

Relatório de Leitura

Lucas Buligon Antunes

Data de finalização: 19/05/2023

Artigo base:

MARENGONI, Maurício; STRINGHINI, Stringhini. Tutorial: Introdução à Visão Computacional usando OpenCV. Revista de Informática Teórica e Aplicada, v. 16, n. 1, p. 125–160, 2010. Disponível em: <https://seer.ufrgs.br/rita/article/view/rita_v16_n1_p125>. Acesso em: 18 maio 2023.

1. Introdução

Por muitas vezes, termos na área de computação acabam sendo confundidos, principalmente devido a sua similaridade. Assim, os termos processamento de imagens e visão computacional se enquadram nesse caso. O processamento de imagens é um processo onde a entrada do sistema é uma imagem e a saída é um conjunto de valores numéricos, que podem ou não compor uma outra imagem. Por sua vez, a visão computacional procura emular a visão humana, portanto também possui como entrada uma imagem, porém, a saída é uma interpretação da imagem como um todo, ou parcialmente.

Os processos de visão computacional geralmente iniciam com o processamento de imagens

De modo que o espectro que vai do processamento de imagens até a visão computacional pode ser dividido em três níveis:

- **Baixo-nível:** operações primitivas, tais como a redução de ruído ou melhoria no contraste de uma imagem;
- **Nível-médio:** operações do tipo segmentação (particionamento da imagem em regiões) ou classificação (reconhecimento dos objetos na imagem);
- **Alto-nível:** processos relacionados com as tarefas de cognição associadas com a visão humana.

Tratando-se sobre visão computacional, uma operação muito comum de ser observada é a aplicação de um operador que extrai a placa do veículo e identifica as letras e números da placa, possibilitando que os dados do veículo sejam encontrados em um banco de dados. Para realizar esse tipo de operação se faz necessário o uso de algumas ferramentas, como o OpenCV, por exemplo.

1.1. OpenCV

O OpenCV (*Open Source Computer Vision*) é uma biblioteca de programação, de código aberto, desenvolvida inicialmente pela Intel Corporation. Ele implementa uma variedade de ferramentas de interpretação de imagens. Ao ser executado, invoca automaticamente uma DLL (*Dynamic Linked Library*) que detecta o tipo de processador e carrega, por sua vez, a DLL otimizada para este.

A biblioteca está dividida em cinco grupos de funções: Processamento de imagens; Análise estrutural; Análise de movimento e rastreamento de objetos; Reconhecimento de padrões e Calibração de câmera e reconstrução 3D.

2. Desenvolvimento

Os processos de visão computacional, muitas vezes, necessitam de uma etapa de pré-processamento envolvendo o processamento de imagens. As imagens das quais serão extraídas alguma informação, em alguns casos, precisam ser convertidas para um determinado formato ou tamanho e, além disso, precisam ainda ser filtradas para remover ruídos provenientes de seu processo de aquisição. É importante ressaltar que ruído não é apenas interferência no sinal de captura da imagem, mas também interferências que possam atrapalhar a interpretação ou o reconhecimento de objetos na imagem.

2.1. Processamento de Imagens

Para a remoção dos ruído são utilizados os filtros, os quais podem ser classificados de duas maneiras: espaciais¹, filtros que atuam diretamente na imagem; ou frequência, onde a imagem é inicialmente transformada para o domínio de frequência usando da transformada de Fourier (geralmente através da transformada de Fourier discreta) e então é filtrada neste domínio, em sequência sendo transformada de volta para o domínio de espaço.

2.1.1. Domínio espacial

Os processos no domínio espacial são definidos pela equação 01:

$$g(x, y) = T(f(x, y)) \quad (01)$$

Onde: $f(x,y)$ é a imagem original, $T(.)$ é uma transformação na imagem e $g(x,y)$ é a imagem transformada. T é uma operação definida sobre uma vizinhança de influência do pixel que está localizado na posição x, y . A ideia de vizinhança de influência considera os pixels ao redor da posição x, y . Esta vizinhança é definida por uma região quadrada (ou retangular) e de tamanho (lado) ímpar.

Estas regiões, que definem matrizes nas operações de transformação, também são chamadas de máscaras.

