UNIDADE 6 - Reconhecimento e Detecção

Uma parte significativa dos sistemas de visão computacional possui como objetivo final ou intermediário o reconhecimento de objetos, seres vivos, cenas, defeitos, estruturas ou qualquer tipo de coisa que possa ser classificada de alguma forma. O alvo a ser reconhecido pode já estar isolado em um única imagem ou ocorrer juntamente com outras coisas e neste caso, além de reconhecer, temos o problema de detectar uma ou mais ocorrências do alvo na imagem. Nas próximas seções serão apresentados três grandes grupos de técnicas que podem ser utilizadas para resolver o problema de reconhecimento e detecção.

6.1 Casamento de modelos

A técnica de casamento de modelos é bastante utilizada em problemas de detecção e consiste basicamente em manter um modelo do objeto que se pretende detectar e percorrer a imagem ou algum conjunto de atributos extraídos da imagem em busca de regiões na imagem que se enquadrem ou que se aproximem do modelo do objeto de interesse. Para identificar se o modelo se enquadra a uma determinada região da imagem utiliza-se algum tipo de medida de distância ou de similaridade entre o modelo e cada região da imagem. As regiões em que essa medida é suficientemente alta (similaridade alta), indicando que a região e o modelo são similares, são reportadas para o usuário ou para algum outro módulo do sistema, como sendo regiões em que o objeto está presente. O valor acima do qual deve se reportar uma ocorrência do objeto é chamado de limiar e pode variar de problema para problema, tendo que ser definido experimentalmente ou através de conhecimento prévio sobre o tipo de imagem a ser tratada pelo sistema (é comum utilizar-se o inverso da similaridade de forma que o valor 0 represente a máxima similaridade ou a distância mínima entre duas imagens). Além do limiar, duas outras coisas que precisam ser definidas experimentalmente são o tipo de medida de similaridade e o tipo de modelo a ser adotado para representar o objeto alvo. Nos dois casos, existem muitas opções disponíveis. Como medida de similaridade podemos citar a distância Euclidiana, de Manhattan, de Mahalanobis, e da

Escavadeira (*Earth-Mover*). Já entre os modelos, o mais simples seria uma única imagem contendo apenas o objeto alvo. Esse modelo tem aplicação muito restrita, pois dependendo do objeto, simples rotações e mudanças de escala (tamanho) fariam com que a similaridade entre o modelo e a região em que o objeto se encontra (rotacionado ou com tamanho diferente) fosse muito baixa. Para fazer com que o sistema encontre o objeto mesmo que ele esteja rotacionado ou em tamanhos diferentes, uma solução seria guardar várias imagens do objetos, rotacionado em diferentes ângulos e em diferentes tamanhos ou então trabalhar com modelos mais sofisticados ou ainda com comparação da atributos extraídos da imagem, ao invés de trabalhar diretamente com os pixels. A capacidade de um sistema de visão computacional detectar objetos mesmo quando são rotacionados ou estão em diferentes escala é chamada de invariância à rotação e invariância à escala, respectivamente.

