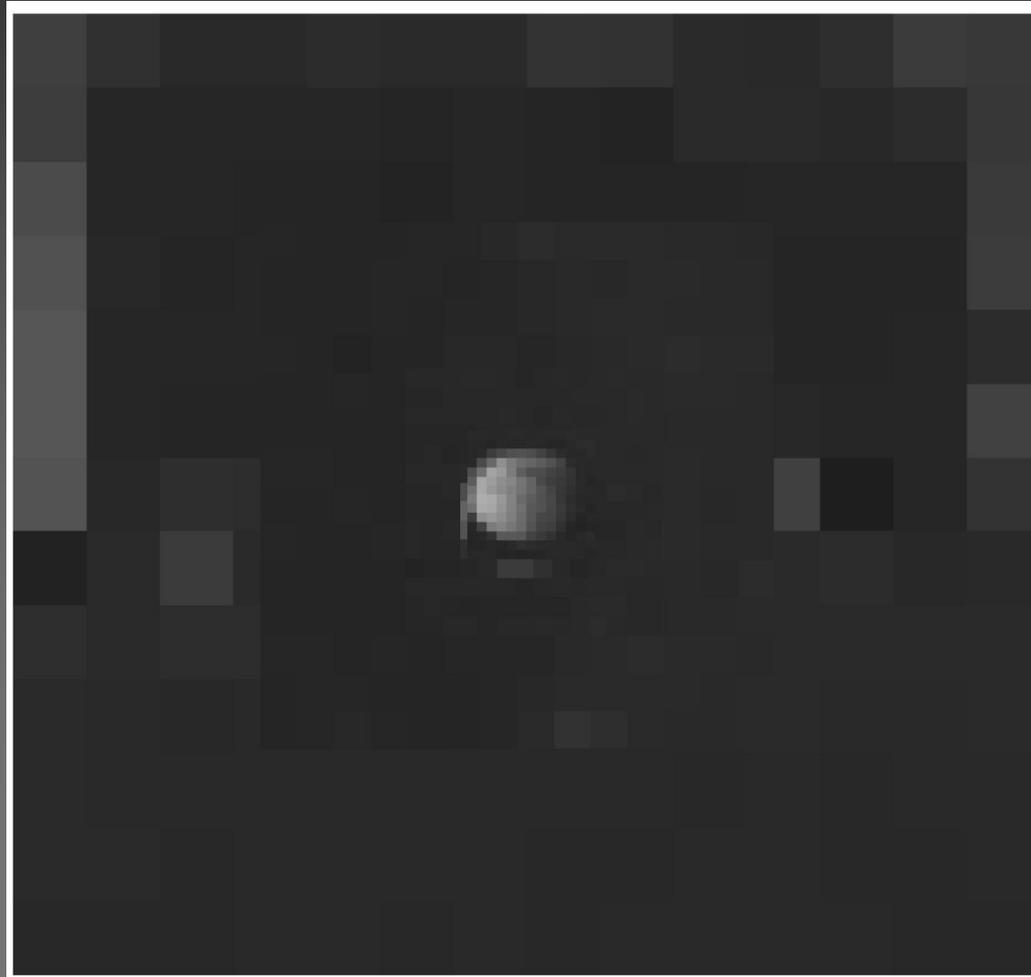
## Redução de dados e abstração de características

- Imagem em fóvea MR



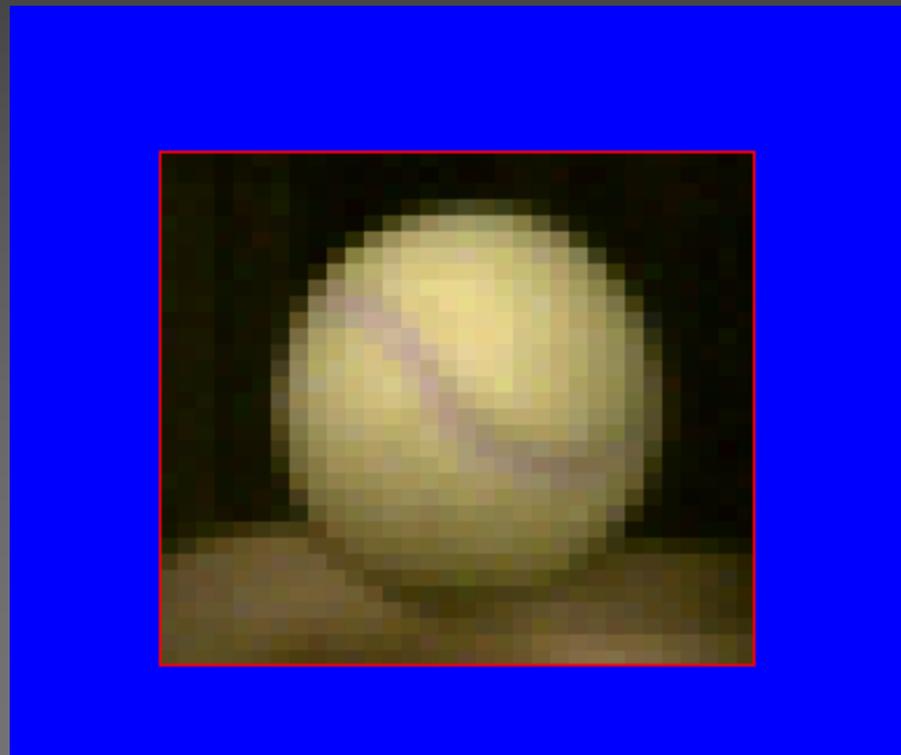
## Redução de dados e abstração de características

- Imagem reconstruída da fóvea MR



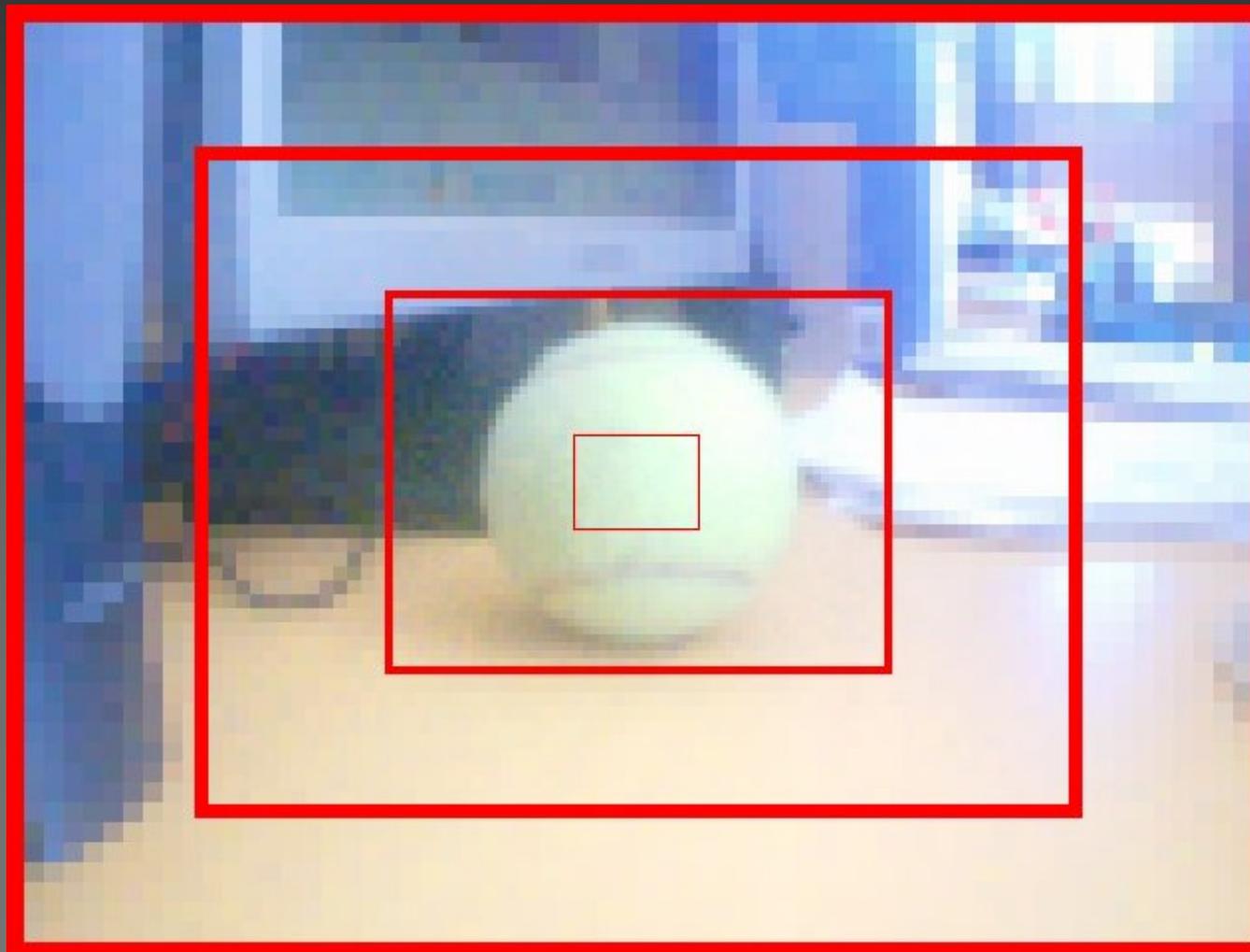
## Redução de dados e abstração de características

- Imagem em fóvea MR e zoom do nível 3



## Redução de dados e abstração de características

- Imagem reconstruída (reescaladas)



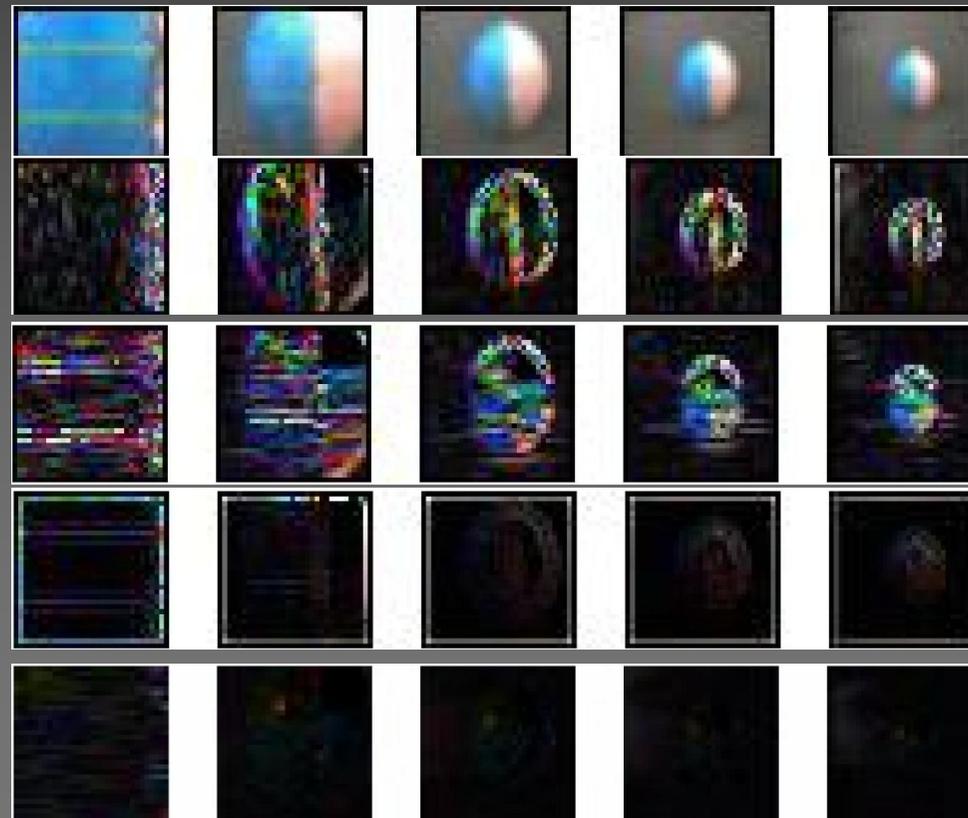
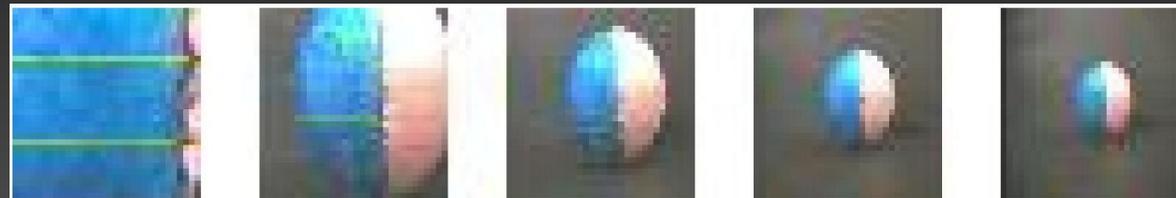
# Melhorando a estrutura

- Tempos de processamento

Etapa	Multiresolução ( $\mu s$ )	Original ( $\mu s$ )
Multiresolução	244	–
Gaussiano	584	10480
Gradiente	1169	21020
Laplaciano	579	10506
Movimento	62	5355
<b>Total</b>	<b>2638</b>	<b>47361</b>

# Melhorando a estrutura

- Aplicação da fóvea MR



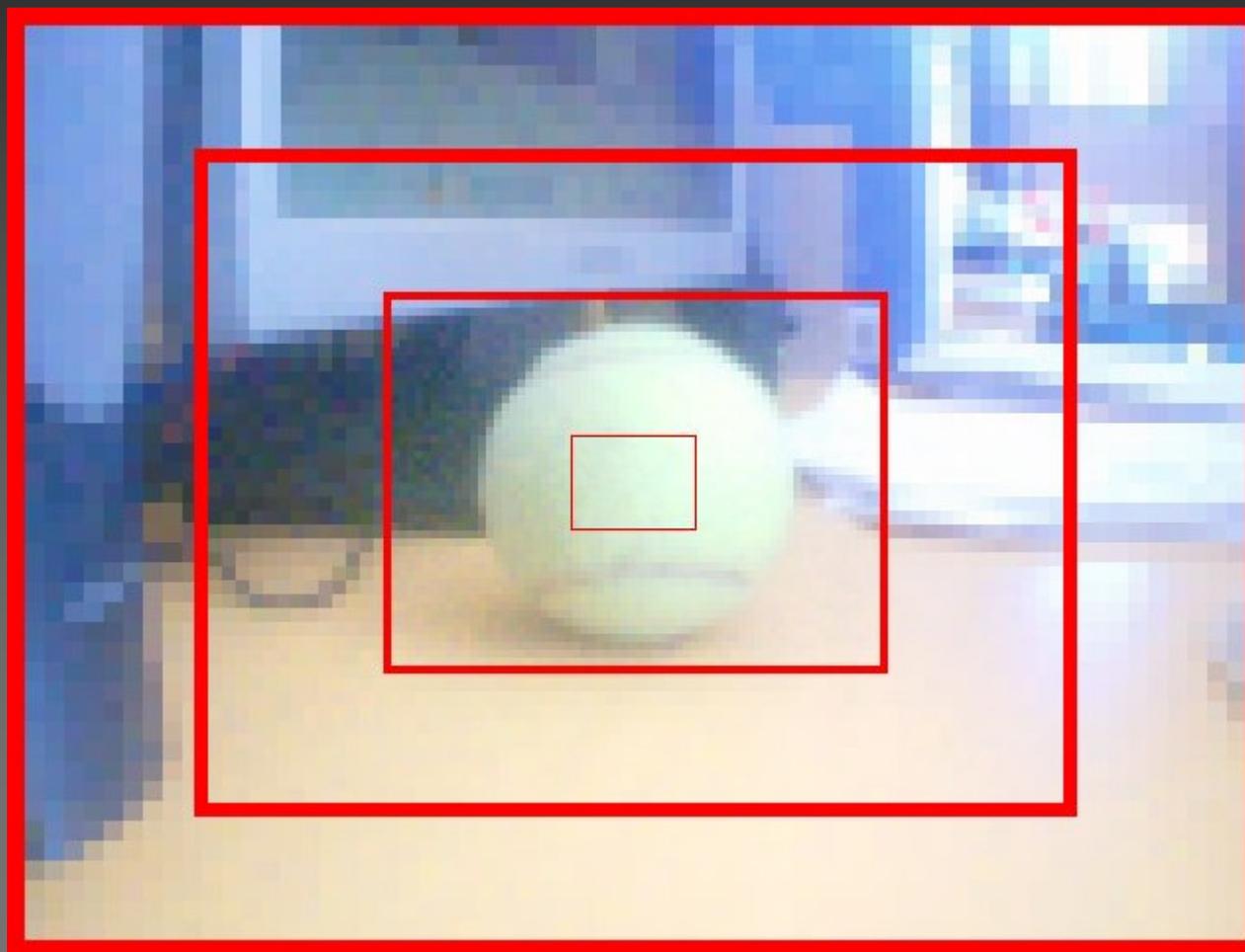
# Introduzindo a fóvea móvel

- Voltar a atenção (obter detalhes) para algum local fora do centro da imagem:
  - ▣ pode-se adquirir outra imagem, colocando esta nova região no centro da imagem (movendo as câmeras) e recalcular uma nova estrutura em fóvea MR usando o processo acima.
  - ▣ recalcular somente a estrutura em multiresolução passando como parâmetro uma outra posição, para a fóvea não muito na periferia da imagem.