No caso mais simples, o operador T é computado em uma vizinhança de tamanho 1×1 , ou seja, apenas o valor do pixel no ponto é suficiente para determinar o valor na imagem processada. Essa operação é chamada de função de transformação de intensidade², sendo utilizada para alterar a intensidade da imagem. De modo que é amplamente utilizada, por exemplo, na binarização de uma imagem.

2.1.1.1. Histogramas

Os histogramas são ferramentas de processamento de imagens que possuem grande aplicação prática. Os histogramas são determinados a partir de valores de

¹ O termo domínio espacial se refere à imagem em si, e métodos que atuam no domínio espacial estão baseados na manipulação direta dos pixels da imagem.

² Quantidade de brilho ou cor em um pixel de uma imagem. É um valor numérico que representa a quantidade de luz emitida ou refletida por um determinado ponto na imagem.

intensidade dos pixels. Entre as principais aplicações dos histogramas estão a melhora da definição de uma imagem, a compressão de imagens, a segmentação de imagens ou ainda a descrição de uma imagem.

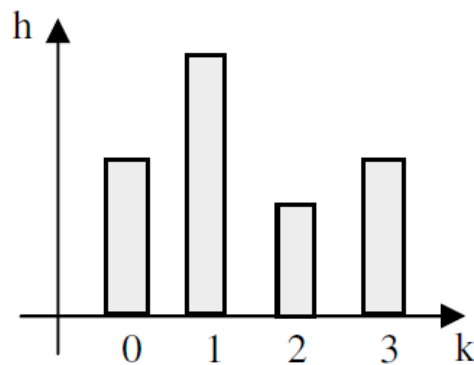


Figura 01 - Representação gráfica de um histograma

Uma operação bastante comum utilizando histogramas é o ajuste dos valores de intensidade de forma a melhorar o contraste em uma imagem. Esta operação é chamada de equalização de histogramas.

2.1.1.2. Filtros Estáticos

Os filtros estatísticos são os filtros espaciais mais comuns, dentre eles é possível citar os filtros de :

- Média: também chamado de passa-baixa, tem efeito de suavização da imagem e minimização dos ruídos, porém o efeito acaba sendo de embasamento ou borramento da imagem;
- Mediana: tende a reduzir o efeito de ruído de pulso, do tipo *salt and pepper*³;
- Máximo: tende a clarear a imagem;
- Mínimo: sendo o contrário do filtro de máximo, tende a escurecer a imagem;
- Moda: tende a homogeneizar os valores na imagem.

2.1.1.3. Filtros Gaussianos

Os filtros Gaussianos são filtros de média e são utilizados para suavizar a imagem de forma ponderada e simétrica. Um exemplo disso é o filtro do tipo passa-alta, utilizado para realçar bordas ou regiões de interesse abruptas de intensidade, entretanto também realçando ruídos da imagem.

2.1.2. Domínio de Frequência

É possível fazer uma troca de base em uma imagem e representá-la em termos de uma soma ponderada infinita de um conjunto de senóides. Esta representação mostra que mudanças rápidas na imagem são representadas por frequências altas e, por outro lado, mudanças suaves são representadas por frequências baixas. Esta mudança de base pode ser

³ Esse efeito apresenta-se por meio de alguns pixels brancos (*salt*) e pretos (*pepper*) na imagem.

feita utilizando a transformada de Fourier. As expressões que computam a transformada de Fourier no modo contínuo e no modo discreto⁴ são apresentadas nas equações 02 e 03:

Contínuo:

$$F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-j2\pi(ux+vy)} \quad (02)$$

Discreto:

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} f(x, y) e^{-j2\pi(ux/M+vy/N)} \quad (03)$$

Definir um filtro no domínio de frequência corresponde a encontrar uma máscara para ser utilizada em conjunto com a imagem transformada e assim obter a imagem filtrada desejada.

2.2. Visão Computacional

2.2.1. Segmentação

O processo de segmentação consiste em particionar uma imagem em regiões, ou objetos distintos. Este processo é geralmente guiado por características do objeto ou região, como, por exemplo, cor ou proximidade. Assim, o nível de detalhamento em um processo de segmentação depende da tarefa a ser executada e da resolução da imagem que se tem.

