



**UNIVERSIDADE FEDERAL
DO PARANÁ
SETOR DE TECNOLOGIA
DEPARTAMENTO DE
ENGENHARIA ELÉTRICA**

SEGMENTAÇÃO DE IMAGENS E ALGORITMO K-MEANS

Trabalho de graduação apresentado à disciplina
Processamento Digital de Sinais do curso de
Engenharia Elétrica do Setor de Tecnologia da
Universidade Federal do Paraná.

Orientador: Prof. Marcelo Rosa
Aluno: George Kolossoski

CURITIBA
2007
Introdução

A visão é um dos mais poderosos e complicados sentidos que o ser humano possui. Através da visão ele obtém as posições e propriedades dos objetos, assim como suas relações entre os mesmos e o ambiente que os cerca. A visão permite realizar três tarefas básicas, conforme [3]:

- Percepção do mundo;
- Concepção de uma estratégia para tomada de decisão;
- Execução de uma ação.

Enquanto esta visão é muito simples para o ser humano, foi provado por muitos especialistas que é muito complexo ensinar um computador a funcionar com um sistema de visão (mesmo rudimentar). Uma causa disto é que até hoje é difícil explicar precisamente o processo da percepção. Não se podem recuperar todas as informações de uma cena apenas pela sua intensidade. Esta intensidade é, na verdade, resultado de uma combinação de fatores como a superfície de um objeto, fonte e direção de iluminação, luz ambiente, condição atmosférica, entre outros. O olho humano é sempre comparado a uma câmera e, em muitos casos, as similaridades são bem evidentes. Há, entretanto, uma grande diferença: o processo de visualização. Na câmera a filme, as imagens são produzidas no filme através das mudanças fotoquímicas, enquanto que no olho humano as mudanças fotoquímicas provocam impulsos nervosos que são transmitidos ao cérebro. Assim, o cérebro interpreta estes impulsos como provenientes de objetos situados fora do corpo; sendo impossível perceber as imagens como se estivessem na retina. Até mesmo as imagens geradas pelo efeito pós-imagem são projetadas para fora do corpo (ainda que de olhos fechados).

Apesar da visão humana ser bem descrita num nível neuroanatômico, o processamento da informação realizado pela retina e pelo córtex visual do cérebro permanece, de acordo com [4], ainda hoje obscuro. O conhecimento acerca de visão biológica ainda é muito limitado, desconexo e especulativo. Esta hipótese tem motivado pesquisadores em visão computacional a propor teorias computacionais sobre o que seria o processo de visão. Tais teorias têm evoluído ao longo dos anos, baseadas na crescente compreensão deste processo. Entre os adventos produzidos nestas últimas décadas está o processamento digital de imagens.

O processamento digital de imagens conhecido também como processamento de imagens, surgiu da teoria de processamento de sinais e se tornou uma das múltiplas facetas da teoria da informação. Imagens são como a representação ou descrição de um objeto, pessoa ou cena, trazendo informações através de distribuições de intensidade de luz.

O interesse em métodos de processamento de imagens digitais decorre de duas áreas principais de aplicação:

- Melhoria da informação visual para a interpretação humana.
- Processamento de dados de cenas para percepção automática através de máquinas.

O processamento digital de imagens é uma área de concentração do conhecimento humano, que tem como objetivos a manipulação e análise de imagens por computador visando a extração de informação destas. Os recursos disponibilizados pelo processamento de imagens são utilizados em várias atividades, entre as quais estão a medicina, a robótica e a meteorologia.

O interesse em métodos de Processamento Digital de Imagens surgiu, principalmente, da necessidade de melhorar a qualidade da informação pictorial para interpretação humana. Uma das primeiras aplicações das técnicas de PDI foi a melhoria de ilustrações de jornais enviados por cabo submarino entre Londres e New York por volta de 1920. Um dos processos mais importantes de detecção de partes constituintes de uma imagem é chamado de Segmentação, e utiliza conceitos de casamento de padrões e técnicas de correlação ou convolução para detectar bordas,

cantos, buracos e texturas. O processo de segmentação tenta distinguir "backgrounds" e objetos através da detecção de regiões "homogêneas" da cena ou imagem observada, gerando dados simbólicos provenientes de dados icônicos.