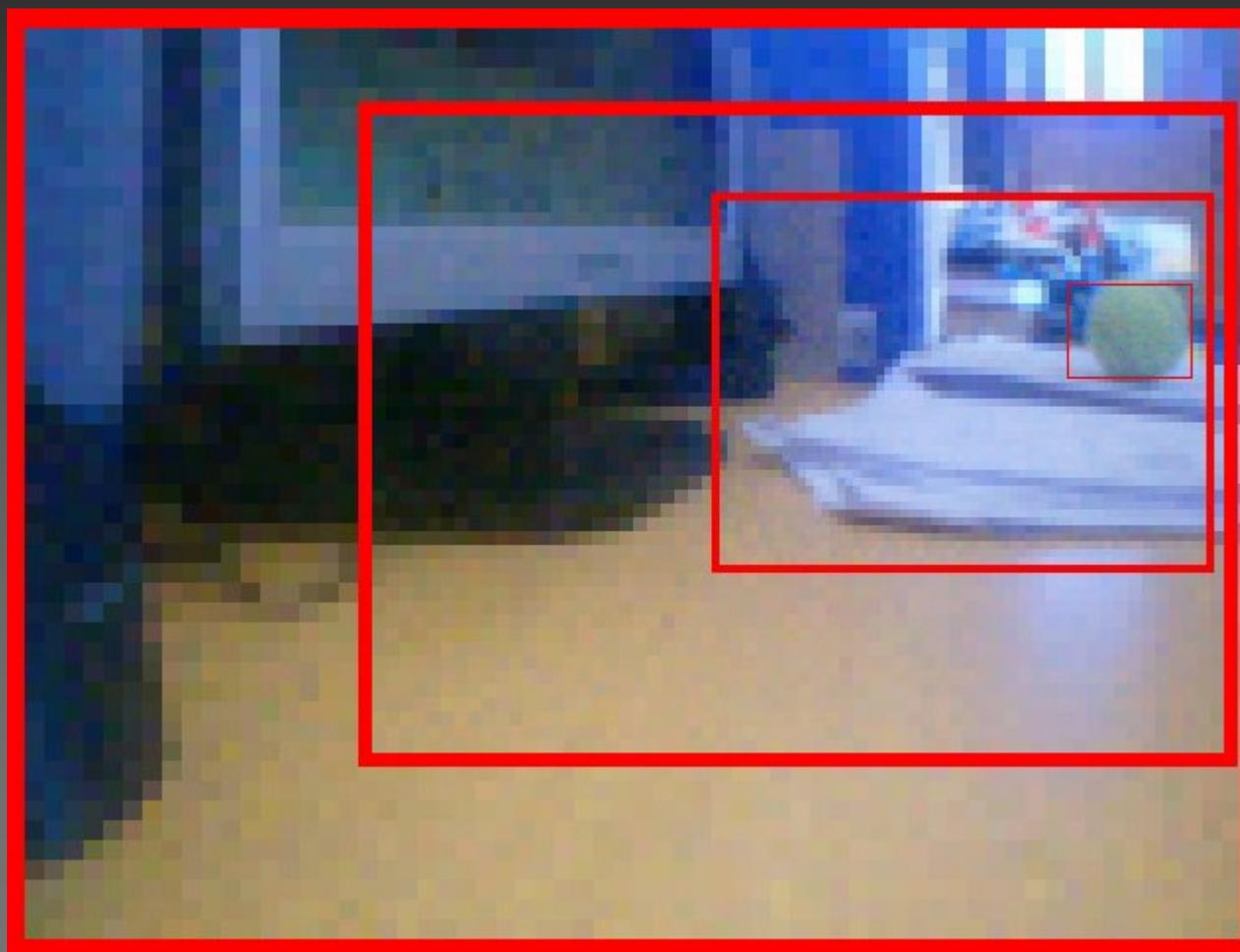
# Introduzindo a fóvea móvel

- Se a fóvea atinge a periferia da imagem:
  - sistema de decisão solicita movimento dos recursos robóticos, colocando a nova fóvea no centro das imagens.
  - recalcula a estrutura em multiresolução passando como parâmetro o centro da imagem.
  - isso pode demorar (envolve movimentos do hardware)