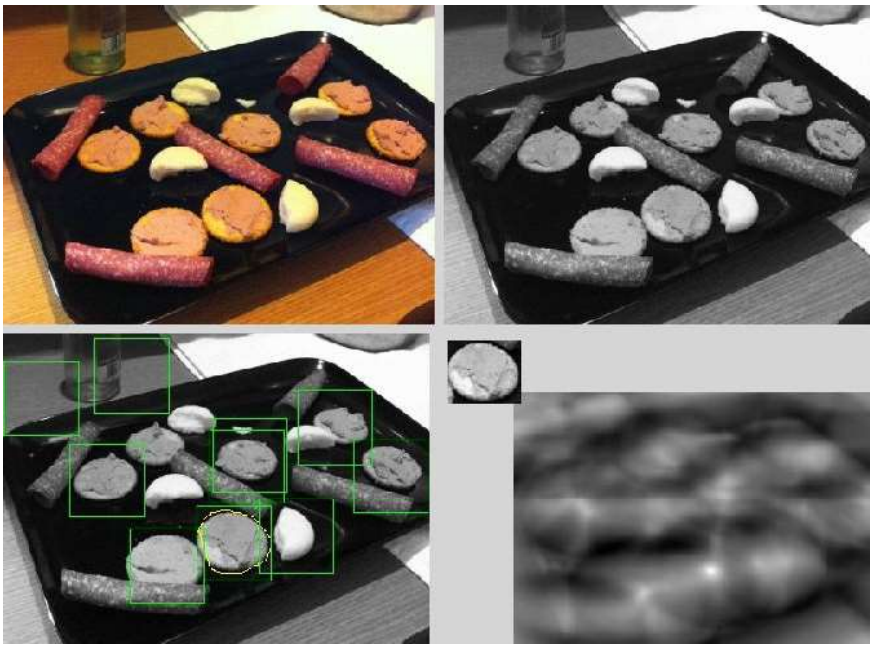

Figura 39: Aplicação de casamento de modelos para detecção de biscoitos

A Figura 39 mostra o funcionamento do tipo mais básico de casamento de modelos, quando o objeto é representado por um única imagem dele mesmo e a imagem original é percorrida completamente calculando-se a similaridade entre cada região (de tamanho igual ao modelo) e o modelo. Temos acima, à esquerda, a imagem original e a direita a imagem em tons de cinza, pois a técnica adotada,

disponível como um plugin para o ImageJ (*template matching*) não utiliza informação de cor. Trata-se de uma bandeja com biscoitos redondos cobertos por patê e rolinhos de salame. O problema em questão é detectar os biscoitos redondos cobertos por patê. Na parte de baixo, à direita, temos o modelo utilizado (imagem de um dos biscoitos isolados, quase no centro da Figura 39) e uma imagem que mostra através da intensidade de cada pixel o centro das regiões da imagem original em que similaridade é maior (quanto mais claro, maior a similaridade). À esquerda temos, apresentados dentro de quadrados verdes, os biscoitos detectados. Embora o biscoito utilizado como modelo e mais alguns outros tenham sido detectados, é possível notar que houveram alguns falsos positivos (regiões marcadas como biscoito mas que não tinham biscoito) e também um falso negativo (biscoito que não foi detectado). Utilizando um conjunto maior de modelos (ou outros tipos de modelos), outras medidas de similaridade, algum tipo de pré-processamento ou pós-processamento para eliminar falsos positivos, por exemplo, melhores resultados poderiam ser obtidos.

6.2 Aprendizagem automática

A aprendizagem automática pode ser organizada em 3 grandes grupos de técnicas. No primeiro grupo temos a aprendizagem supervisionada, quando o sistema tem acesso a amostras ou exemplos daquilo que ele precisa aprender. No segundo, chamado de aprendizagem não-supervisionada, temos os exemplos, mas eles não estão classificados ou marcados com a resposta que o sistema precisa dar e no terceiro grupo temos alguns exemplos marcados e outros não. Chamamos a este terceiro grupo de aprendizagem semi-supervisionada. Alguns autores ainda distinguem um quarto grupo de técnicas, que não são tão exploradas em visão computacional, chamado de aprendizagem por reforço, onde o sistema é de alguma forma “punido” (E.g: recebe um número negativo) ou “recompensado” (E.g: recebe um número positivo) dependendo da resposta que oferece na fase de aprendizagem.

A aprendizagem supervisionada é a mais utilizada em problemas de reconhecimento na visão computacional pois geralmente não é complicado produzir

imagens marcadas ou classificadas indicando quais os tipos de objetos que precisamos reconhecer. É preciso ter em mente, no entanto, que a qualidade dos resultados é muito dependente da qualidade do conjunto de exemplos que é preparado para "ensinar" o sistema. Como a grande a maioria das técnicas de aprendizagem automaticamente tem como suporte teórico a estatística, existe uma relação muito forte entre aprendizagem automática e inferência estatística. Todos aqueles cuidados na obtenção das amostras, necessários para a realização de um experimento estatístico confiável, se aplicam também à aprendizagem automática. Existem centenas de técnicas de aprendizagem automática disponíveis que podem oferecer resultados melhores ou piores dependendo do problema. O vídeo da Figura 40 apresenta uma das técnicas mais básicas de aprendizagem automática e pode ser muito útil para se ter uma noção de como um computador pode aprender.