Processamento de imagens

Conforme encontrado em [2], podemos dividir o processamento de imagens em algumas etapas. A primeira delas é a aquisição da imagem digital. Para isso, é necessário um sensor para imageamento, que pode ser uma câmera de TV monocromática ou colorida, ou também uma câmera de varredura por linha, que produz uma única linha de imagem por vez, por exemplo. Tão importante quanto o sensor é a capacidade de digitalizar o sinal produzido pelo mesmo.

Depois de obtida a imagem digital, o próximo passo é o *pré-processamento* da imagem. A função chave no pré-processamento é a melhoria da imagem de forma que as chances de sucesso dos processos seguintes sejam maiores. Por exemplo, o pré-processamento pode envolver técnicas para o realce de contrastes, remoção de ruído e isolamento de regiões cuja textura indique a probabilidade de informação alfanumérica.

A próxima etapa é a segmentação, que divide uma imagem de entrada em várias partes ou objetos constituintes. Em geral, a segmentação automática é uma das tarefas mais difíceis no processamento de imagens digitais. Este é o estágio do Processamento de Imagens que será abordado neste documento.

O estágio anterior produz uma saída constituída tipicamente por dados em forma de *pixels*, que correspondem tanto à fronteira de uma região como a todos os pontos dentro da mesma. É necessário converter esses dados para uma forma adequada ao processamento computacional. A primeira decisão que precisa ser feita é se os dados devem ser representados como fronteiras ou como regiões completas, e também deve ser especificado um método para descrever os dados, de forma que as características de interesse sejam enfatizadas.

O próximo processo, de descrição, também chamado seleção de características, procura extrair características básicas para discriminação entre classes de objetos, ou que resultem em alguma informação quantitativa de interesse. Quando se trata de reconhecimento de caracteres, descritores tais como buracos e concavidades são características poderosas que auxiliam na diferenciação entre uma parte do alfabeto e outra.

O último estágio envolve reconhecimento e interpretação. Reconhecimento é o processo que atribui um rótulo a um objeto, baseado na informação fornecida pelo seu descritor. A interpretação envolve a atribuição de significado a um conjunto de objetos reconhecidos.

A figura abaixo mostra esquematicamente os passos fundamentais no processamento de imagens:

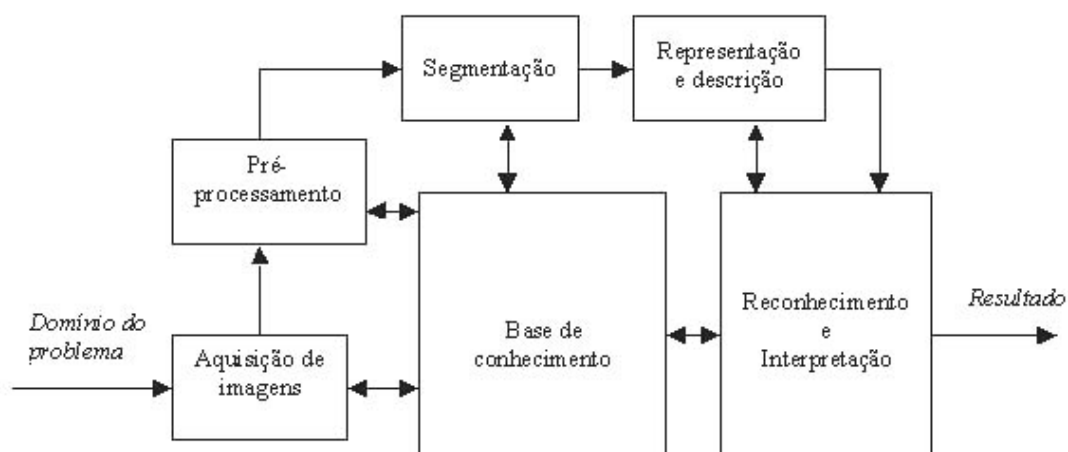


Figura 1: Esquema dos passos fundamentais no processamento de imagens.

O Problema da Segmentação de Imagens

Na área de detecção e reconhecimento de imagens não basta simplesmente representar uma imagem com diferentes cores ou graduações de cinza. Também é necessário identificar regiões e estabelecer subdivisões na imagem em sua unidade básica (pixel) para que possa ser interpretada de acordo com uma finalidade específica. A identificação de regiões ou segmentos (não sobrepostos) na imagem é chamada de segmentação.