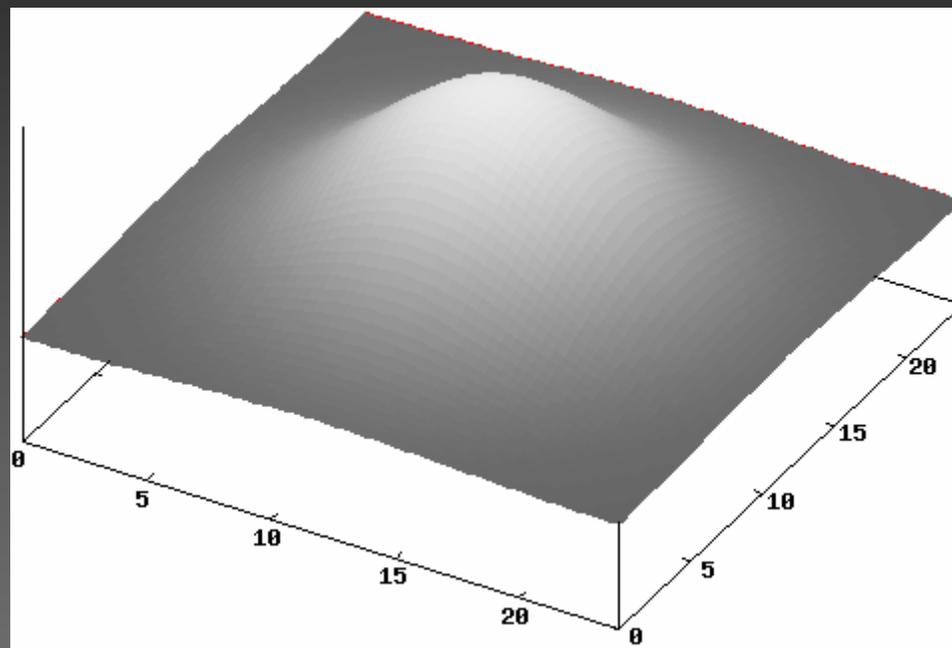
# Introduzindo a fóvea móvel



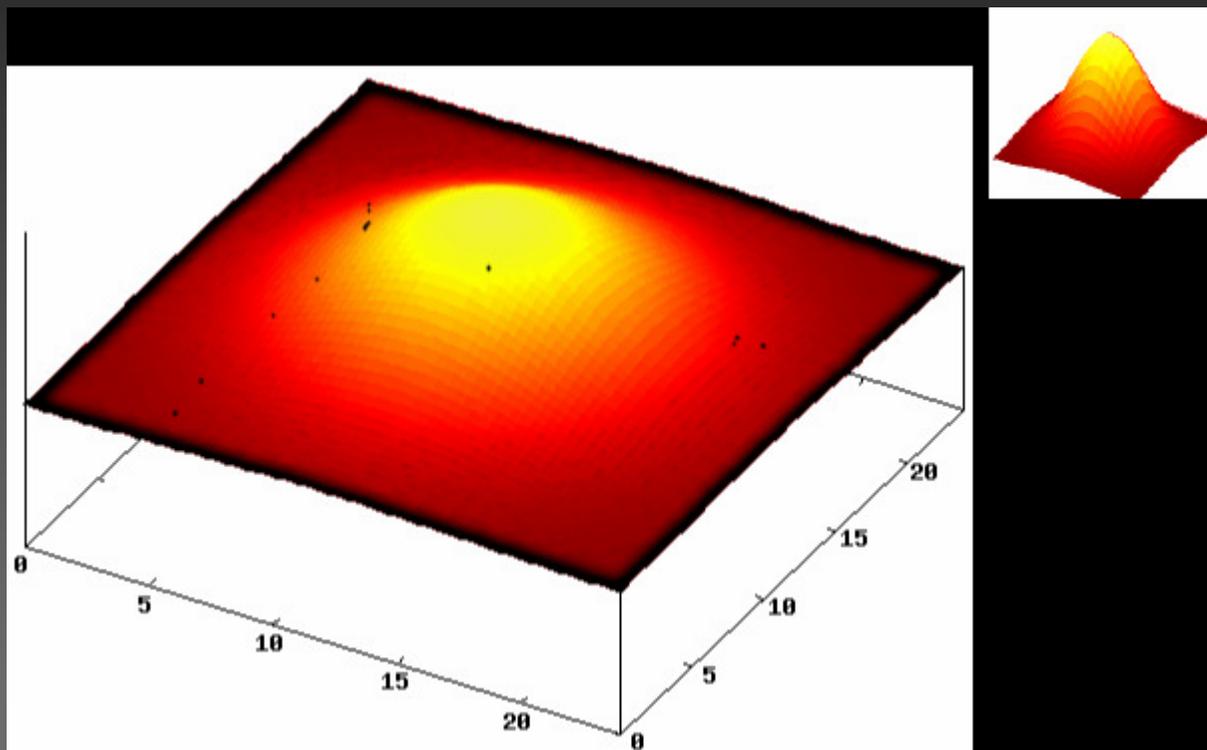
# Introduzindo a fóvea móvel



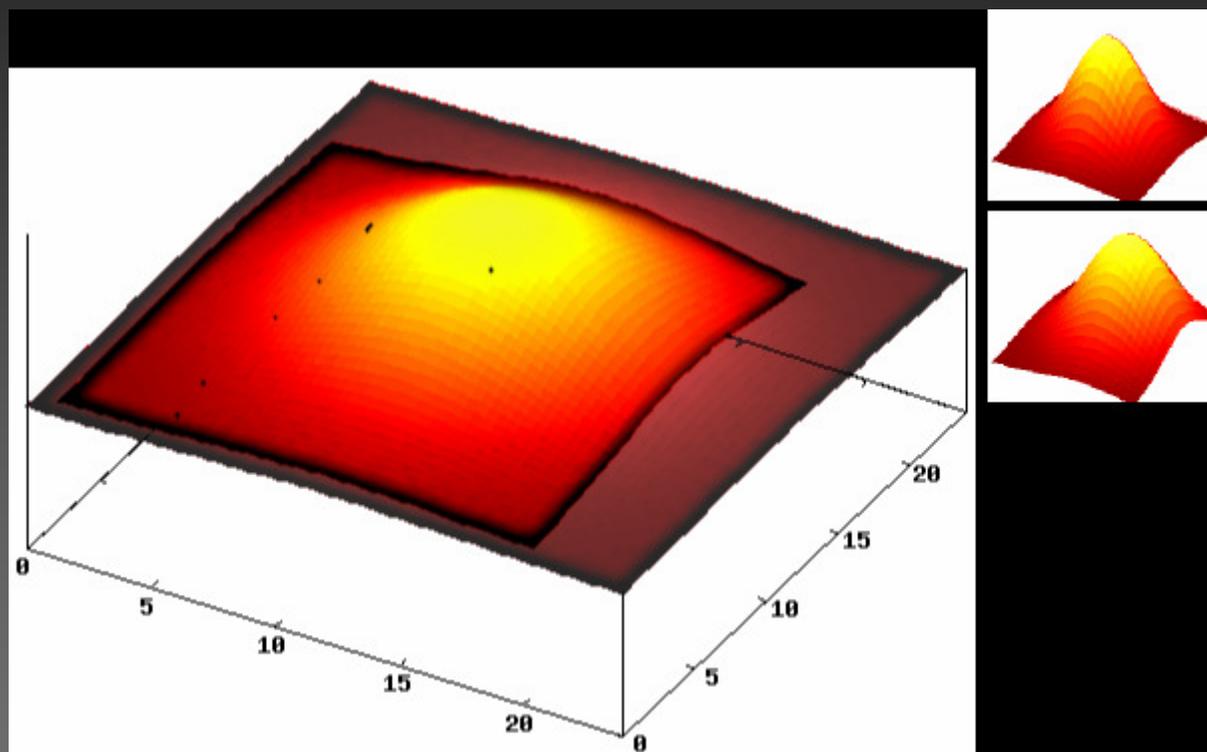
# Introduzindo a fóvea MR móvel



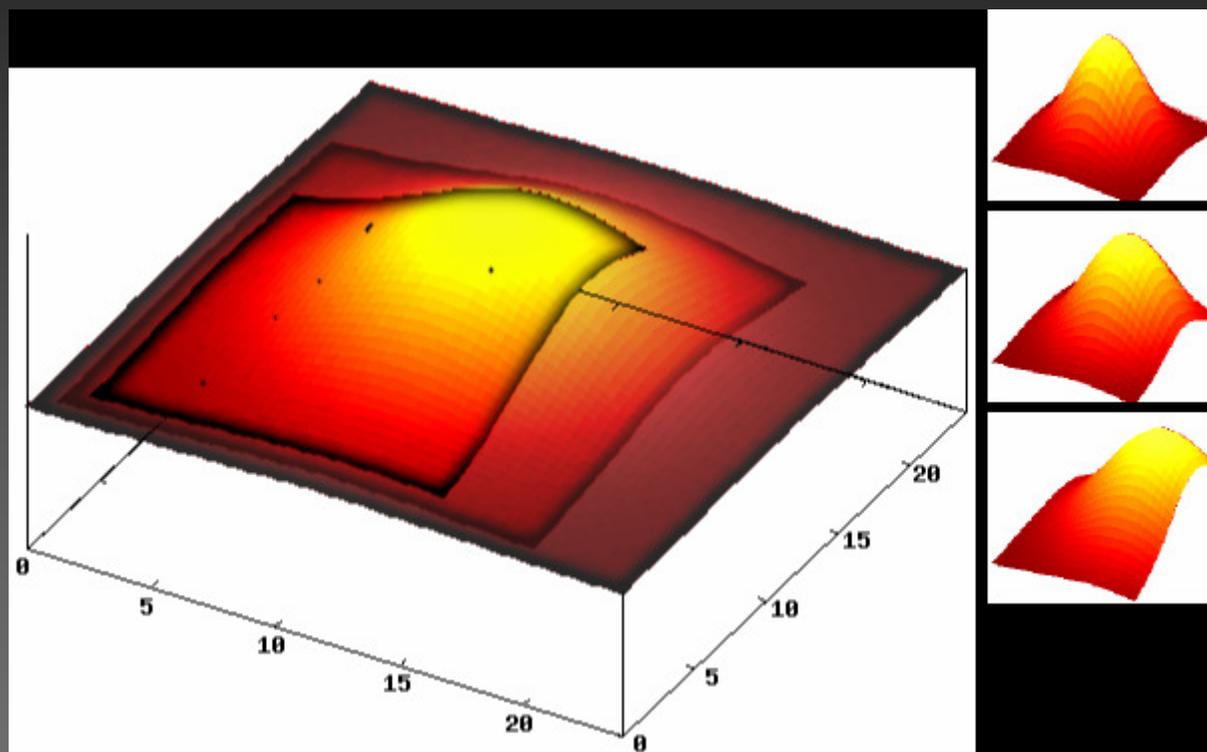
# Introduzindo a fóvea móvel



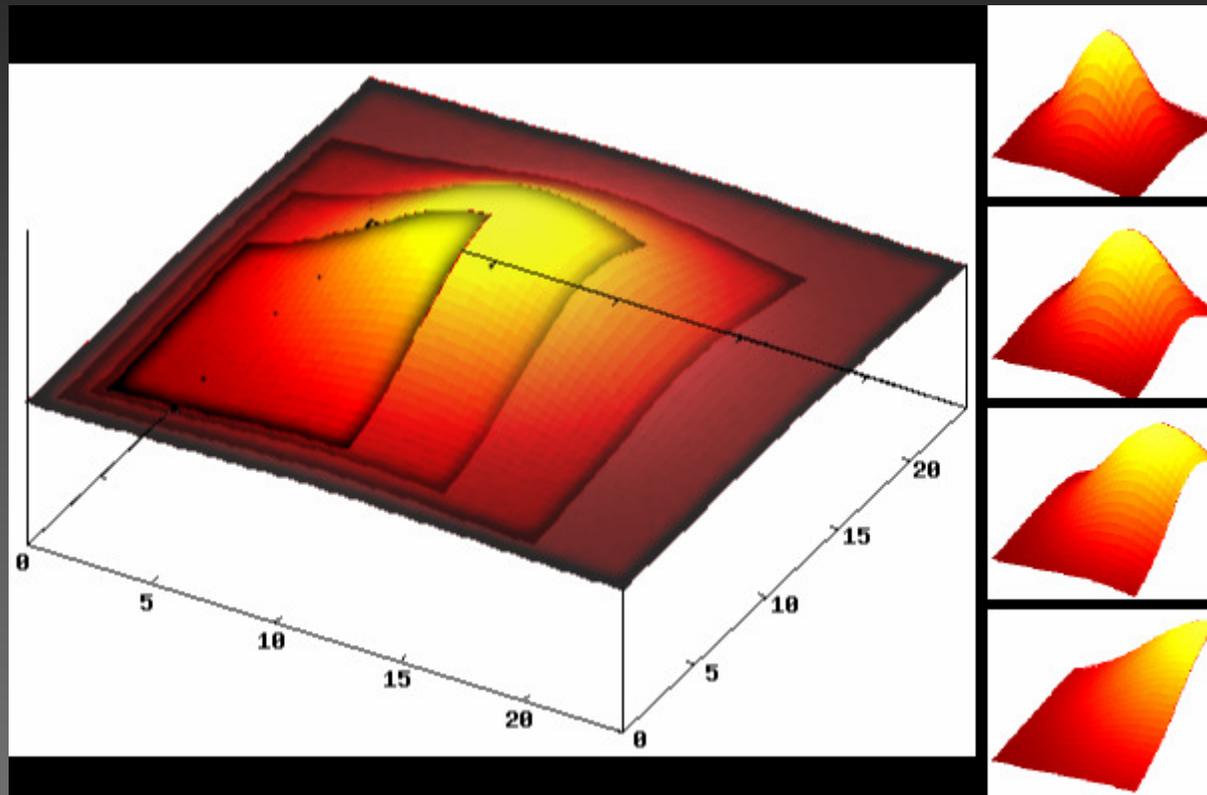
# Introduzindo a fóvea móvel



# Introduzindo a fóvea móvel



# Introduzindo a fóvea móvel



# Introduzindo a fóvea móvel

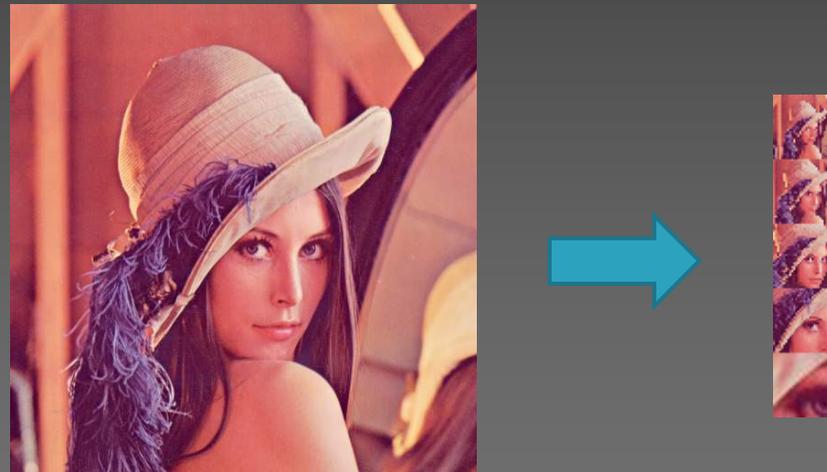
- Mapeamento de um nível na estrutura fóvea MR para a imagem original  $I$



$$\phi_k(p) = \frac{k(W-1)(U-W) + 2k(W-1)F + 2p(mU - kU + kW - m)}{2m(W-1)}$$

# Introduzindo a fóvea móvel

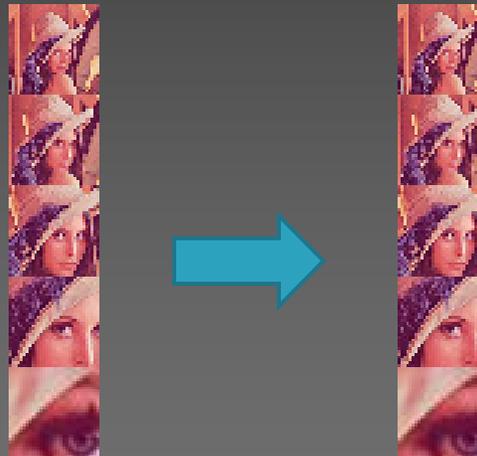
- Mapeamento da imagem original  $I$  para um nível  $R_k$  na estrutura fóvea MR



$$\Phi_k^{-1}(p) = \frac{2m(W-1)p - k(W-1)(U-W) - 2k(W-1)F}{2(mU - kU + kW - m)}$$

# Introduzindo a fóvea móvel

- Mapeamento entre níveis da fóvea MR



$$\omega_{k,j}(p) = \frac{(W-1)(U-W)(k-j) + 2(W-1)(k-j)F + 2p(mU - kU + kW - m)}{2(mU - jU + jW - m)}$$

# Introduzindo a fóvea móvel

- Posicionando a fóvea:
  - ▣ Considerar vários aspectos da atenção
- Dicotomias do processo de atenção visual:
  - ▣ Atenção guiada pela tarefa (*top-down*)
  - ▣ Atenção guiada pelo estímulo (*bottom-up*)
  - ▣ Atenção encoberta (*covert*)
  - ▣ Atenção descoberta (*overt*)

# Introduzindo a fóvea móvel

- Determinando o foco de atenção:
  - modelos biológicos
  - modelos estatísticos e/ou computacionais
  - mapa de saliências
  - busca por objeto realizando um acompanhamento (ou *tracking*)
- Não é o foco desta aula

# Introduzindo a fóvea móvel

- Caso de duas ou mais câmeras (estéreo):
  - fixar a fóvea em uma das imagens (dominante) e determinar a posição na outra imagem (mesmo ponto da cena)
- Correspondência:
  - do nível mais grosseiro para o de maior resolução
- Com a fóvea determinada em cada nível pode-se extrair as características

# Abstração de dados (feature)

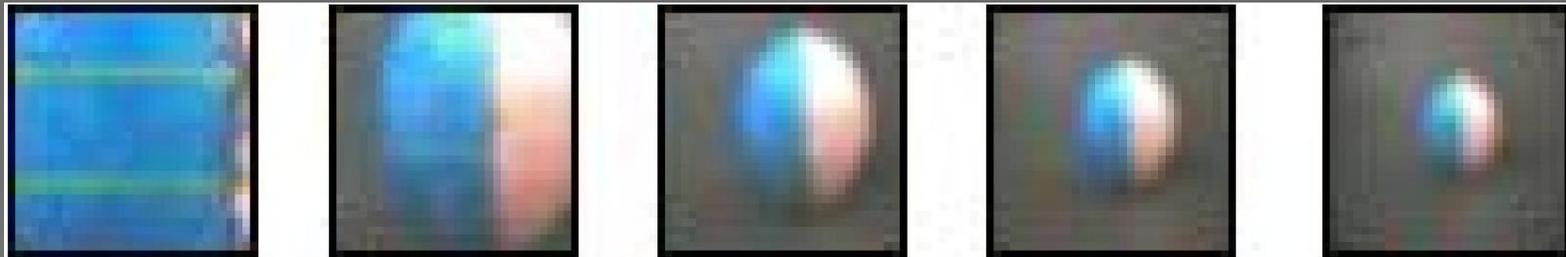
- Pixel (imagem):
  - ▣ codificação dos atributos de cor ou brilho de uma pequena porção da superfície de um objeto discretamente amostrada.
- Características:
  - ▣ textura, forma e posição (ou outras, com significado e detectável)
- Como extrair:
  - ▣ Análise das imagens, através de manipulação computacional dos valores de níveis de cinza

# Abstração de dados (feature)

- Filtro gaussiano

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

$$\frac{1}{16} \begin{array}{|c|c|c|} \hline 1 & 2 & 1 \\ \hline 2 & 4 & 2 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$



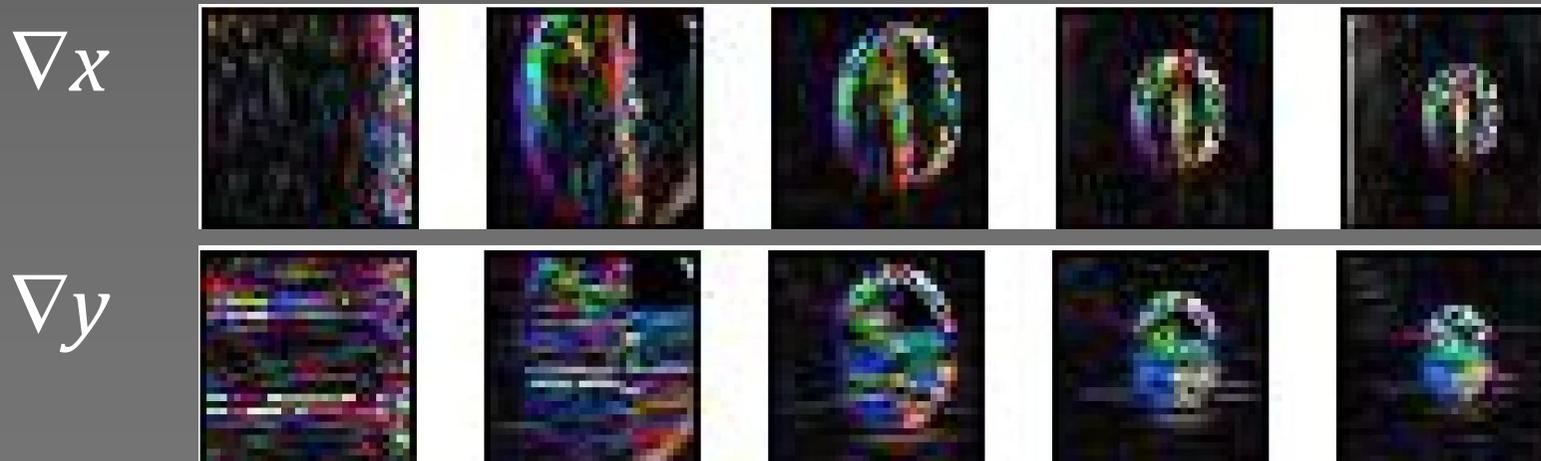
# Abstração de dados (feature)

- Filtro gradiente

$$\nabla f = \text{mag}(\nabla \mathbf{f}) = \left[ \left( \frac{\partial f}{\partial x} \right)^2 + \left( \frac{\partial f}{\partial y} \right)^2 \right]^{1/2}$$

$$\nabla_y \begin{array}{|c|c|c|} \hline -1 & -2 & -1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array} \nabla_x$$

$$\begin{array}{|c|c|c|} \hline -1 & 0 & -1 \\ \hline -2 & 0 & -2 \\ \hline -1 & 0 & -1 \\ \hline \end{array}$$



# Abstração de dados (feature)

- Filtro laplaciano

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

$$\nabla^2(x,y)$$

0	-1	0
-1	4	-1
0	-1	0

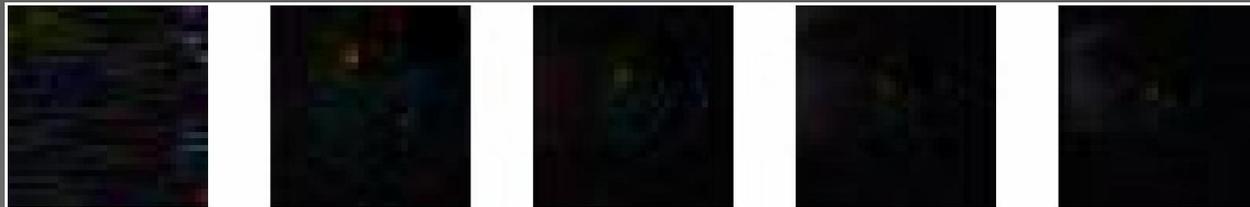


$$\nabla^2(x,y)$$

# Abstração de dados (feature)

- Movimento:  $\Delta I = I(t) - I(t-1)$

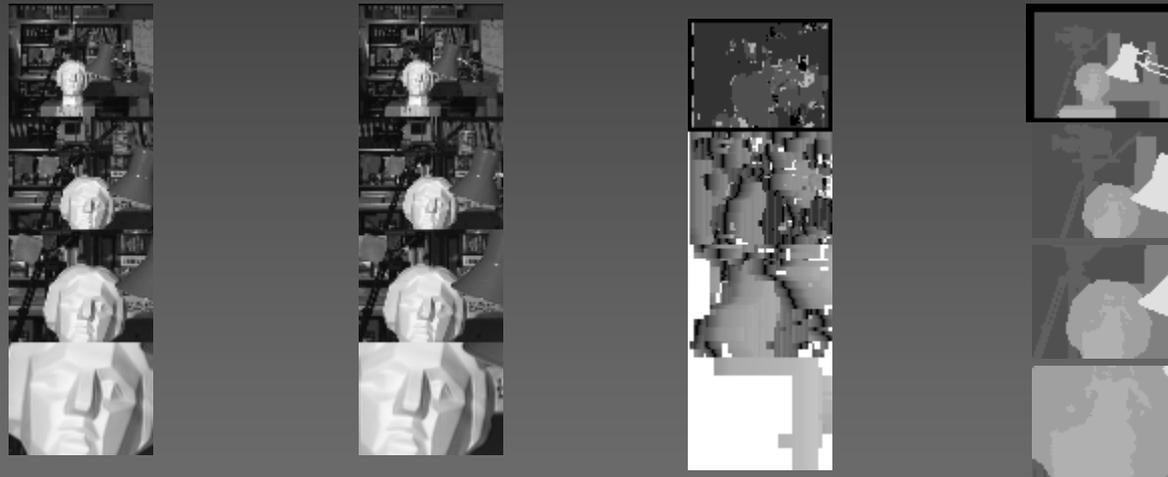
$$M_d = g_d^{(1)} * [I_t - I_{t-1}]$$



$$\nabla(I(t) - I(t-1))$$

# Abstração de dados (feature)

- Disparidade estéreo foveada



- Esquerda Direita Calculada Ideal

# Abstração de dados (feature)

- Tempo (294 x 294)

