

O que se entende por visão por computador?

- Área da ciência da computação focada na **extração** de "**informação útil**" de imagens e vídeos.
- Objetivo: "**descobrir a partir de imagens o que está presente no mundo, onde estão localizadas as coisas, que acções estão a ser executadas**" (Marr,1982).
- Exemplos de "informação útil" incluem, por exemplo, o cálculo da geometria 3D de um objecto presente numa imagem, a detecção e identificação de faces ou de gestos humanos, o seguimento de pessoas ou de veículos numa sequência de vídeo.
- Os algoritmos da visão por computador têm **imensas aplicações** em muitas áreas, desde a **industrial, militar ou médica** até à sempre crescente área do entretenimento.

Funcionamento de uma camara digital (RGB)

- **Regista imagem** através de um **sensor** – do tipo **CMOS** ou **CCD** – e **armazena** imagem em **cartões de memória**
- Suporta uma só ou vários tipos de memória
- **Sensor de imagem digital**: funciona como uma retina de olho, capta a luminosidade das imagens que são projetadas sobre ele continuamente e dá início ao processo de captura de uma instancia ou de uma sequencia de instancias da imagem consecutivamente.

Parâmetros possíveis de ajustar numa camara digital

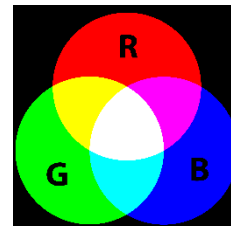
- **Abertura da lente**: lente (abertura através da qual a imagem chega ao sensor); Sistema de abertura é controlado pela camara. A camara envia o comando e o sistema fecha ou abre o mecanismo que abre o orifício por onde a “imagem passa”
- Espécie de membrana que pode ser mais ou menos fechada, deixando passar mais ou menos luz
- Quanto maior a abertura do diafragma, mais a imagem ficará exposta
- **Tempo de exposição**: quantidade de tempo que o obturador da maquina leva a abrir e a fechar, deixando passar a luz que irá sensibilizar a pelicula fotografia ou o sensor digital e formar a imagem.
- Quanto menor o tempo de exposição, menos luz é absorvida no interior da maquina, maior abertura do diafragma necessária para se obter uma exposição correta.
- **Distancia focal**: é através dela que o user define a maior ou menor aproximação de uma imagem, ou ainda o campo de visão que deseja trabalhar.

Definição de espaço de cor

- **Modelo matemático** (abstrato) usado para **descrever** cada **cor** a partir de **formulas**
- Formalizar a descrição de cores através de **tuplos de números**, tipicamente formados por **3 ou 4 elementos**
- Sistema definido por uma **base representativa dos componentes**, de acordo com a definição do **espaço considerado**.
- A representação de qualquer cor pode então ser feita à custa da combinação desses componentes. Normalmente são tridimensionais.

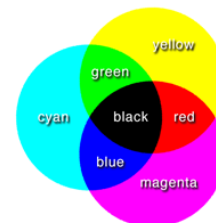
Modelo de cor: RGB

- Sistema de cores **aditivas** - Red Green Blue
- **Objetivo**: reprodução de cores em dispositivos eletrônicos (e.g. TV, PC, retroprojetores, câmeras digitais, tradicionais...)



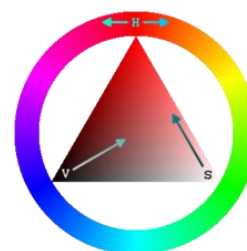
Modelo de cor: CMY

- Sistema de cores **subtrativas** - Ciano Magenta Amarelo – preto (key)
- **Objetivo**: funciona devido à absorção de luz, pelo fato de que as cores que são vistas vêm da luz que não é absorvida.
- É usado em impressoras e derivadas que reproduz a maioria das cores do espectro visível.



Modelo de cor: HSV ou HSB

- Sistema de cores formadas pelas componentes - tonalidade, saturação e brilho
- **Tonalidade**: verifica o tipo de cor, abrangendo todas as cores do espectro, desde o vermelho até ao violeta, mais o magenta.
- **Saturação**: também chamado de pureza. Quanto menor esse valor, mais com tom de cinza aparecerá a imagem. Quanto maior o valor, mais pura é a imagem
- **Brilho**: define o brilho da cor



Tipos de filtros

- Para que serve? Suaviza uma imagem e permite remover ruídos, perdendo pequenos detalhes
- **Média:** mais fácil de implementar. O kernel é uma matriz com todos os valores igual a um (o pixel é substituído por uma média dos vizinhos $N \times M$).