A segmentação de imagens traz como resultado um conjunto de regiões/objetos ou contornos extraídos da imagem. Assim, os pixels em uma determinada região são similares em alguma característica ou propriedade computacional, tais como cor, intensidade, textura ou continuidade. Em relação às mesmas características, regiões adjacentes devem possuir diferenças significativas entre umas e outras.

Há uma certa dificuldade no processo de segmentação de imagens nos computadores. A identificação dos segmentos deve obedecer a algumas características, encontradas em [9]. Os pixels devem possuir alguma propriedade em comum dentro da imagem. Essa propriedade pode ser, por exemplo, uma superfície que representa um osso dentro de uma radiografia, uma peça sendo submetida a um controle de qualidade ou um mapa ilustrando alguma característica de uma foto. Dentre as propriedades desejáveis de uma imagem são destacadas algumas a seguir:

- Homogeneidade da região representada pelos pixels.
- Os segmentos são regiões fechadas e devem ser delimitadas por bordas ou outros segmentos.
- Associadas ao segmento existem certas propriedades matemáticas. As propriedades mais comuns utilizadas para representar as regiões são os diagramas de Venn, sendo P_n os pixels representativos de uma região R_n , contendo N pixels.
- Cada pixel deve pertencer somente a uma região, não havendo regiões adjacentes com pixels em comum.
- Os segmentos, com relação a níveis de cinza e textura, devem ser uniformes e homogêneos.
- As regiões devem ser simples e não devem apresentar buracos pequenos.
- Regiões adjacentes devem possuir diferenças significativas.
- Os segmentos devem ter bordas precisas.

As características mencionadas acima, na prática, são utópicas por que superfícies homogêneas são geralmente cheias de furos e as bordas em geral são irregulares. Além disso, podem ocorrer a fusão e perda de bordas nas regiões adjacentes. Como regra geral a identificação de segmentos é específica e típica para cada aplicação. Geralmente destaca-se a separação da região de interesse buscando-se descontinuidade e similaridade nos diferentes tons da imagem. As descontinuidades são representadas pelas mudanças bruscas nos tons das cores como linhas e bordas. As similaridades baseiam-se nos limiares dos tons, subdivisão da imagem em regiões homogêneas e crescimento de regiões.

Dificuldades inerentes

Existem algumas dificuldades inerentes à segmentação de imagens. Antes que um processo de segmentação seja executado, alguns fatores devem ser considerados. Dentre eles podemos destacar os seguintes:

- Segmentação de forma autônoma em larga escala: Quando a segmentação de imagens envolve processos automáticos, há uma grande necessidade de controlar o ambiente de onde a imagem é retirada. Ambientes bem controlados (grandes contrastes) tendem a facilitar a interpretação das imagens. Ambientes externos, dependentes do clima, iluminação e outros fatores apresentam várias dificuldades.

- Controle da luminosidade: Conforme a aplicação envolvida, a existência de sombras tende a dar uma falsa impressão acerca do tamanho real da região a ser segmentada.
- As bordas das regiões são muitas vezes irregulares e imprecisas.
- A precisão do resultado depende da qualidade da distinção entre os diferentes elementos da imagem.
- Escolha da melhor estratégia e adequação à aplicação que se deseja.

Vários algoritmos e técnicas de segmentação de imagens foram desenvolvidos, não havendo, porém, uma solução geral para este problema. Muitas vezes para a resolução de um problema de segmentação é necessária a combinação das técnicas, para que ocorra a adaptação ao domínio do problema.

O Algoritmo k-means

Um dos algoritmos desenvolvidos para resolver o problema da segmentação de imagens é o K-Means, também chamado de K-Médias. A idéia do algoritmo é fornecer uma classificação de informações de acordo com os próprios dados, baseada em análises e comparações entre os seus valores numéricos. Assim, o algoritmo fornecerá uma classificação automática sem a necessidade de supervisão humana, ou seja, sem pré-classificação existente. Por causa desta característica, o K-Means é considerado como um algoritmo de mineração de dados (data mining) não supervisionado.