Phase	Multiresolução ( $\mu s$ )	Original ( $\mu s$ )
Fóvea	244	–
Gaussiano	584	10480
Gradiente	1169	21020
Laplaciano	579	10506
Movimento	62	5355
Estéreo (3x3)	2916	1653555
Estéreo (5x5)	5081	3053555
<b>Total (sem estéreo)</b>	<b>2638</b>	<b>47361</b>

# Abstração de dados (feature)

- Tempo (câmeras web, 352 x 288)

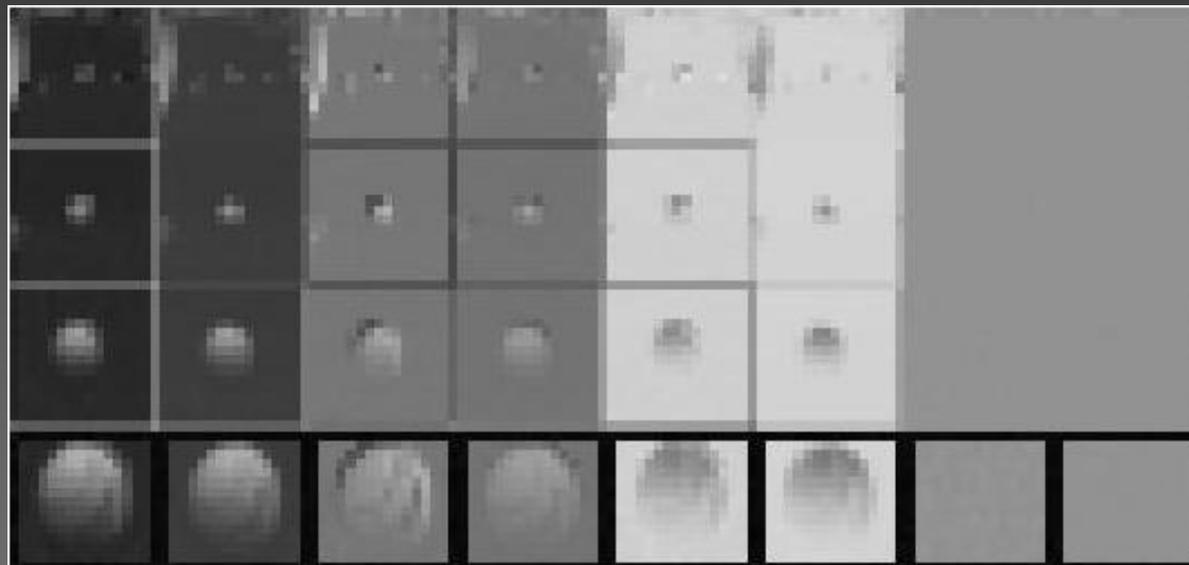
Fase	Multiresolução (ms)
Aquisição	21.8
Cópia p/ memória	1.5
Fóvea	1.2
Filtragem	3.6
Matching estéreo	2.9
Total (sem aquis.)	9.2

# Aplicações para a estrutura foveada

- Captura das imagens;
- Pré-processamento
  - distorção, equalização;
- Extração de características:
  - filtros (gaussiano, Sobel, laplaciano);
  - análise de movimento;
  - cálculo de disparidade estéreo.
- Tarefas de mais alto nível (reconhecimento, atenção, rastreamento, rendering, transmissão)

# Aplicações para a estrutura foveada

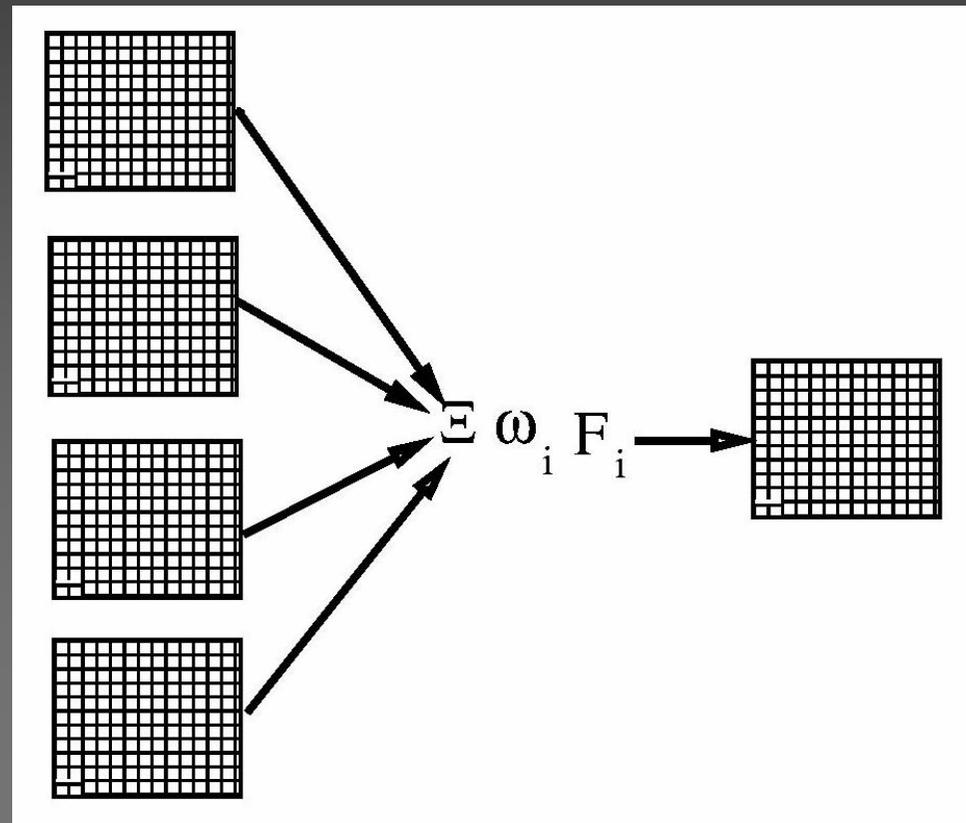
- Features (atenção e reconhecimento)



- Gaussianos, gradientes de gaussianos, laplacianos de gaussianos e gradientes da diferença entre quadros

# Aplicações para a estrutura foveada

- Atenção visual

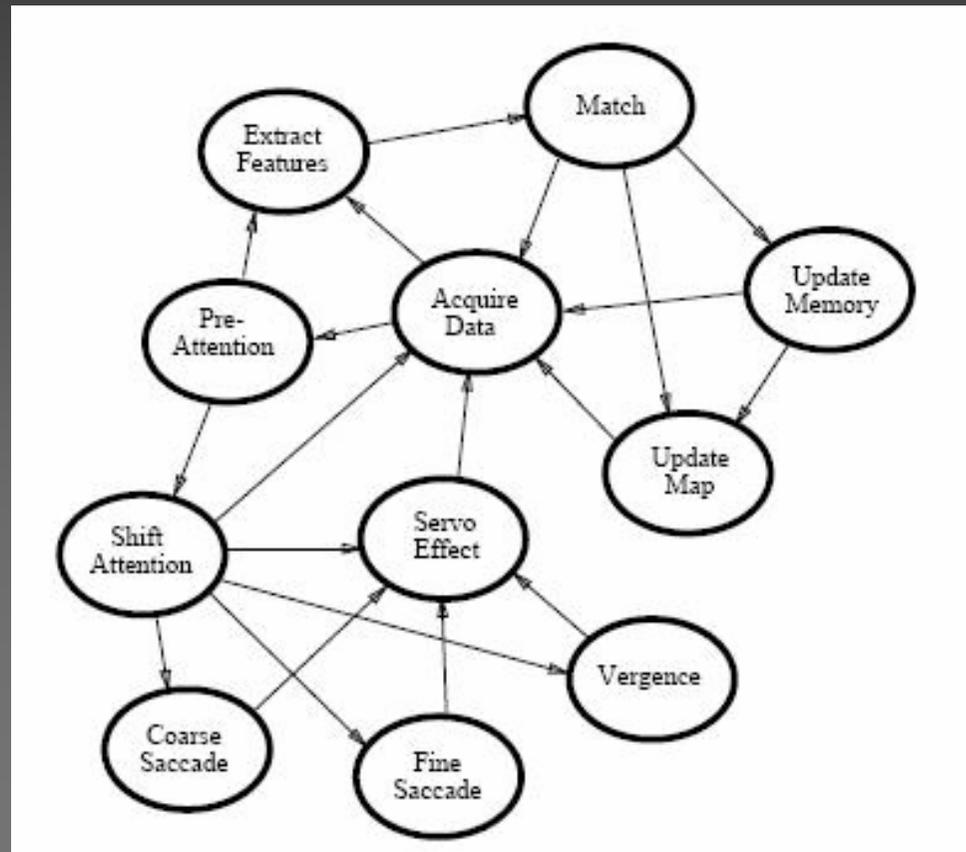


# Aplicações para a estrutura foveada

- Vigilância (atenção e categorização)
  - capturar imagens
  - criar a representação em forma de fóvea MR (hardware dedicado)
  - extrair características
  - gerar mapas de saliência
  - escolher a região “mais saliente”
  - mover as câmeras robóticas (sacádico)
  - identificar/reconhecer e inserir na memória
  - atualizar (ou não) mapa do ambiente