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

25	25	27
24	25	26
22	21	23

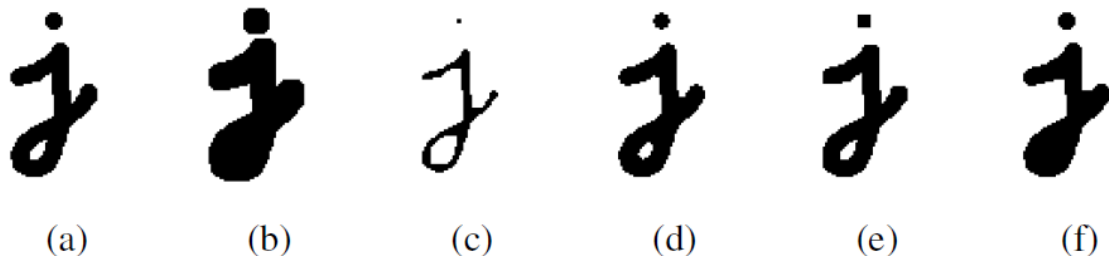
 \rightarrow

	24	

- **Binomial:** usa coeficientes binomiais como pesos para dar mais ênfase a pixels próximos ao centro da vizinhança $N \times M$.
- **Gaussiana:** usa a função gaussiana para definir os pesos de vizinhança.

Operações binárias

- São conhecidas como operações morfológicas. A operação consiste numa convolução da imagem binária com um elemento estruturador binário.
- **Dilatação:** estampar o elemento estruturante em cada pixel do objeto; versão expandida
- **Erosão:** operação dual da dilatação; dilatação do fundo; versão contraída
- **Abertura:** erosão seguida de dilatação
- **Fecho:** dilatação seguida de erosão – dual à abertura



Histogramas e suas operações

- Exibição gráfica de **frequências**
- Tipicamente representado por um **gráfico de barras**
- Um histograma de imagens permite-nos ver a **distribuição** de **cores** ou **intensidades**
- **Operações** – permitem aumentar a visibilidade dos objetos de interesse numa imagem
- **Histograma Equalização** - melhora o contraste de uma imagem, a fim de esticar a faixa de intensidade.
- **Equalização do Histograma Local** - aumenta a quantidade de realce observando as propriedades de intensidade local (dividindo uma imagem em regiões e realizando o histograma de equalização em cada sub-imagem ou usando estatísticas locais).
- **Histograma Comparação** - obtém um parâmetro numérico que expressa o quão bem dois histogramas correspondem uns aos outros (ex .: Correlação, Qui-Quadrado, Intersecção, ...).
- Soma, subtração...

Segmentação

- Processamento intermediário para o reconhecimento de objetos.
- Localizar regiões com propriedades comuns.
- Partição sobre o conjunto de pixels; dividir imagem em regiões que tem algum significado
- Propriedades de agrupamento habituais (nível de cinzento, cor, textura).
- Muitas vezes requer pré-processamento.
- A segmentação de imagens não é trivial - tarefa difícil.
- A precisão de segmentação determina o sucesso

Segmentação: RGB ou HSV

- O HSV ao contrario do RGB separa a imagem, ou pela intensidade, chroma ou da info da cor
- Isso é muito útil em muitas aplicações. Por exemplo, se você deseja fazer a equalização do histograma de uma imagem colorida, provavelmente deseja fazer isso apenas no componente de intensidade e deixar os componentes de cores sozinhos. Caso contrário, você terá cores muito estranhas.
- Na visão por computador, muitas vezes você quer separar os componentes da cor da intensidade por várias razões, como robustez para mudanças de iluminação ou remoção de sombras.
- Note, no entanto, que HSV é um dos muitos espaços de cores que separam cor de intensidade. H SV é frequentemente utilizado simplesmente porque o código para converter entre RGB e HSV está amplamente disponível e também pode ser facilmente implementado.
- RGB tem a ver com "detalhes de implementação" sobre a forma como RGB exhibe cor, e H SV tem a ver com os componentes de "cor real". Outra maneira de dizer isso seria RGB é a forma como computadores trata cores, e H SV tentar capturar os componentes da forma como os seres humanos percebem a cor.