Existem diversas técnicas que podem ser utilizadas para a segmentação, dentre elas valem ser citadas as segmentações por: detecção de borda, corte e crescimento de região.

2.2.1.1. Segmentação por Detecção de Borda

Uma borda em uma imagem é caracterizada por uma mudança, normalmente abrupta, no nível de intensidade dos pixels. Os detectores de borda são definidos para encontrar este tipo de variação nos pixels e quando estes pixels estão próximos eles podem ser conectados formando uma borda ou um contorno e assim definindo uma região ou objeto.

Variações nos níveis de intensidade dos pixels podem ser determinadas pelas derivadas primeira e/ou derivada segunda. Alguns métodos de determinação de borda utilizam estas técnicas. O processo consiste em definir máscaras que caracterizam estas variações e em seguida fazer a convolução da imagem pela máscara.

Os operadores mais comuns baseados na derivada primeira (operadores de gradiente) de uma imagem são os operadores de Prewitt e de Sobel. É possível ainda definir operadores que atuam utilizando a derivada segunda dos pixels (operadores Laplacianos), estes operadores são baseados no fato que quando existe uma borda a derivada segunda atravessa o eixo das abscissas (este efeito é conhecido como zero-crossing).

⁴ A expressão discreta é a utilizada na DFT (*Discrete Fourier Transform*) que geralmente é utilizada na implementação em computador da transformada de Fourier.

2.2.1.2. Segmentação por Corte

A segmentação de imagens por corte é simples de ser implementada, é rápida em termos computacionais e utiliza de propriedades intuitivas para criar a imagem segmentada. Ela particiona uma imagem diretamente em regiões baseado simplesmente nos valores de intensidade e/ou propriedades destes valores.

A idéia central deste tipo de operador é a de verificar no histograma da imagem quantas regiões existem (picos e vales) e segmentar a imagem baseado nesta informação. Utilizando esta técnica é possível definir mais de uma região e criar uma imagem segmentada que não seja simplesmente uma imagem binária. É possível ainda fazer tratamentos no histograma para minimizar efeitos de ruídos que apareçam na imagem.

2.2.1.3. Segmentação Baseada em Crescimento de Região

Esta técnica de segmentação encontra regiões diretamente na imagem agrupando pixels ou sub-regiões em regiões maiores baseado em critérios de crescimento pré-definidos. O procedimento parte de um conjunto de pontos, chamados de sementes, e, a partir destes pontos vai agrupando pontos utilizando uma vizinhança de influência, formando as regiões. Nesta vizinhança são analisadas propriedades e são medidas similaridades para determinar se o pixel faz parte ou não da região sendo considerada. As propriedades normalmente consideradas são: cor, intensidade de nível de cinza, textura, momentos, entre outras.

Detalhes importantes desta técnica são a definição das sementes e a definição de um critério de parada para o crescimento de regiões.

2.2.2. Reconhecimento de Padrões

Reconhecer significa conhecer de novo, processo onde existe algum conhecimento prévio e algum tipo de armazenamento do conhecimento sobre o objeto a ser reconhecido.

Para fazer o reconhecimento um sistema de visão necessita uma base de conhecimento dos objetos a serem reconhecidos, esta base de conhecimento pode ser implementada diretamente no código ou pode ser aprendida a partir de um conjunto de amostras dos objetos a serem reconhecidos utilizando técnicas de aprendizado de máquina.

O reconhecimento de objetos é uma das principais funções da área de visão computacional e está relacionado diretamente com o reconhecimento de padrões.

Um objeto pode ser definido por mais de um padrão (textura, forma, cor, dimensões, etc) e o reconhecimento individual de cada um destes padrões pode facilitar o reconhecimento do objeto como um todo. As técnicas de reconhecimento de padrões podem ser divididas em dois grandes grupos: estruturais, onde os padrões são descritos de forma simbólica e a estrutura é a forma como estes padrões se relacionam; o outro grupo é baseado em técnicas que utilizam teoria de decisão, neste grupo os padrões são descritos por propriedades quantitativas e deve-se decidir se o objeto possui ou não estas propriedades.