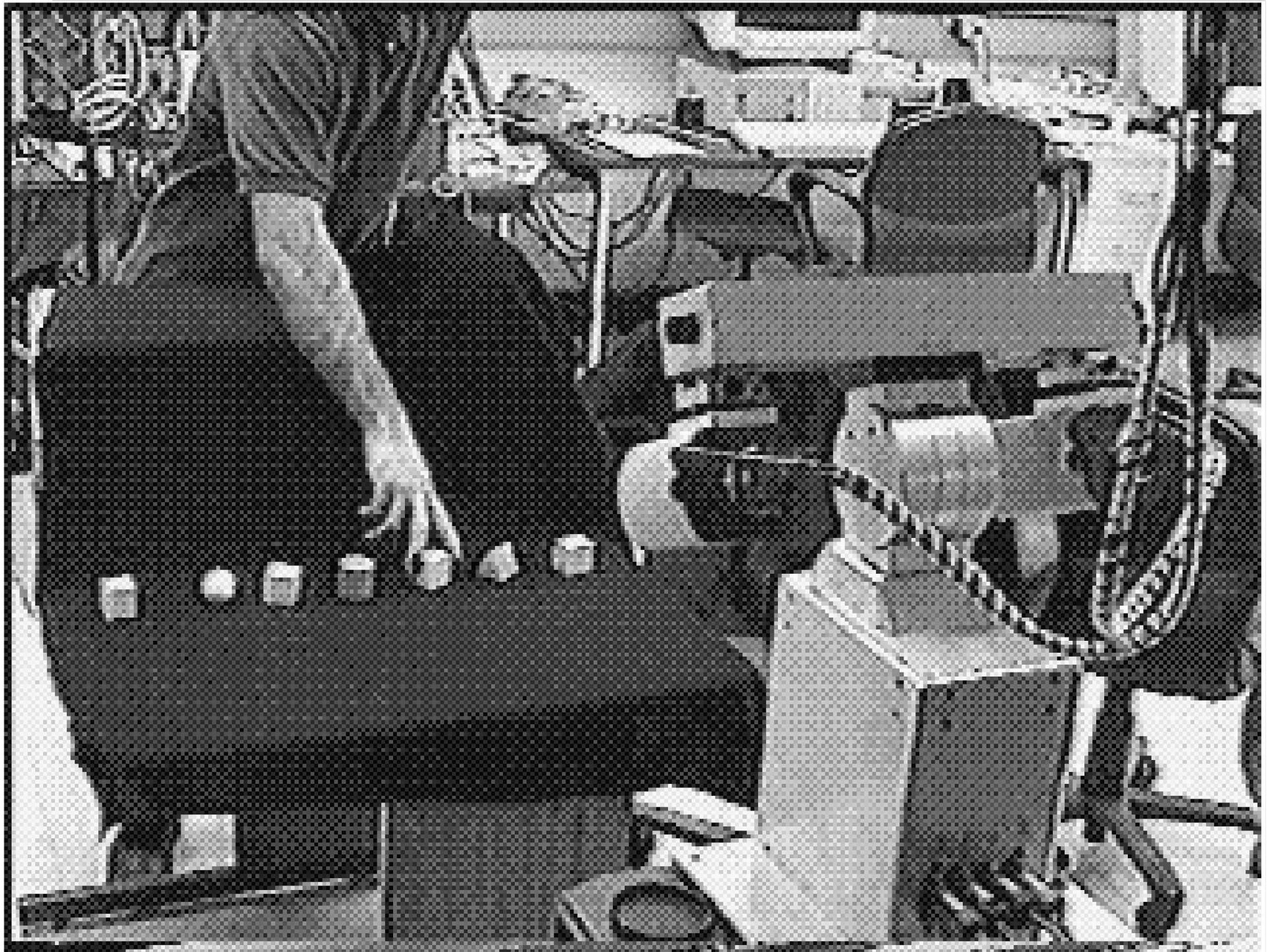
# Aplicações para a estrutura foveada

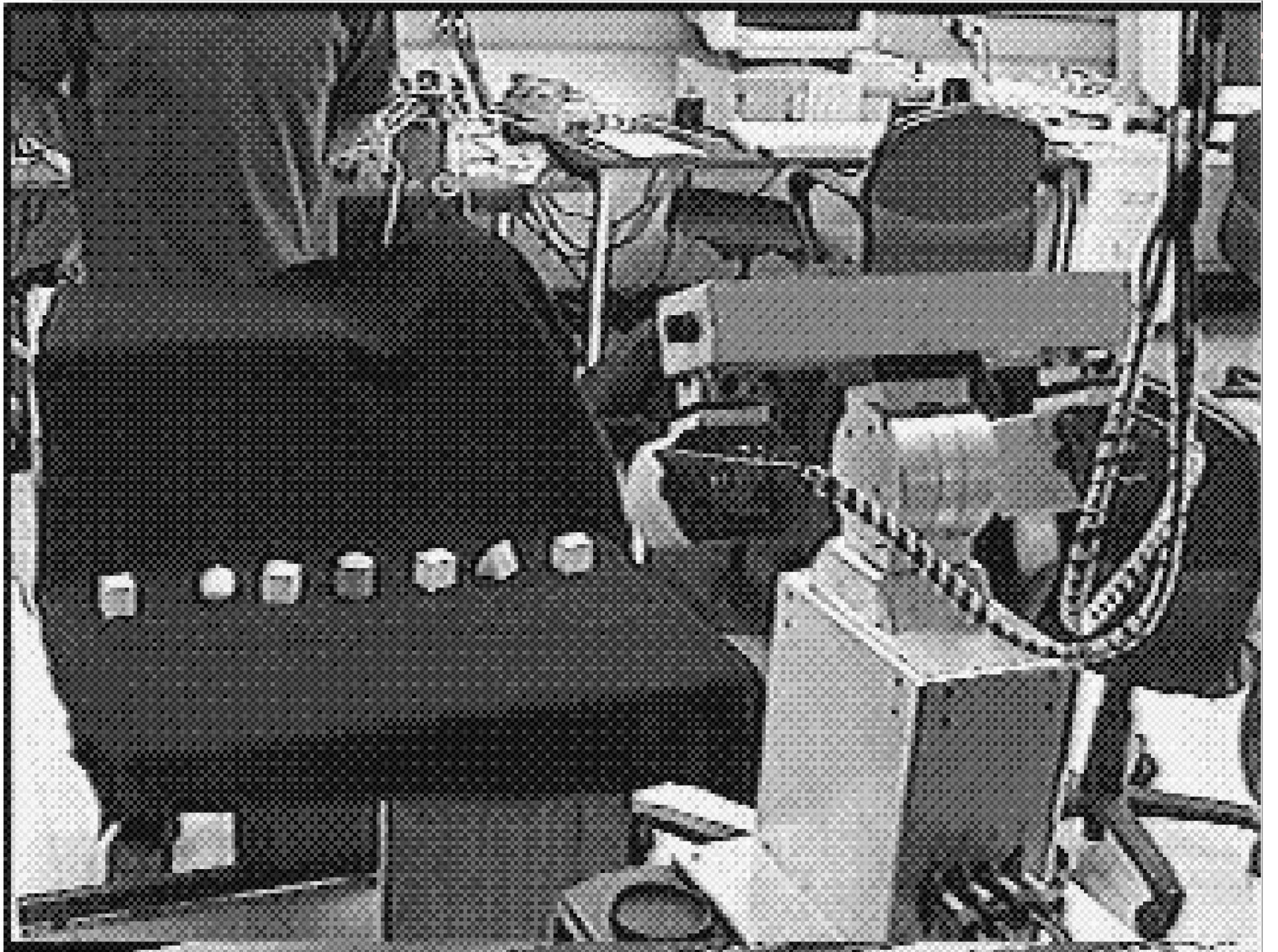
- Programação baseada em comportamentos

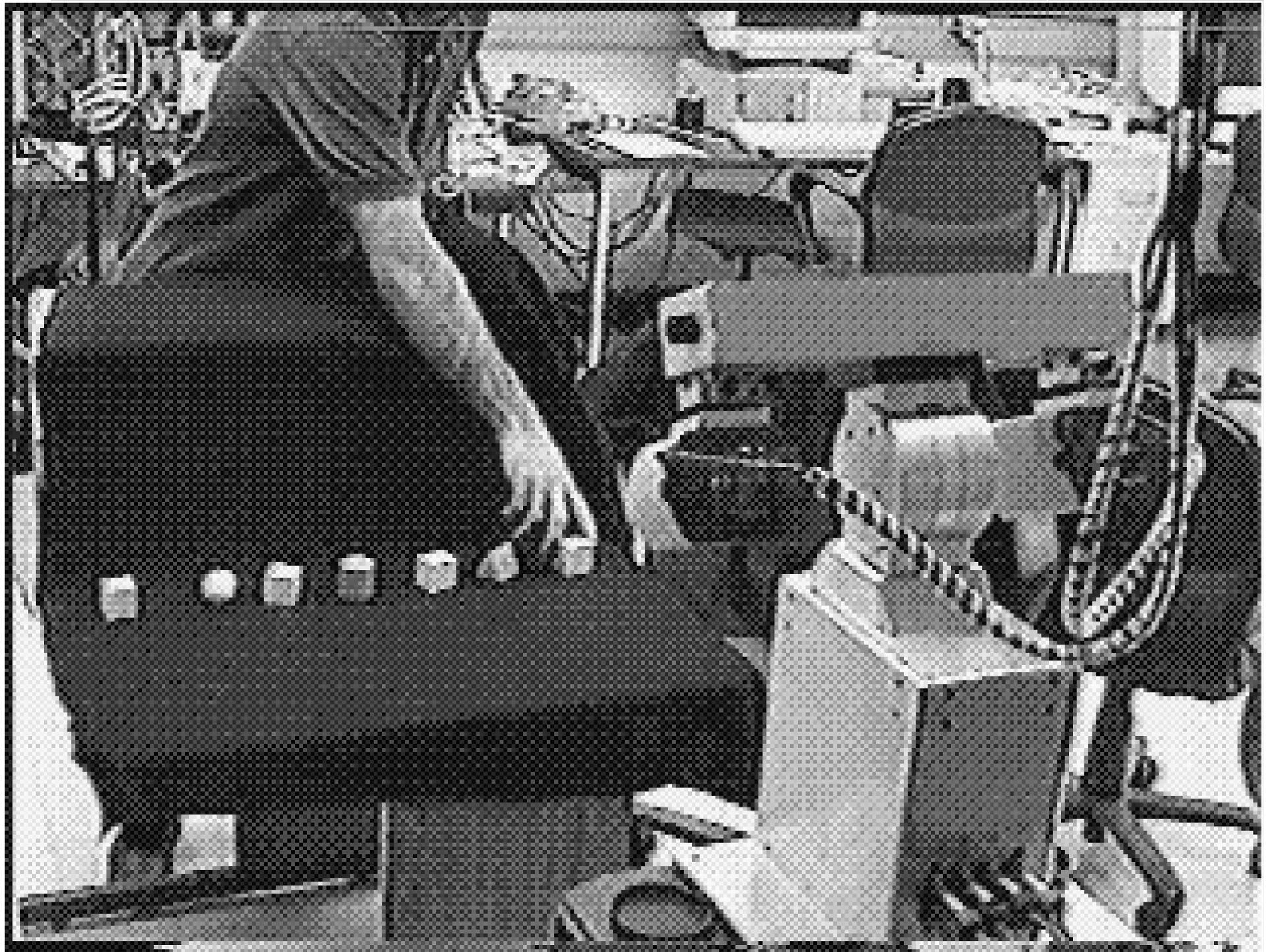


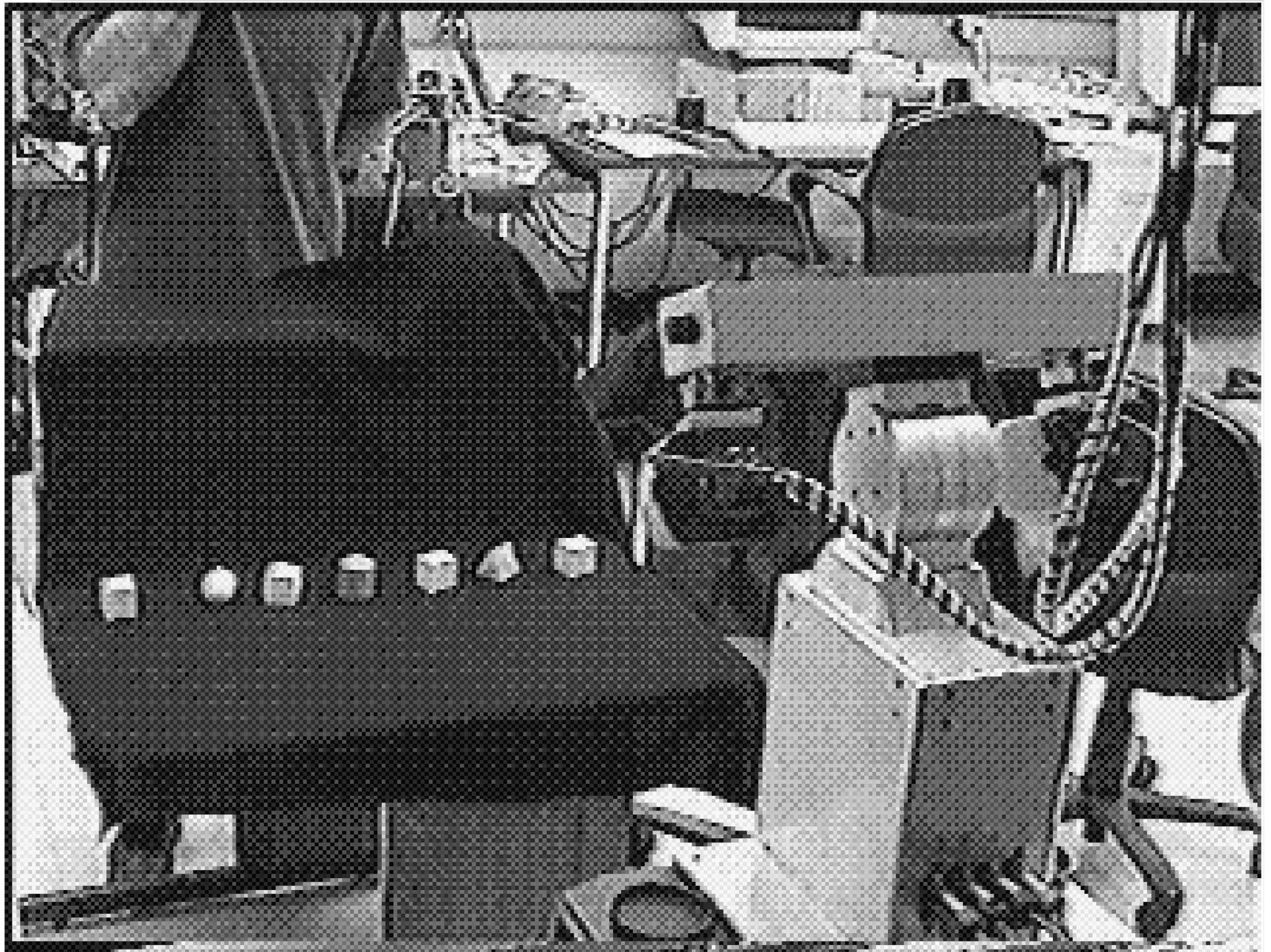
# Aplicações para a estrutura foveada

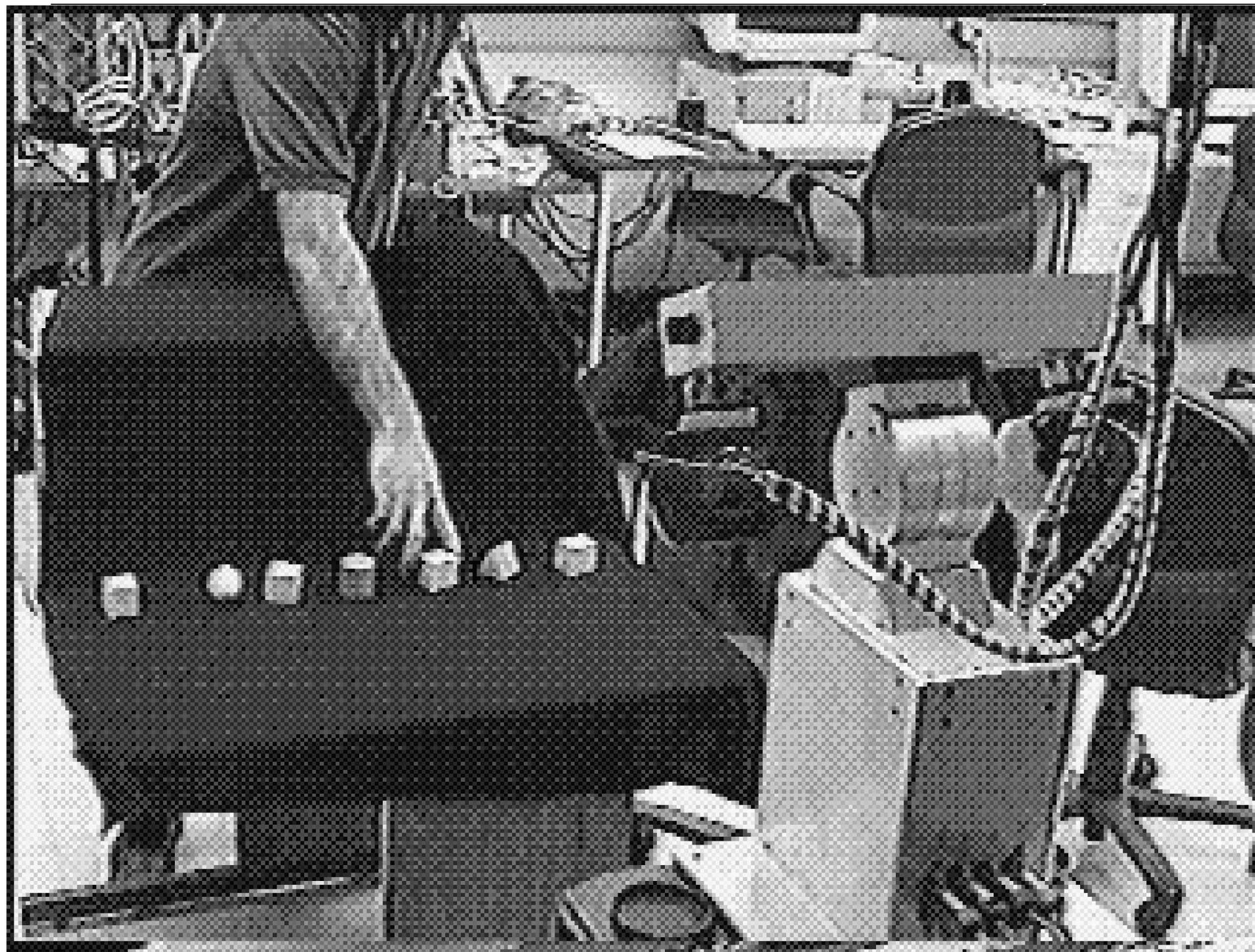
- Procedimento de atenção em execução

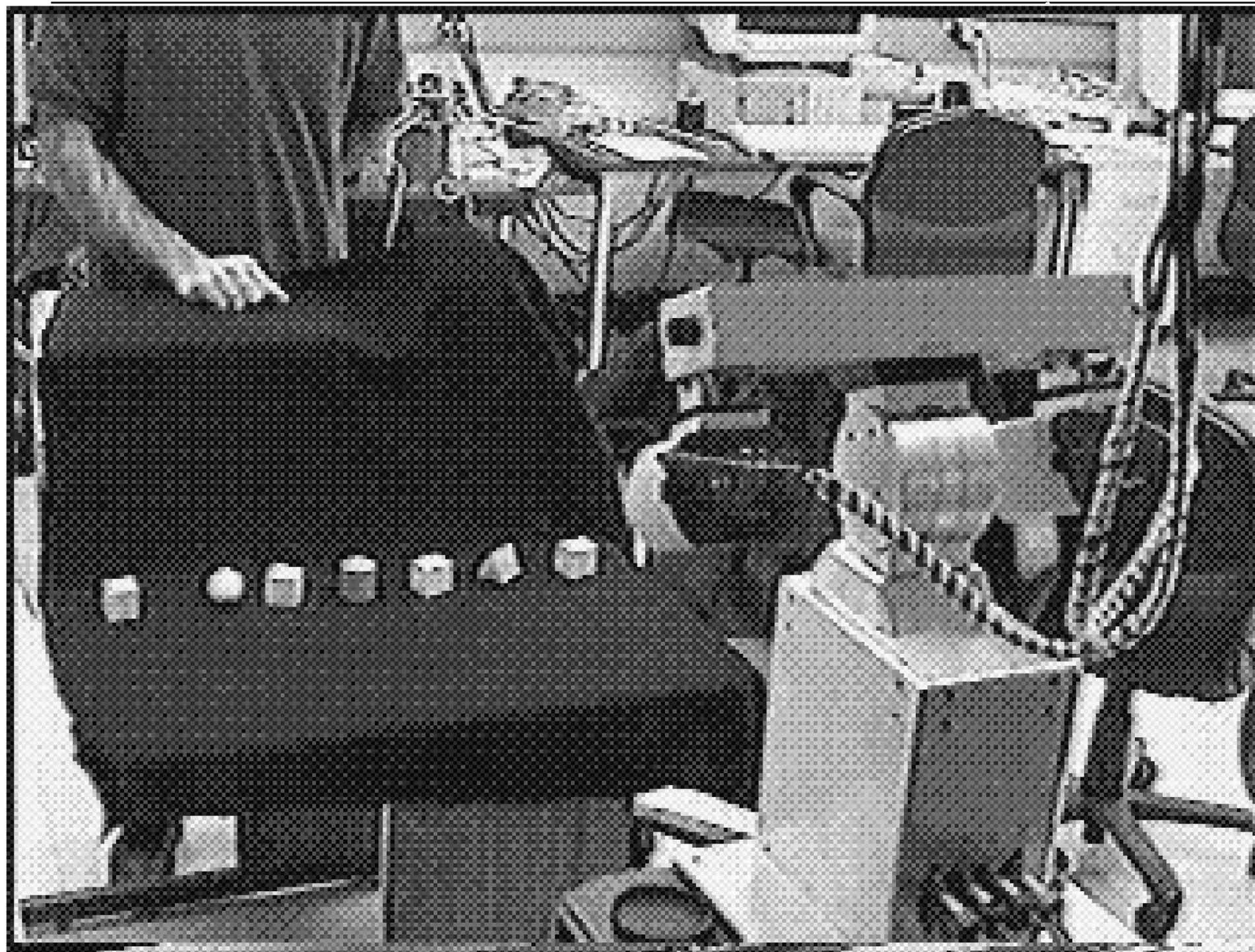


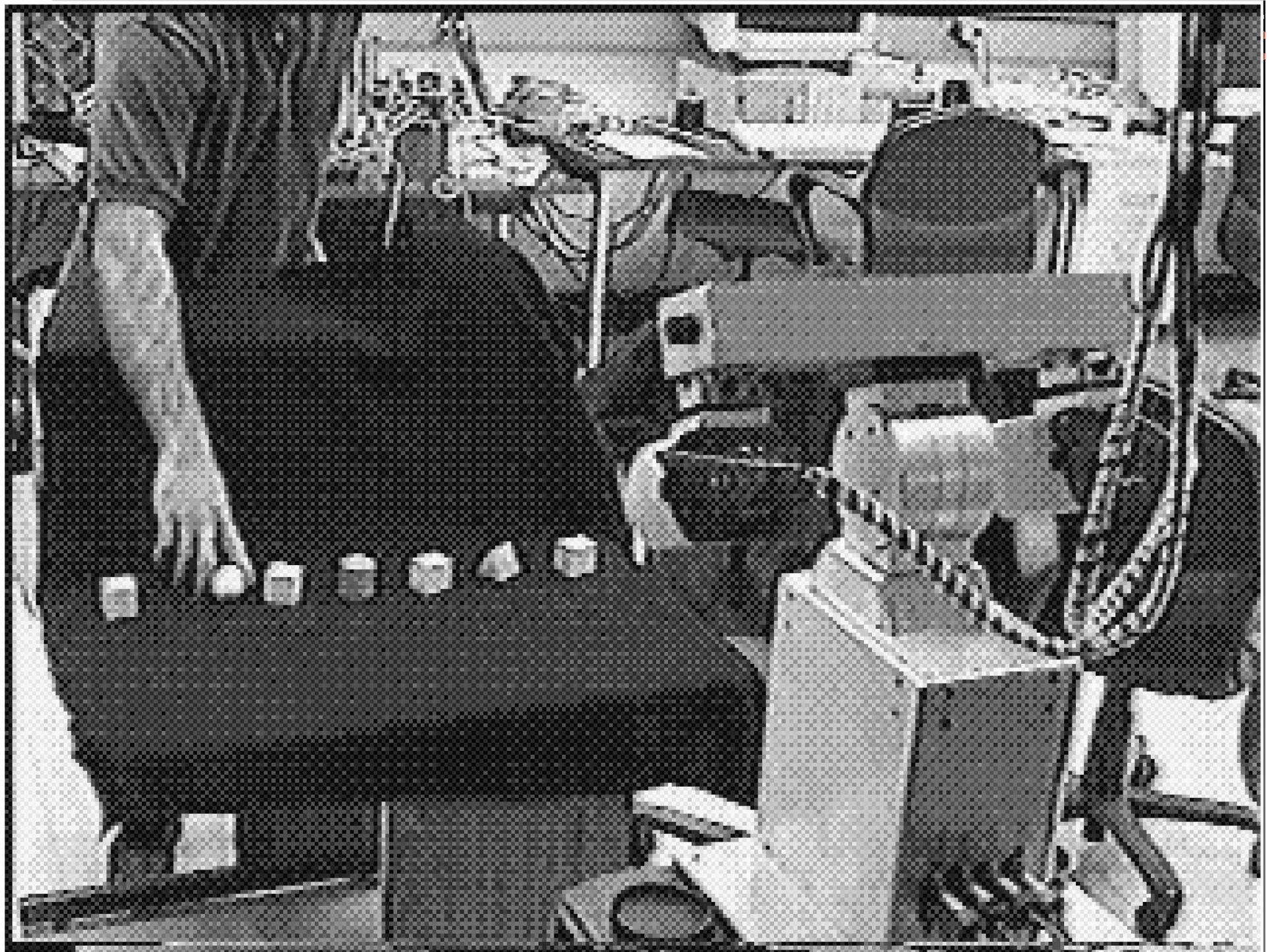