Thresholding

- A base de muitos algoritmos de segmentação baseados em regiões.
- O passo mais imediato e computacionalmente atraente.

Imagens grayscale/vantagens

- Permite maior facilidade, na construção e manipulação de matrizes de características

Edge Detection

- Permite determinar **pontos** de uma imagem digital em que a **intensidade luminosa muda repentinamente**. – permite localizar **mudanças de funções de intensidade**.
- Mudanças repentinas em imagem geralmente refletem eventos importantes no cenário, como a **descontinuidade da profundidade** (transição entre o objeto e o fundo), **descontinuação da orientação da superfície**, mudança das **propriedades do material** ou variações na **iluminação da cena**.
- A sua utilização **reduz significativamente a quantidade de dados** a serem **processados**, e descarta informação que é considerada **menos** relevante, ainda que **preservando** importantes **propriedades** estruturais de uma imagem.
- Detecção pode ser **prejudicada** pela detecção de **falsas bordas** devido a **ruídos** existentes (digitalização, compressão...) – pode ser resolvido retirando ruído.

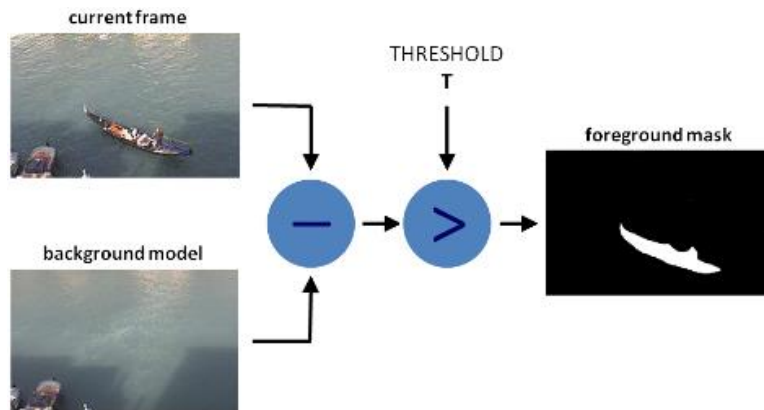
Edge Detection: Canny

- Utiliza um algoritmo multi-estágios para **detetar** uma **ampla margem de bordas** na imagem
 - O critério de detecção exprime o facto de não se deverem perder pontos importantes e de não haver respostas espúrias.
 - O critério de localização diz que a distância entre a posição real e a posição localizada da aresta deve ser mínima.
 - O critério de uma resposta minimiza respostas múltiplas a uma única aresta.

Corner Detection

- Abordagem utilizada em VC para **extrair** certas **características** e **inferir** o **conteúdo** de uma imagem.
- Detecção de cantos sobrepõe-se à extração de pontos de interesse.
- É **usado** normalmente: detecção de movimento, registo de imagens, controlo de vídeo, modulação 3d e reconhecimento de objetos
- Um **canto** pode ser definido como uma **intersecao de duas bordas** ou por um **ponto** que tem **duas direções** de borda.
- **Ponto de interesse**: ponto numa imagem que tem uma **posição bem definida** e pode ser **robustamente detectado**. Ex: cursa, ponto isolado de baixa/elevada intensidade, fim de linha...
- **Exemplo**: Harris Corner Detection

