

Introdução

Lista de exercícios referente ao primeiro projeto da disciplina de Visão Computacional do programa de pós-graduação do Centro de Informática. A lista é composta por quatro questões, as quais foram resolvidas usando MATLAB R2012a.

Códigos fonte estão disponíveis online, juntamente com as imagens citadas nas resoluções abaixo e os enunciados das questões da referida lista no repositório do github: github.com/pauloserrafh/visaocomputacionalCIn, dentro da pasta relativa ao primeiro projeto, separadas em sub-pastas para cada questão. As imagens relativas à cada exercício estão numeradas de forma crescente, sendo a 1 sempre a imagem original e mostram a sucessão de resultados após cada etapa descrita na resolução da questão abaixo.

Resoluções

1) Identificar parafusos e porcas

A imagem de referência é utilizada para determinar as características de um parafuso e de uma porca. Para separar o background dos objetos, é feita uma binarização da imagem, ou seja, todos os pixels que tem um valor maior que o threshold definido são transformados em branco, caso contrário o pixel é preto. De uma maneira geral, operações em imagens consideram que os elementos que se deseja identificar, ou ativos, em uma imagem estão em branco enquanto o fundo é preto. Nas imagens originais os objetos estão em tons mais escuros do que o fundo, portanto ambas as imagens são binarizadas e têm as cores invertidas.

Da imagem de referência, os elementos conectados são extraídos. Elementos conectados são conjuntos contínuos (em qualquer direção) de pixels ativos. Como existem apenas dois objetos na imagem, sabemos que o componente com maior tamanho será a referência para parafusos e o com o menor será a referência para porcas.

De forma similar, são extraídos os elementos conectados da imagem a ser avaliada. Cada elemento é classificado como parafuso ou porca baseado na diferença absoluta entre o tamanho do elemento e as referências de tamanhos de parafuso e porca. São encontrados 3 parafusos e 4 porcas.

2) Extrair objetos em primeiro plano

A imagem original é convertida para tons de cinza e em seguida é aplicado um filtro passa baixa gaussiano. Filtros passa-baixa (FPB), como o nome sugere, atenuam as altas frequências, as quais, em imagens, caracterizam bordas e detalhes finos. Portanto, o resultado do filtro gaussiano é um embaçamento da imagem, aumentando ainda mais o efeito de “desfoque” no background.

Em seguida, os contornos restantes na imagem suavizada são encontrados. Usando um raciocínio análogo ao de filtros passa-baixa, filtros passa-alta acentuam as altas frequências, ou seja, bordas e detalhes finos. O filtro utilizado foi o Laplaciano de Gaussiana. Como a imagem já havia passado por um FPB no passo anterior, apenas alguns ruídos são identificados pelo FPA no plano de fundo da imagem.

Do passo anterior, resultam os contornos dos componentes em primeiro plano, porém bastante desconexos. Para preencher as falhas e tornar os elementos adjacentes conectados, é realizada uma operação morfológica de fechamento, que é o resultado da aplicação em cascata de uma operação de dilatação seguida de erosão. Uma dilatação atua na imagem de forma que o pixel de saída ficará ativo se qualquer pixel na vizinhança do pixel de entrada estiver ativo; caso contrário, o pixel de saída ficará inativo. A operação de erosão por sua vez resulta em pixels de saída ativos se todos os pixels da vizinhança do pixel de entrada estiverem ativos; caso contrário, o pixel de saída ficará inativo. A região de vizinhança, chamada elemento estruturante, pode ter qualquer forma, desde que seja a mesma tanto para a dilatação quanto para a erosão.

A região de interesse na imagem original é o maior elemento conectado da imagem atual após o fechamento. Alguns elementos ruidosos nas bordas da imagem ainda aparecem conectados à região de interesse, então a imagem é tratada para que as bordas sejam removidas. O resultado após todos os passos descritos anteriormente é uma imagem cujo único elemento ativo é o componente de primeiro plano da imagem original, porém com algumas falhas, o que é resolvido com uma operação de preenchimento do elemento.

O preenchimento atua removendo os conjuntos de elementos inativos que estão dentro de componentes conectados. Por fim, o resultado é utilizado como uma máscara para a imagem original. Tomando como base a imagem resultante da aplicação de todas as operações descritas anteriormente, a imagem original é modificada usando uma comparação simples: Os valores da imagem original são mantidos em pixels que estão ativos na imagem tratada; caso contrário, o pixel da imagem original fica preto.

3) Extrair contorno do elemento de interesse

A imagem é convertida de rgb para hsv. O espaço de cores definidos pelo sistema hsv é composto por três parâmetros: matiz (hue), saturação (saturation) e valor (value). A matiz indica o tipo de cor, saturação define o percentual de cinza, quanto menor o valor, mais acinzentada a cor, e valor indica o brilho da cor. Portanto, a segmentação desejada pode ser obtida trabalhando apenas com a matiz.

Sabendo que a maior parte da imagem é composta pelo plano de fundo e que ele tem aproximadamente a mesma cor, é possível avaliar o histograma da matriz para identificar o valor que mais ocorre e remover elementos dessa tonalidade da imagem. Como existem pequenas diferenças entre valores de pixels que são visualmente semelhantes, não são removidos apenas os pixels que tem um valor igual ao máximo indicado no histograma e sim valores que estão dentro de um intervalo que contém este valor.

Após a remoção do background, a imagem é binarizada. Usando um raciocínio similar ao aplicado à questão 2, é realizada uma operação de fechamento com o intuito de

remover os pequenos ruídos existentes após a binarização, e em seguida o maior elemento conectado é extraído. Por fim, é aplicado um FPA Laplaciano de Gaussiana para extrair as bordas da imagem tratada.

Dentre as 3 imagens, o melhor resultado visual é obtido com a imagem dos cavalos, pois há uma diferença grande na cor entre background e elementos desejados e todo o fundo é composto por cores próximas (verde e amarelo). Na imagem do canguru, não foi possível separar a “mancha” de terra próxima à pata dianteira do canguru devido à cor muito próxima e diferente do background e alguns elementos que já eram escuros naturalmente, como olhos, focinho e parte interna da orelha, foram perdidos após a binarização. Na imagem da águia, existe um meio termo entre ambas: o background é bastante diferente do elemento desejado, porém alguns pontos mais escuros naturalmente na imagem foram perdidos.

4) Contar número de andares

A imagem é binarizada de forma que a imagem resultante contenha apenas o prédio e pequenos ruídos. Em seguida, é aplicada uma operação morfológica de erosão, deixando visíveis apenas as janelas do prédio. Usando um elemento estruturante linear é aplicada uma dilatação, o que conecta todas as janelas de um mesmo andar. Vale notar que elementos estruturantes diferentes foram utilizados nas operações de erosão e dilatação, portanto não fica caracterizada uma operação de abertura.

Como todas as janelas do andar estão conectadas e separadas por andar, é correto dizer que cada uma das “linhas” formadas pela união de janelas corresponde a um andar. Portanto, o número de andares é igual ao número de elementos conectados na imagem final. São encontrados 5 andares.