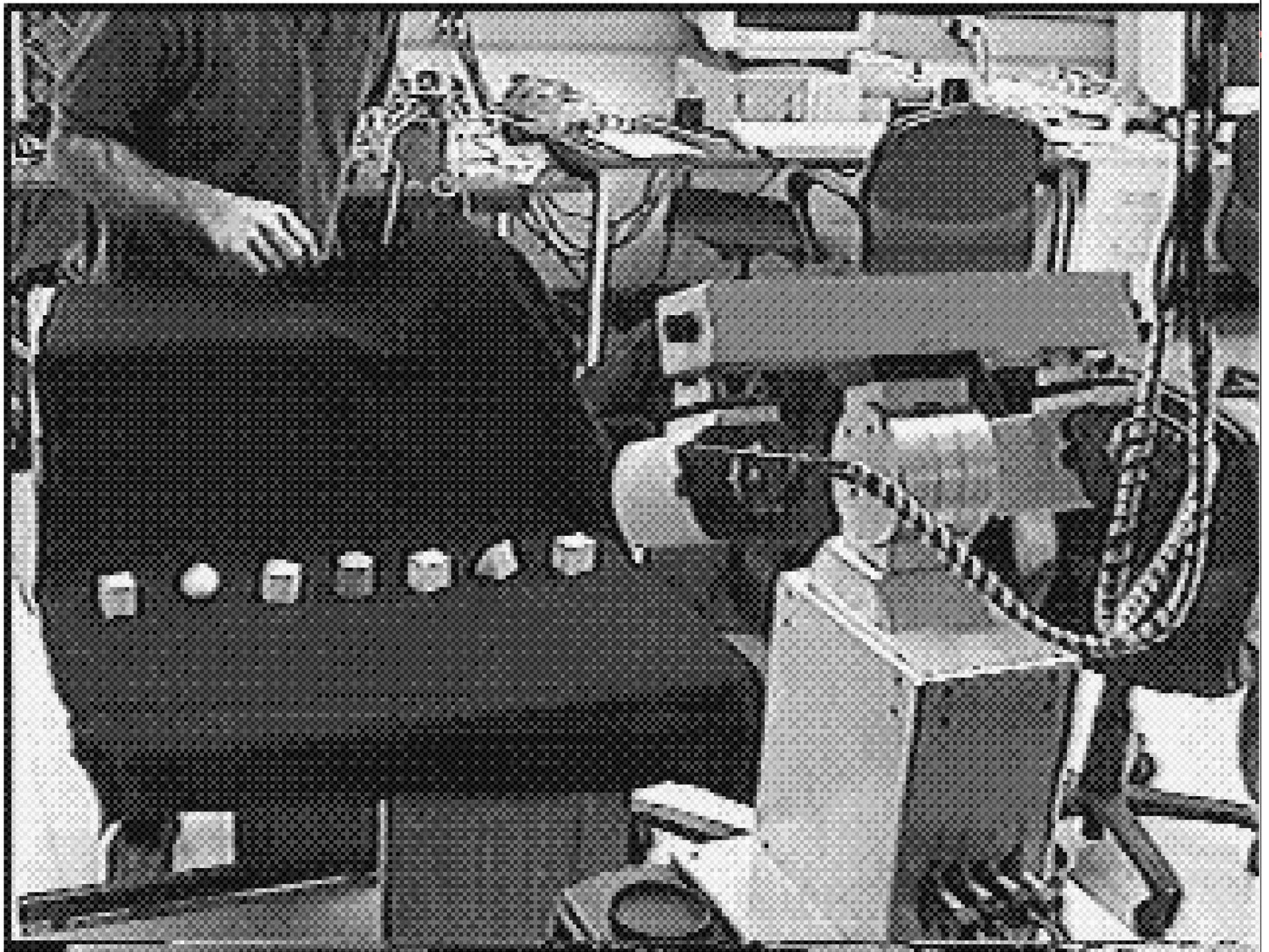






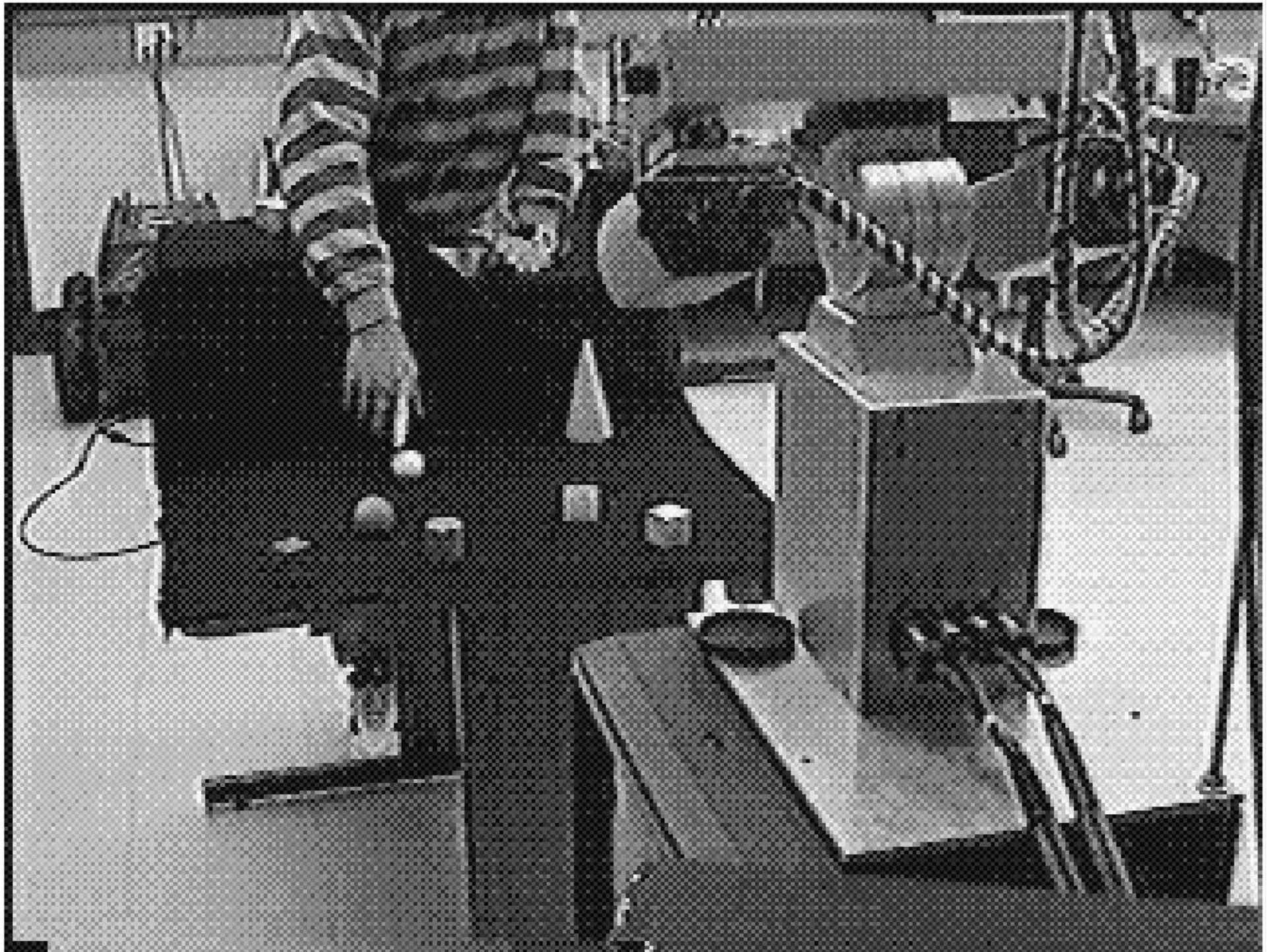




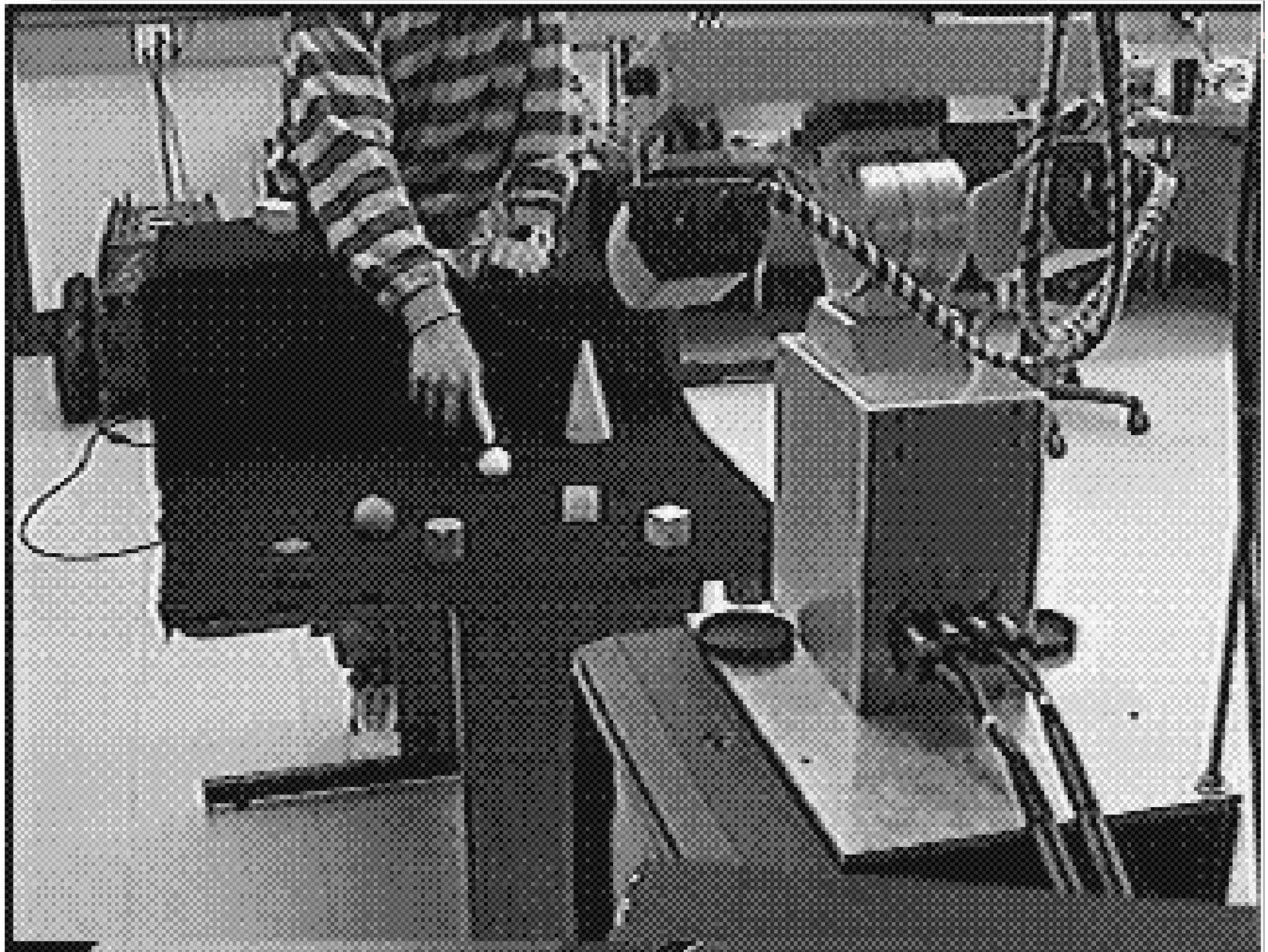


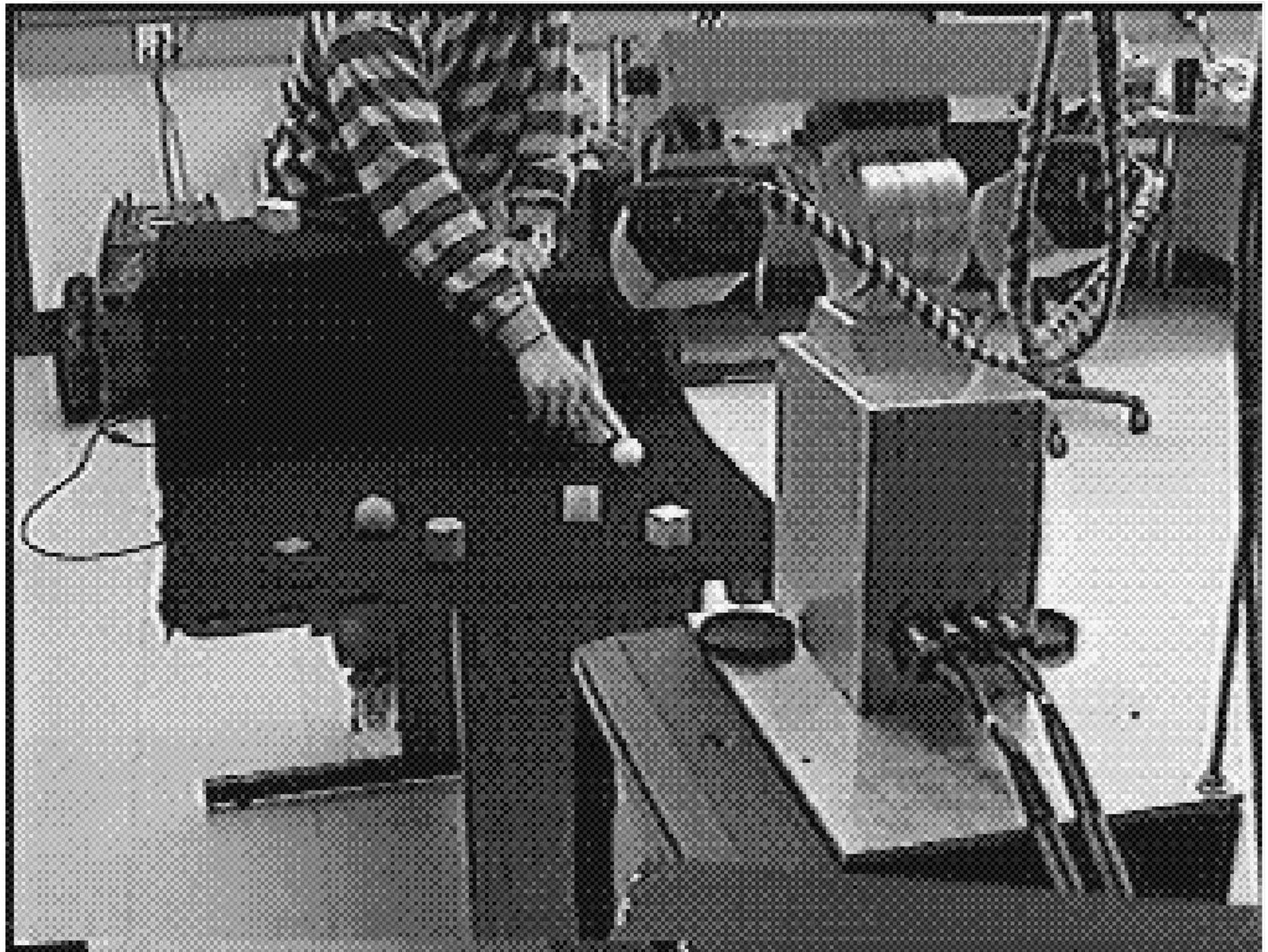
# Aplicações para a estrutura foveada

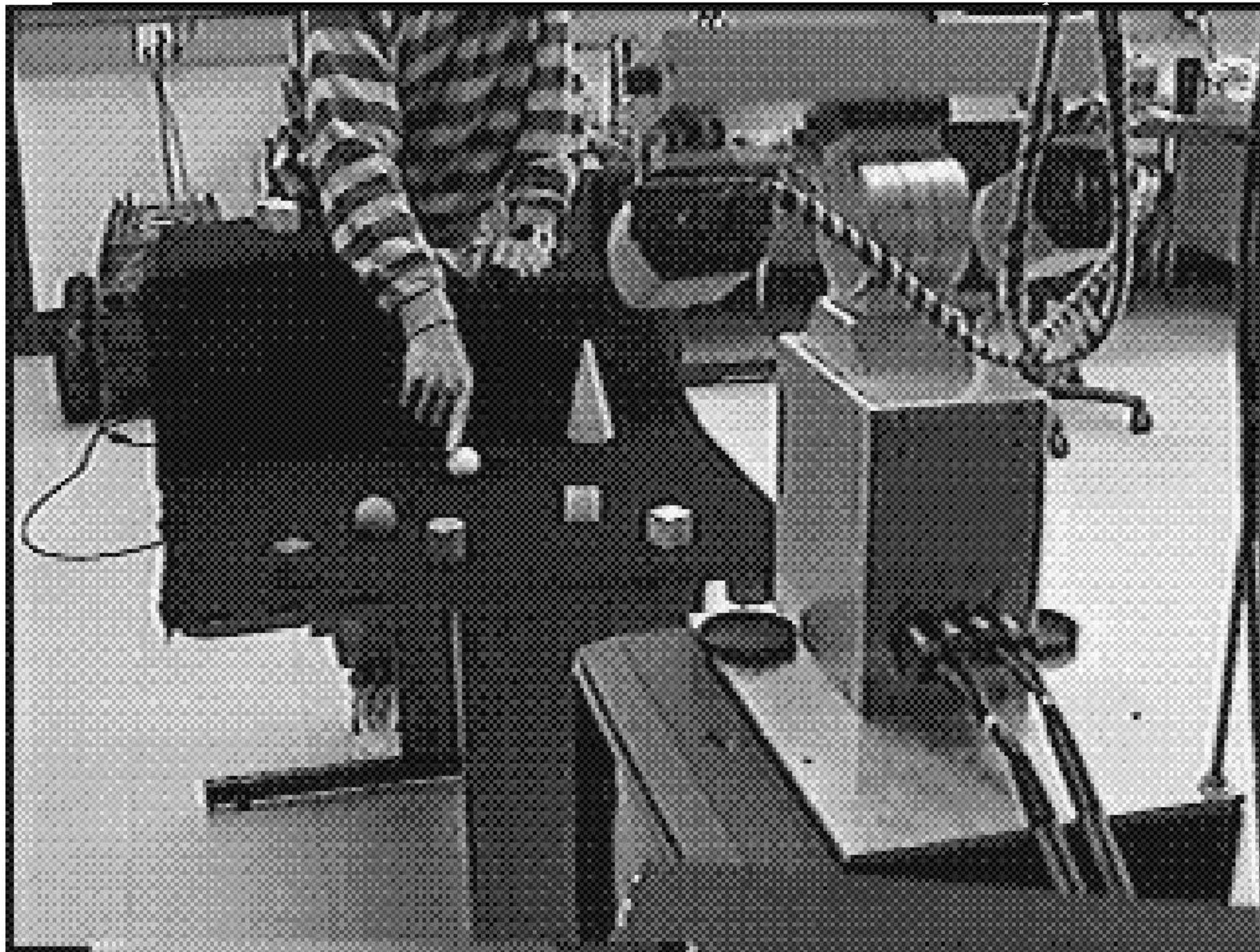
- Atualizando mapas do ambiente (tracking e reconhecimento , simultâneos)





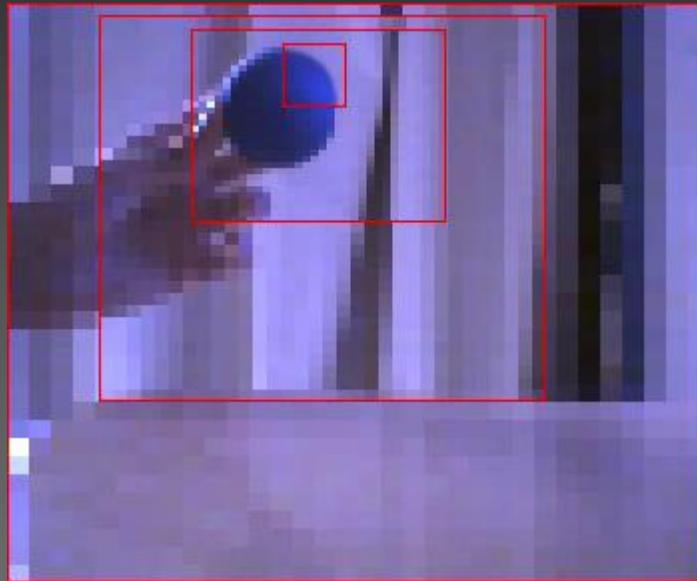


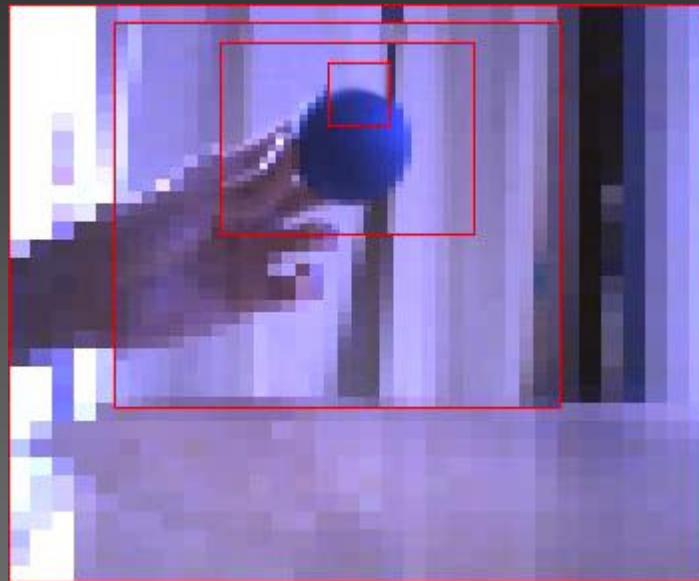