Background subtraction



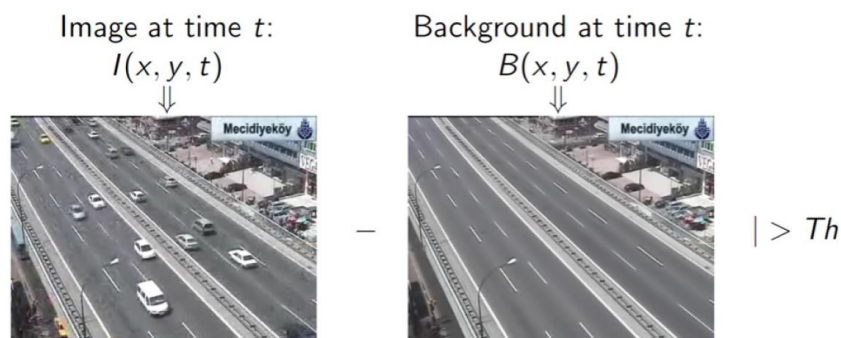
- É uma técnica comum e amplamente utilizado para gerar **uma máscara de primeiro plano** (ou seja, uma imagem binária que contém os **pixels pertencentes a movimentação de objetos na cena**) usando **câmaras estáticas**.
- Consiste em **comparar** uma **imagem observada** com uma mesma imagem retirada do mesmo local sem **nenhum objeto de interesse**.
- As áreas do plano de imagem onde há uma **diferença** significativa entre as imagens observadas e estimadas indicam a **localização dos objetos de interesse**.
- O BS vem da técnica simples de subtrair a imagem observada da imagem estimada e limiar o resultado para gerar os objetos de interesse.
- **Segmentar objetos de interesse** numa cena para aplicações como: Vigilância; Visão do robô; Rastreamento de objetos; Aplicações de tráfego; Captura de movimento humano; Realidade aumentada

Problemas importantes

- **Foreground detection** - como as áreas de objeto são distinguidas do fundo;
- **Background maintenance** - como o fundo é mantido ao longo do tempo;
- **Post-processing** - como as áreas de objeto segmentado são pós-processadas para rejeitar falsos positivos.

Algoritmo genérico

- Crie uma imagem do fundo estacionário fazendo uma média de uma sequência longa.
- Diferença de um quadro a partir do quadro de fundo conhecido
- **Algoritmos de detecção de movimento como estes só funcionam se a câmera estiver parada e objetos estiverem movendo contra um fundo fixo**

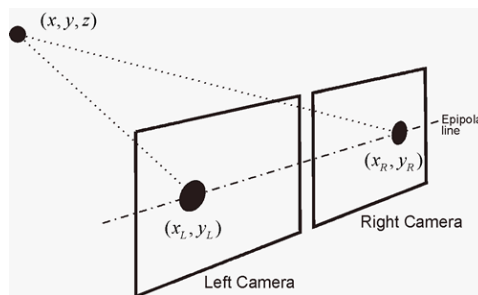


Stereo Vision

- Imita a forma como os humanos vêem os objetos
- Cada um dos olhos vê uma imagem ligeiramente diferente (parallax - deslocamento horizontal)
- A quantidade de deslocamento depende da "distância inter-pupilar" (IPD) (varia para cada pessoa na faixa de 53-73 mm)
- O cérebro combina essas duas imagens o que nos dá uma percepção 3D

Em visão por computador...

- Olhos: duas camaras colocadas em posições estratégicas para obter duas imagens dif.
- SW calcula a posição 3D dos objetos baseada nas diferenças entre essas duas imagens
- Problemas: dependendo das posições dos objetos é preciso mudar os ângulos entre as camaras e o seu foco. É sempre difícil obter um modelo 3D realmente bom.



Capacidade de definir a profundidade a partir de 2 imagens

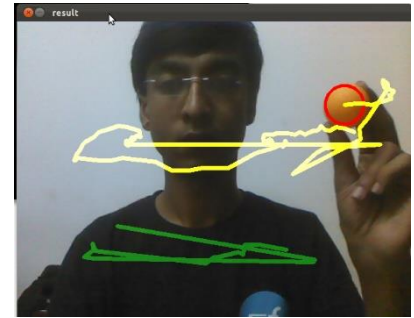
Possível por computação de correspondências entre duas imagens

Passos

<ul style="list-style-type: none"> – Remove lens distortion (<i>undistortion</i>) – Adjust distance and angles to obtain rectified images (<i>rectification</i>) – Find stereo correspondences to produce the disparity map (difference between x coordinate in rectified images) – Transform disparity map in depth map if camera parameters available 	<ul style="list-style-type: none"> • Calibrate cameras • Rectify images • Compute disparity • Estimate depth
---	--

Tracking

- É uma **questão crucial de pesquisa em visão computacional**, especialmente para as aplicações onde o **ambiente está em contínua mudança**: Robot Vision - navegação robô móvel, aplicações que devem lidar com apertos instáveis,... Vigilância; Aplicações de tráfego; Captura de movimento humano;



Detecção: Detectamos o objeto independentemente em cada frame e podemos registrar sua **posição ao longo do tempo**, por exemplo, com base nas coordenadas da janela de detecção

Rastreamento com dinâmica: usamos medidas de imagem para estimar a posição do objeto, mas também incorporamos a **posição prevista pela dinâmica**, isto é, a expectativa do padrão de movimento do objeto.