É possível usar o exemplo de uma tabela com linhas e colunas que contêm os dados a serem classificados, para melhor entendimento de como o algoritmo funciona. Nesta tabela, cada coluna é chamada de dimensão e cada linha contém informações para as dimensões, que também são chamadas de ocorrências ou pontos. Geralmente, trabalha-se com dados contínuos neste algoritmo, mas nada impede que dados discretos sejam utilizados, desde que eles sejam mapeados para valores numéricos correspondentes. Como foi dito, o algoritmo analisa todos os dados desta tabela e cria classificações, isto é, o algoritmo indica uma classe (cluster) e diz quais linhas pertencem a esta classe. A quantidade de classes desejadas deve ser fornecida pelo usuário. Este número de classes que deve ser passado para o algoritmo é chamado de k e é dele que vem a primeira letra do nome do algoritmo: K-Means.

O algoritmo compara cada valor de cada linha por meio da distância e gera as classes, além de classificar as ocorrências. Geralmente, para calcular a que distância uma ocorrência está da outra, é utilizada a distância euclidiana. A maneira de calcular esta distância depende da quantidade de atributos da tabela fornecida. Após o cálculo das distâncias o algoritmo calcula centróides para cada uma das classes. Conforme as iterações do algoritmo, o valor de cada centróide é refinado pela média dos valores de cada atributo de cada ocorrência pertencente a este centróide. Com isso, o algoritmo gera k centróides e coloca as ocorrências da tabela de acordo com sua distância dos centróides.

Para simplificar, a explicação de como o algoritmo funciona pode ser apresentada em cinco passos, como encontrado em [1]:

1) Os centróides recebem valores iniciais.

Neste passo os k centróides devem receber valores iniciais. No início do algoritmo geralmente escolhem-se os k primeiros pontos da tabela. Também é importante colocar todos os pontos em um centróide qualquer para que o algoritmo possa iniciar seu processamento.

2) Gera-se uma matriz de distância entre cada ponto e os centróides.

Neste passo, é calculada a distância entre cada ponto e os centróides. A parte que exige mais processamento para os cálculos ocorre neste passo, pois se existem N pontos e k centróides deve se calcular $N \times k$ distâncias neste passo.

3) Classificar cada ponto de acordo com as suas distâncias dos centróides.

Aqui, os pontos são classificados de acordo com sua distância dos centróides de cada classe. A

classificação funciona assim: o centróide que está mais perto deste ponto vai 'incorporá-lo', ou seja, o ponto vai pertencer à classe representada pelo centróide que está mais perto do ponto. É importante dizer que o algoritmo termina se nenhum ponto 'mudar' de classe, ou seja, se nenhum ponto for 'incorporado' a uma classe diferente da que ele estava antes deste passo.

4) Cálculo dos novos centróides para cada classe.

Nesta etapa, os valores das coordenadas dos centróides são refinados. Para cada classe que possui mais de um ponto o novo valor dos centróides é calculado fazendo-se a média de cada atributo de todos os pontos que pertencem a esta classe.

5) Repetir até a convergência.

O algoritmo retorna ao passo 2 repetindo iterativamente o refinamento do cálculo das coordenadas dos centróides, até que mais nenhum ponto mude de classe.

Nota-se que desta maneira se obtém uma classificação que coloca cada ponto em apenas uma classe, portanto pode se dizer que este algoritmo faz uma classificação *hard* (hard clustering). Outros algoritmos trabalham com o conceito de classificação *soft* onde existe uma métrica que diz o quão 'dentro' de cada classe o ponto está.

O algoritmo k-means é utilizado para segmentar imagens, baseado em seus atributos, em k pedaços. Ele assume que os atributos dos pontos da imagem formam um espaço vetorial. O objetivo do algoritmo é minimizar a variância dos atributos dos pontos que estão dentro de um determinado segmento. Matematicamente podemos dizer que o k-means minimiza a função erro quadrático:

$$V = \sum_{i=1}^k \sum_{x_j \in S_i} |x_j - \mu_i|^2$$

onde existem k segmentos S_i , com $i = 1, 2, \dots, k$, e μ_i é o centróide, ou centro de um conjunto de pontos x_j , pertencentes a S_i .