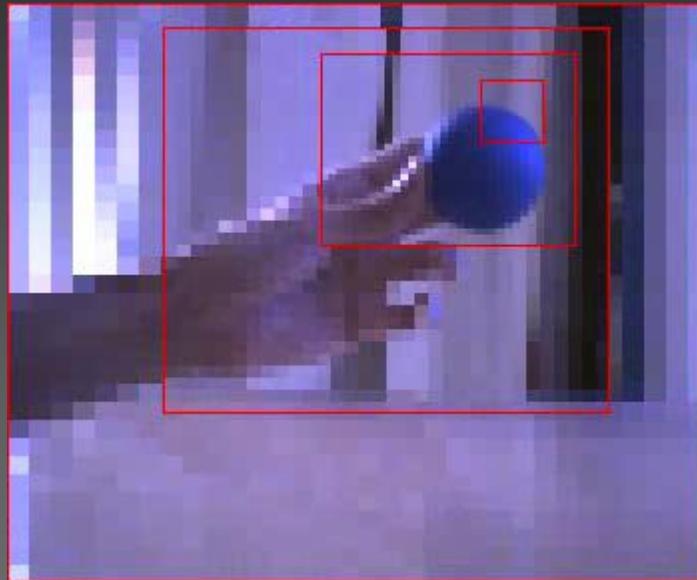


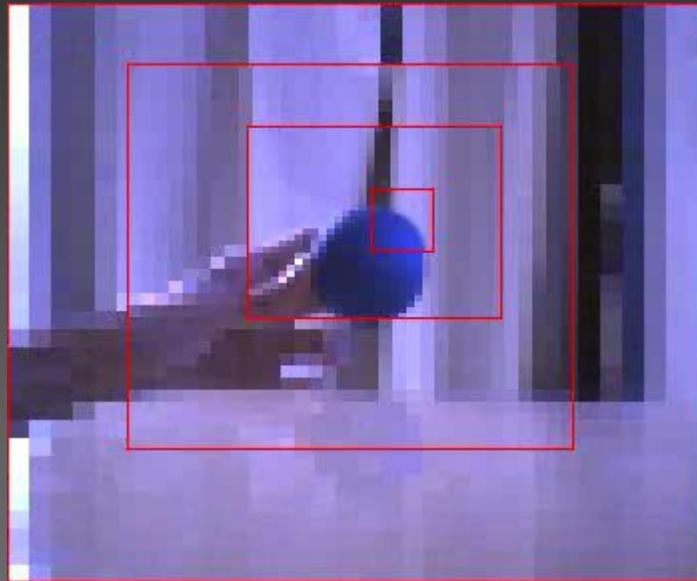
# Aplicações para a estrutura foveada

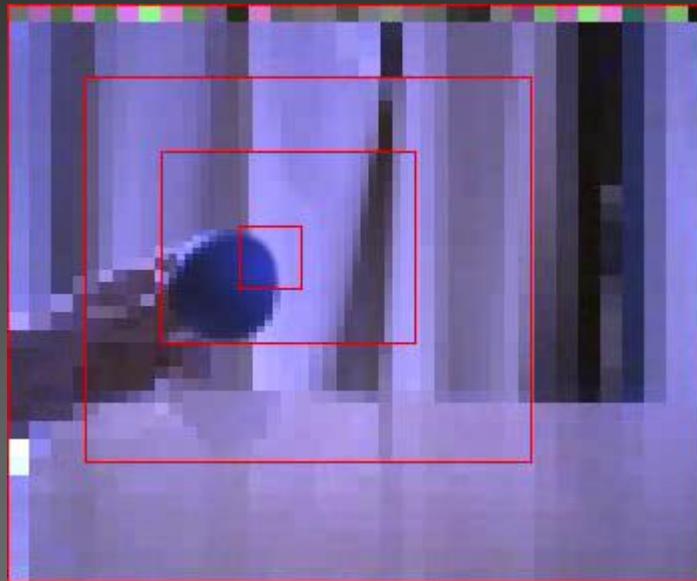
- Fóvea MR móvel executando tracking

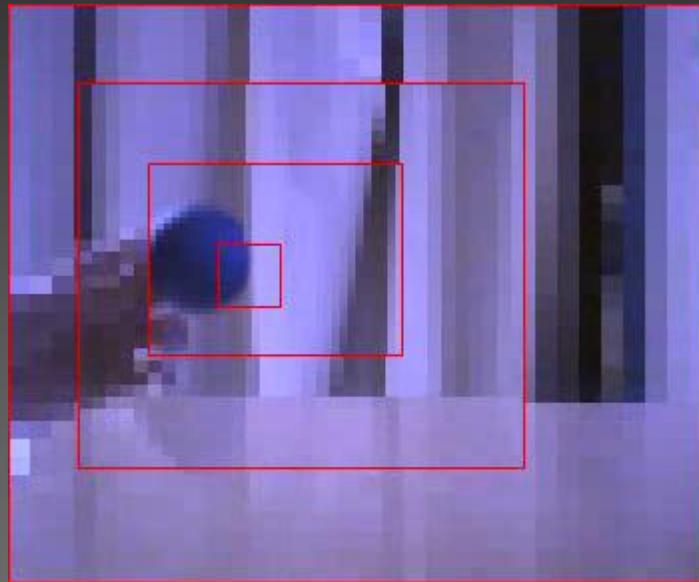








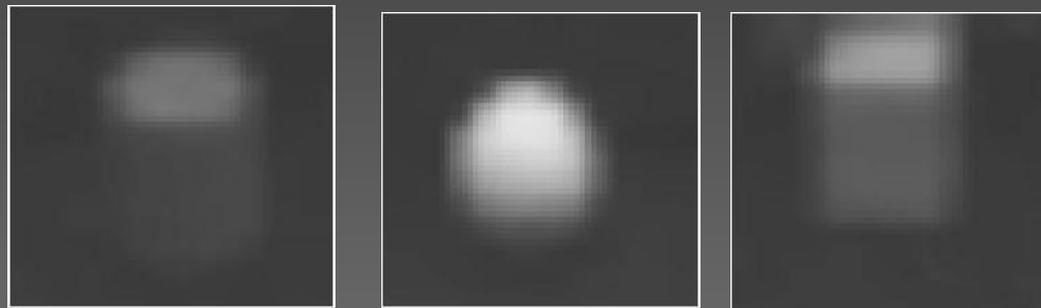




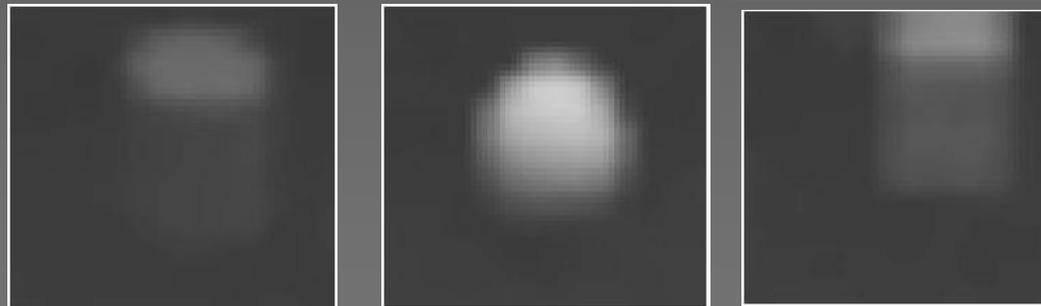
# Aplicações para a estrutura MCMR

- Executando vergência das câmeras (fóvea focalizada no objeto)