Problemas:

- **Inicialização** - Muitas vezes feito manualmente (Subtração de fundo, detecção também pode ser usado)
- **Associação de dados, múltiplos objetos controlados** - oclusões, desordem
- **Objetos deformáveis e articulados**
- Construindo **modelos** precisos de **dinâmica** - (exemplo: parâmetros de montagem para um modelo de dinâmica linear)
- **Deriva** - acumulação de erros ao longo do tempo

Template matching

- Técnica usada para encontrar áreas de uma imagem que correspondam (são semelhantes) a uma imagem do modelo (*template*).
- São necessários dois componentes:
 - **Imagem original:** imagem onde se espera encontrar uma imagem do *template*
 - **Imagem template:** imagem que vai ser comparada com a imagem original, deslizando-se.
- O *template* é comparado a todas as sub-regiões possíveis de uma imagem maior.
- **Problemas:** diferentes rotações / diferentes escalas



Sistema de detecção de objetos

- **Recursos de baixo nível:** refletir diretamente recursos específicos da imagem e vídeo: cor, textura, forma, movimento...
- **Recursos de medio nível:** muita subjetividade; normalmente existem múltiplas soluções; exemplos: Segmentação, Fluxo Óptico, Identificação
- **Recursos de alto nível:** Interpretação Semântica; Conhecimento; Contexto. Os humanos fazem isto muito bem!
- **Várias abordagens para classificação / reconhecimento**
 - **Shape:** descritor de formas parecidas
 - **Aparencia:** valores de pixéis semelhantes
 - **Geometria:** estrutura similar e locais similares com parâmetros similares
 - **Graficos:** relações de peças similares
 - **Bag of words:** descritores de características locais semelhantes

Parâmetros intrínsecos

- Permite relacionar as **coordenadas pixel relativas às imagens** com as coordenadas de **pontos do espaço medidos no sistema referencial com origem no centro da câmara**, ou seja, **transforma as coordenadas 3D no sistema câmara em coordenadas 2D no sistema da memória frame**.
- Estes parâmetros **dependem** das **características físicas** da câmara (da sua geometria interna e do tipo de lente) , isto é, inclui a informação associada à projeção perspetiva e resolução, alinhamento do sensor da imagem e num outro conjunto de coeficientes associados à distorção na imagem, provocada pelo facto da lente não ser perfeita.

Parâmetros extrínsecos

- São utilizados para **transformar as coordenadas 3D no sistema** de coordenadas do **mundo em coordenadas 3D no sistema da câmara**, para cada **posição e/ou orientação da câmara/plano de calibração**.

Transformação de Hough

- Técnica matemática que permite **detetar figuras geométricas** em imagens digitais
- 1ª implementação **só permitia detetar retas**, mais tarde permitiu outras formas geométricas parametrizáveis tais como **círculos**, elipses...

Optical flow

- Padrão de aparente movimento de objetos, superfícies e arestas numa cena visual causada pelo movimento relativo do observador em relação à cena
- Normalmente usa o método luca-kanade.

Funcionamento de uma camara RGBD (Kinect)

- Camara de alta qualidade, capaz de detetar movimentos em 3D aliado a sensores de movimentos
- Sensores:
 - Camara rgb que permite reconhecimento facial
 - Sensor de profundidade IV permite 3D
 - Microfone embutido
 - ...
- Combina uma camara RGB com informações de profundidade

Local features

- **Deteção:** identificar pontos de interesse
 - Laplacian of gaussian – blob detector
 - SIFT ou SURF - keypoints
- **Descrição:** extrair vetor de descrição de características à volta de cada ponto de interesse
- **Matching:** determinar correspondência entre a descrição nas duas visualizações