Este algoritmo se tornou extremamente popular porque ele converge muito rapidamente. Na prática, tem se observado que o número de iterações é tipicamente muito menor que o número de pontos. Contudo, pesquisadores mostraram que existem certos conjuntos de pontos nos quais o k-means leva um tempo acima do polinomial: $2^{\Omega(\sqrt{n})}$ para convergir.

Em relação à performance, não é garantido que o algoritmo retorne a solução mais otimizada. A qualidade da solução final depende muito do número inicial de segmentos, k, e pode, na prática, ser muito pior do que a solução otimizada. Uma das coisas que podem ser feitas para contornar esse problema é, levando-se em consideração que o algoritmo é rápido, executá-lo várias vezes e retornar a melhor solução encontrada.

Outro ponto negativo do k-means é a necessidade do usuário definir o número inicial de segmentos (k). Se a imagem não é naturalmente segmentada, pode se conseguir resultados estranhos. Além disso, o algoritmo funciona muito bem somente quando segmentos estão disponíveis na imagem original.

Demonstração do algoritmo

As figuras seguintes, retiradas de [10], demonstram a utilização do algoritmo k-means para uma imagem constituída de vários pontos vermelhos, distribuídos em conjuntos. Note que existem 5 conjuntos principais de pontos, mas serão utilizados apenas 4 centróides.

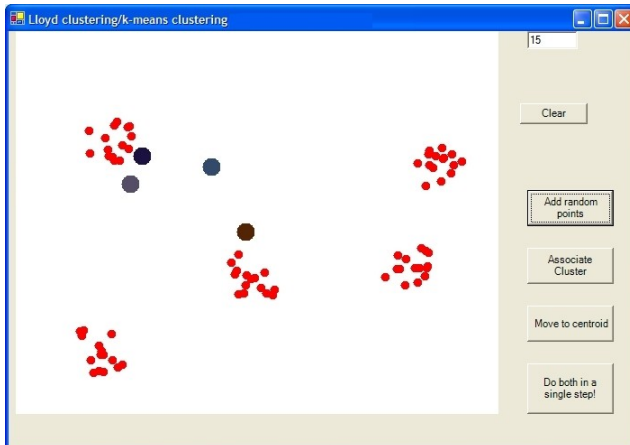


Figura 2: Primeiro os centros (pontos pretos) são escolhidos aleatoriamente.

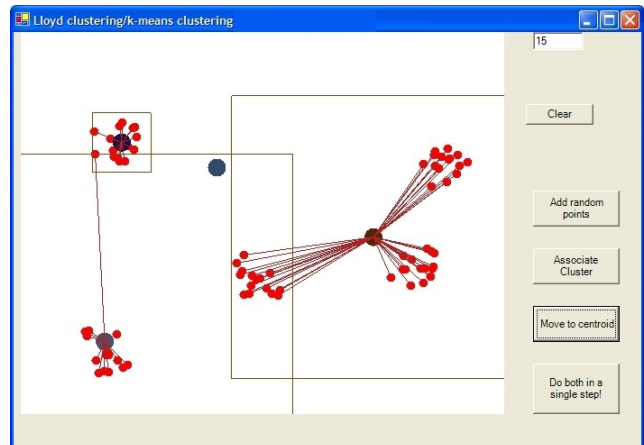


Figura 3: Os pontos vermelhos foram associados aos centros, e estes foram movidos para os respectivos centróides.

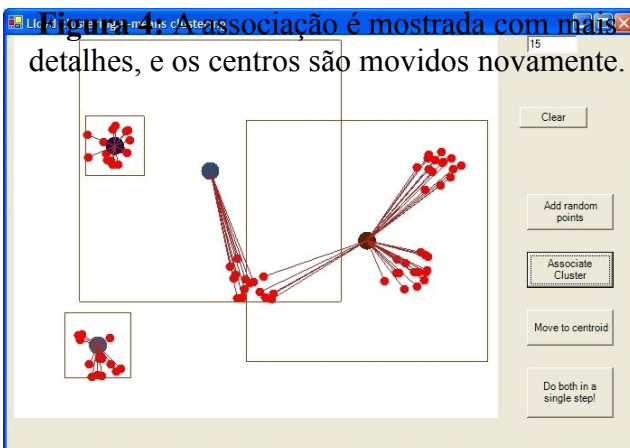


Figura 4: A associação é mostrada com mais detalhes, e os centros são movidos novamente.