Esq



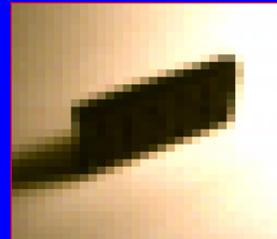
Dir



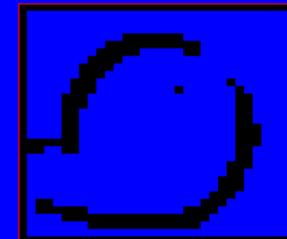
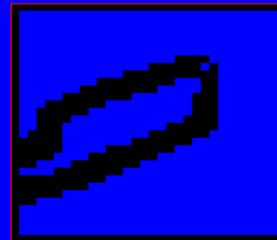
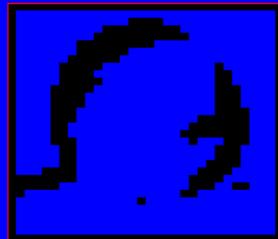
# Aplicações para a estrutura MCMR

- Reconhecimento (acertou mais de 85%)

Objeto



Feature



# Foveated Rendering Research

GTC 2017 in May 2017



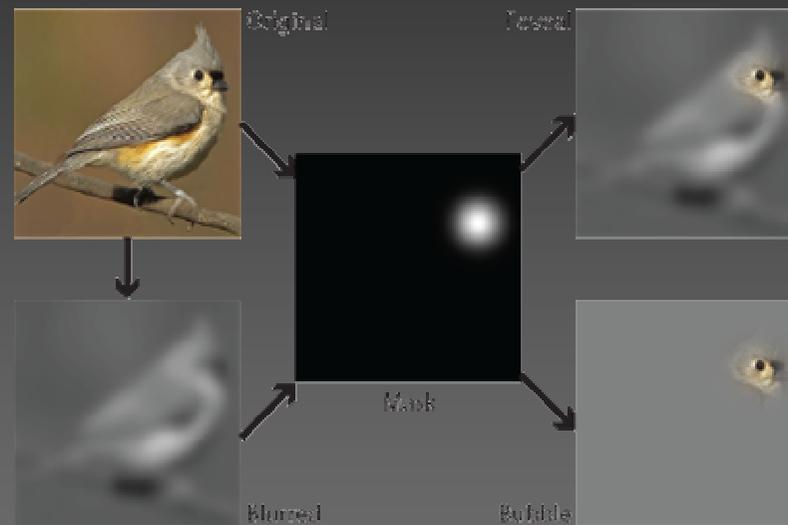
# GTC - 2017

Foveated rendering is a class of algorithms which increase the performance of virtual reality applications by reducing image quality in the periphery of a user's vision.

Anjul Patney: presented results from recent and ongoing work in understanding the perceptual nature of human peripheral vision, and its uses in improving the quality and performance of foveated rendering for virtual reality applications.

# Bubble-net

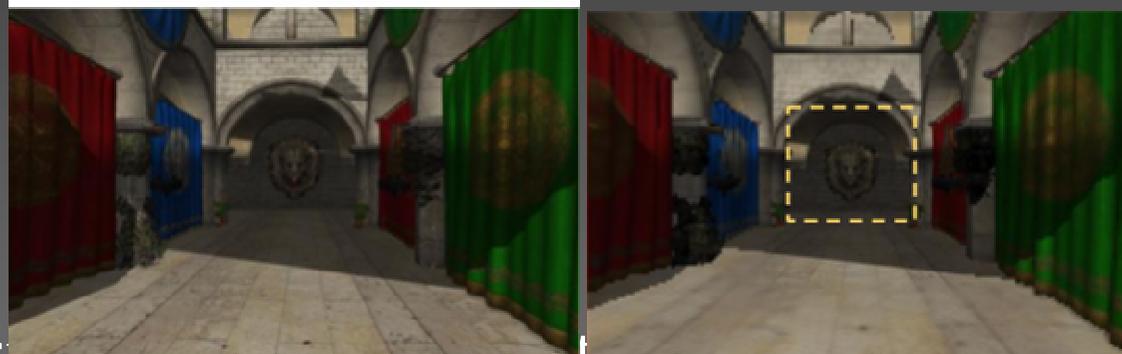
Foveated imaging for visual discovery



<http://www.cs.cornell.edu/projects/bubblenet/>

# Principle of Foveated Rendering

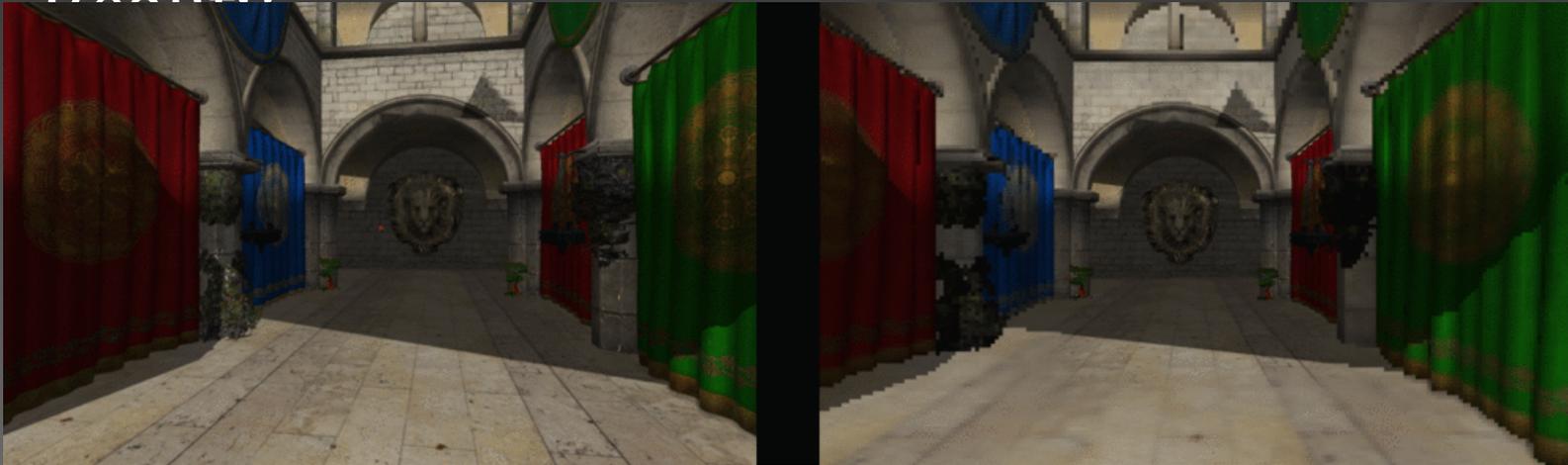
Location of the High Acuity (HA) region needs to be updated with eye-tracking to align with eye saccades, preserving constant high-resolution across the field of view.



In contrast, systems with no eye tracking may need to render a much larger HA region.

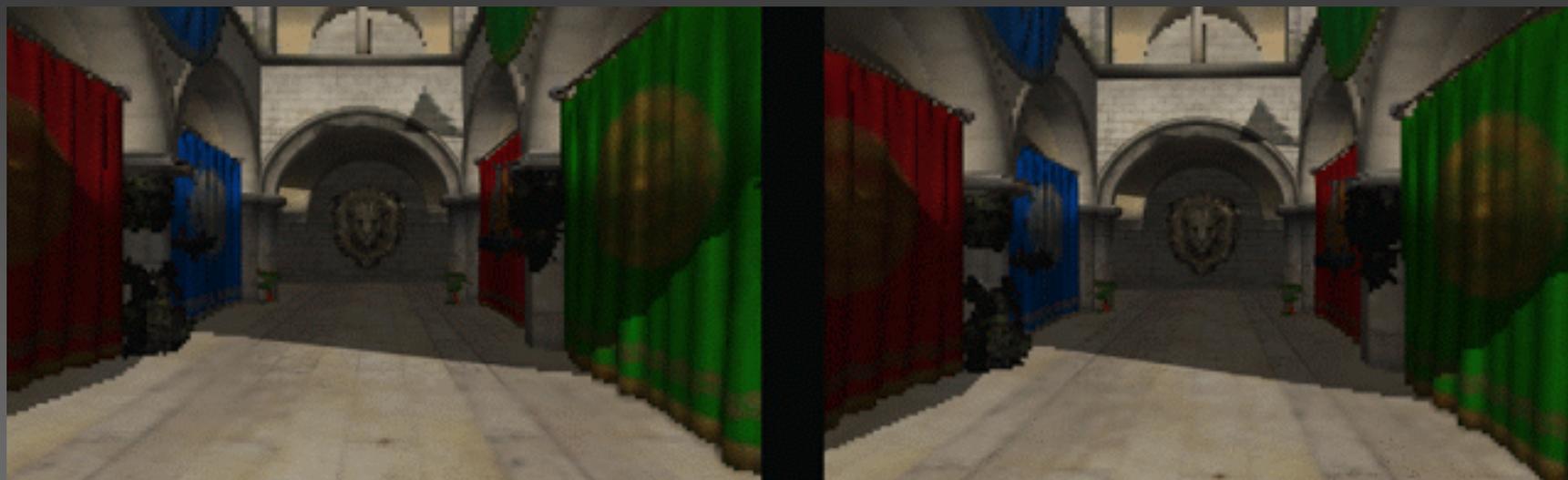
# Google AI Blog

## A New Foveation Pipeline for Virtual/Mixed Reality

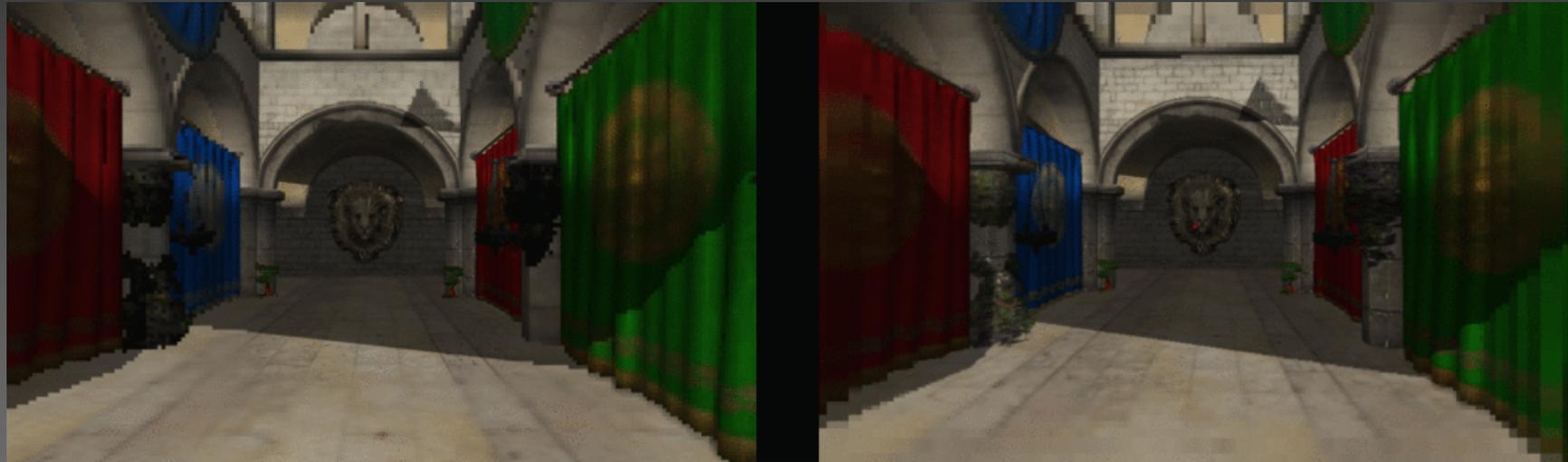


<https://ai.googleblog.com/2017/12/introducing-new-foveation-pipeline-for.html>

# Phase Aligned Rendering



# Conformal rendering



# Foveated transmission



Low spatial resolution content  
outside of fovea

High resolution content at fovea



A hybrid video frame buffer where  
content is no in display resolution  
and holds multiple sub-frames

# Foveated deep learning



# Foveated deep learning



# Considerações e sugestões de leituras

- Resumo do que foi visto na aula:
  - Foram apresentados modelos visando facilitar a redução e abstração de dados
  - Foi apresentada, de forma progressiva uma metodologia para redução e abstração de dados para aplicações de visão robótica utilizando um modelo de fóvea em multiresolução combinado com um modelo de fóvea móvel
  - Este modelo pode ser aplicado em tarefas envolvendo controle da atenção visual, reconhecimento de objetos e navegação.

# Considerações e sugestões de leituras

- Uma estrutura em multiresolução utilizando fóvea móvel pode ser construída em tempo real (cerca de 1.2ms em um PC 2 GHz, para 4 imagens com um tamanho de 64x48 pixels).
- Técnicas para extração de feições (reconstrução estéreo e filtros, movimento) podem ser empregadas em tempo real.
- O foveamento baseado em fóvea MR e fóvea móvel são as duas principais estruturas usadas.
- Futuro: multifoveamento (esperando resultado, artigo submetido).

# Sugestões de leituras

- Autores de trabalhos base para o texto:
  - Leonard Uhr [Uhr 1972], Andrew Witkin [Witkin 1983], Lindeberg, Bay, Lowe, Luiz Gonçalves [Gonçalves et al. 2000, Gonçalves & Grupen 2000a, Gonçalves, Oliveira & Grupen 1999b, Gonçalves, Oliveira & Grupen 1999a, Gonçalves, Giraldi, Oliveira & Grupen 1999, Gonçalves, Wheeler, Oliveira & Grupen 1999, Silva et al. 1999, Gonçalves et al. 2000, Gonçalves & Grupen 2000b, Gonçalves et al. 1998, Gonçalves & Oliveira 1998], Rafael Beserra [Beserra 2008, 2011], José Sávio Segundo [Segundo & Gonçalves 2004] e Marcos Medeiros 2009.

# Agradecimentos

Cópias do texto, incluindo arquivos fontes e imagens, podem ser solicitadas a

[lmarcos@dca.ufrn.br](mailto:lmarcos@dca.ufrn.br)

Agradecimentos aos autores e aos alunos e professores dos Laboratórios de Robótica da UMASS e Natalnet da UFRN que contribuíram com a teoria apresentada.

[www.natalnet.br](http://www.natalnet.br)

Dúvidas e perguntas?