Os processos de reconhecimento de padrões podem ainda ser uma mistura das técnicas utilizadas nestes dois grupos, por exemplo, no processo de reconhecimento de faces.

2.2.3. Rastreamento

O processo de rastreamento é um processo de reconhecer um padrão em uma sequência de imagens. O rastreamento poderia ser feito desta forma, porém, a busca em cada imagem de uma sequência sem o uso de qualquer conhecimento específico é relativamente lenta. Os processos de rastreamento atrelam um conhecimento sobre o movimento do objeto que está sendo rastreado para minimizar a busca entre as imagens em uma sequência.

Os processos de rastreamento podem ser aplicados em diversas áreas, indo de sistemas de segurança/vigilância até o uso em sistemas de interface humano-computador. No openCV as técnicas de rastreamento incluem dois componentes principais: identificação de objetos e modelagem da trajetória. Existem algumas funções que são utilizadas para o rastreamento, baseadas nos algoritmos de “*meanshift*” e “*camshift*”.

Entre as técnicas utilizadas, vale citar:

- *Corner Finding*: pressupõe a busca por um mesmo objeto de interesse em sequências de frames em um *stream* de vídeo. A idéia básica é buscar pontos diferenciados em uma imagem, passíveis de serem novamente encontrados em frames subsequentes;
- *Subpixel Corners*: diferente da técnica anterior, utiliza técnicas de localização de subpixels para obter com maior precisão a localização de detalhes de uma imagem;
- *Optical Flow*: possibilita a identificação de movimento entre sequências de frames sem que se conheça a priori o conteúdo destes. O OpenCV possui funções que implementam técnicas de detecção de movimento esparsas e densas:
 - Algoritmos de natureza esparsa consideram algum conhecimento prévio sobre os pontos que se deseja rastrear, como por exemplo os *corners* descritos nas seções anteriores;
 - Algoritmos densos associam um vetor de velocidade ou de deslocamento a cada pixel na imagem, sendo, portanto desnecessário o conhecimento prévio de pontos específicos da imagem.

Para a maioria das aplicações práticas, entretanto, as técnicas densas possuem um custo de processamento muito alto, sendo preferíveis, portanto, as técnicas esparsas;

- *Mean-Shift e Camshift*: *camshift* (*Continuously Adaptive Mean-SHIFT*) é um algoritmo desenvolvido para o rastreamento de cor, possibilitando também o rastreamento de faces. É baseado numa técnica estatística onde se busca o pico entre distribuições de probabilidade em gradientes de densidade. Esta técnica é chamada de “média por deslocamento” (*mean shift*) e foi adaptada no *Camshift* para tratar a mudança dinâmica das distribuições de probabilidade das cores numa sequência de vídeo;
- *Estimadores*: algoritmos deste tipo são capazes de estimar, por exemplo, a localização de uma pessoa ou objeto em movimento numa sequência de vídeo. Esta tarefa é dividida, basicamente, em duas fases: predição, baseada num

conhecimento prévio de dados da imagem e correção, que usa novas medidas para apurar a predição realizada anteriormente.

3. Glossário

CONVOLUÇÃO - Multiplicação de duas expressões no domínio de frequência.

MÁSCARA - Imagem binária que é usada para selecionar partes específicas de uma imagem original ou para realizar operações seletivas em determinadas áreas da imagem.

INTENSIDADE - Quantidade de brilho ou cor em um pixel de uma imagem. É um valor numérico que representa a quantidade de luz emitida ou refletida por um determinado ponto na imagem.

PROCESSAMENTO DE IMAGENS - É a área que lida com o desenvolvimento e aplicação de algoritmos e técnicas para analisar, melhorar e extrair informações de imagens digitais.

VISÃO COMPUTACIONAL - Refere à capacidade de um sistema computacional processar informações visuais. Ele combina técnicas de processamento de imagens, inteligência artificial e aprendizado de máquina para analisar e compreender o conteúdo visual em imagens ou vídeos. Isso inclui a captura de dados visuais, o pré-processamento desses dados, a extração de características relevantes e a interpretação dessas características para tomar decisões ou realizar tarefas específicas.