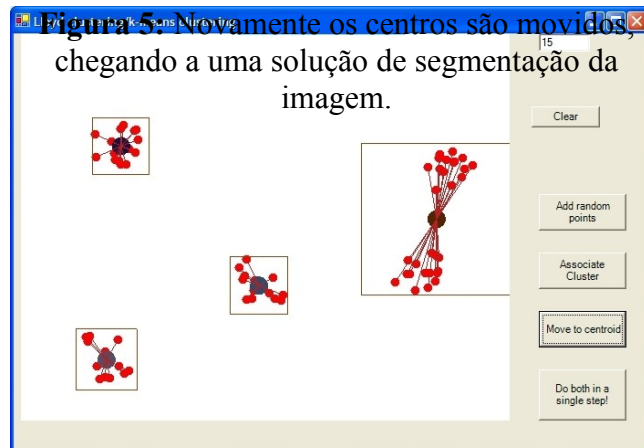


Figura 5: Novamente os centros são movidos, chegando a uma solução de segmentação da imagem.

Melhorias no algoritmo

Em 2006 foi proposta uma nova maneira de escolher os centros iniciais, baseada na idéia de seleccionar os centros de um modo que eles já estão posicionados perto de grandes quantidades de pontos. Esse método de posicionamento traz uma série de melhorias na taxa de erros do k-means. Apesar de levar mais tempo para escolher os centros iniciais, o algoritmo convergirá muito mais rápido, fazendo com que o tempo de processamento diminua. Os autores da nova proposta testaram o novo modo de posicionamento com dados reais e dados produzidos e obtiveram melhorias de 10 a 100 vezes em velocidade e taxa de erros. Os testes mostraram que o novo método é, no mínimo, tão bom quanto o método original em relação a ambos os atributos.

Conclusão

O processamento digital de imagens é uma ferramenta muito importante, pois traz muitas vantagens e melhorias para áreas como a robótica, medicina, fotografia, meteorologia, enfim, tudo o que envolve imagens. Tem como objetivos a manipulação e análise de imagens por computador visando a extração de informação destas, para que os resultados venham a trazer benefícios para as áreas citadas anteriormente, além de muitas outras aplicações.

Dentro do processamento de imagens, levando em consideração a maneira como as imagens são processadas digitalmente, encontra-se o processo de segmentação, que consiste em dividir a imagem em várias partes de acordo com as características dos pontos (pixels). Esse é um processo simples para o ser humano, que consegue segmentar uma cena, obtida através da visão, imediatamente, separar todos os objetos e definir seus contornos. Porém, é um processo complicado ao nível de computação. O algoritmo k-means foi desenvolvido para resolver esse problema, porém ele depende de certas variáveis que, se mal escolhidas, levam a resultados não otimizados. Apesar disso é um algoritmo bastante eficiente em termos de velocidade, e com a sua melhoria descrita no item anterior, poderá ser utilizado para mais aplicações com mais eficiência. Espera-se que mais melhorias sejam desenvolvidas na área de processamento de imagens, para que a visão computacional possa enfim chegar a simular com alto grau de perfeição o aparato visual humano.

Bibliografia

- [1]
http://www.imasters.com.br/artigo/4709/sql_server/data_mining_na_pratica_algoritmo_k-means/
- [2]
<http://atlas.ucpel.tche.br/~vbastos/pi.htm>
- [3]
<http://www.ic.uff.br/~aconci/deteccao.html>
- [4]
<http://www.ele.ita.br/~labvisao/>
- [5]
http://www.ufpel.tche.br/prg/sisbi/bibct/acervo/info/1999/mono_eurico_antunes.pdf
- [6]
<http://www.inf.ufsc.br/~visao/1998/seara/index.html>
- [7]
<http://www.wikipedia.org/>
- [8]
<http://www.inf.ufsc.br/~visao/cores.html>
- [9]
<http://www.inf.ufsc.br/~visao/segmentos.pdf>
- [10]
http://en.wikipedia.org/wiki/K-means_algorithm
- [11]
<http://people.revoledu.com/kardi/tutorial/kMean/index.html>