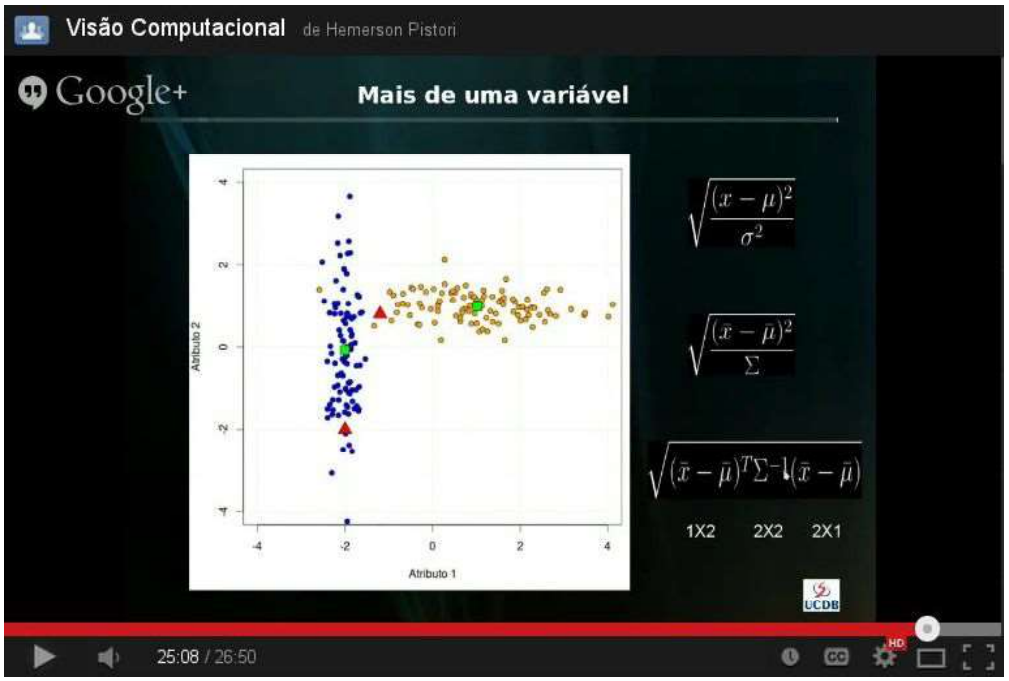

Figura 40: Apresentação sobre um algoritmo simples de aprendizagem automática. Disponível em http://www.youtube.com/watch?v=0tMw0X_Efrw&feature=share

O software Weka é muito útil para avaliar e comparar diferentes técnicas de aprendizagem automática e o vídeo da Figura 41 mostra como utilizar algumas das funcionalidades desse software. Estão disponíveis no weka implementações dos principais algoritmos de aprendizagem supervisionada, como os baseados em máquinas de vetores de suporte, árvores de decisão e redes neurais artificiais.

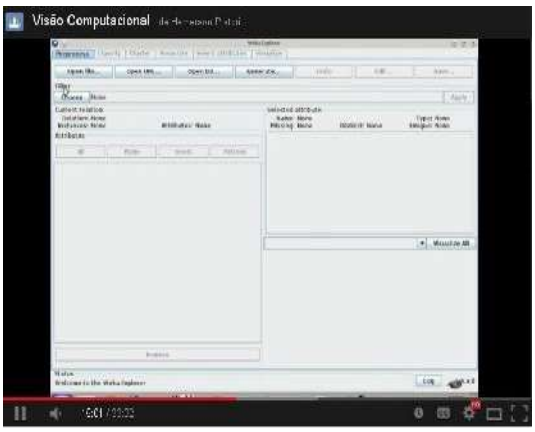

Figura 41: Vídeo sobre o Weka disponível em http://youtu.be/tXFxzbjU0xY

6.3 Métodos sintáticos ou estruturados

Existe um conjunto de métodos que podem ser utilizados na etapa de reconhecimento de um sistema de visão computacional que tem sua origem na área das linguagens. Sabemos que o ser humano possui uma grande capacidade para representar e interpretar informações codificadas em uma sequência de fonemas ou de letras de um alfabeto. É a nossa capacidade de utilizar uma língua, como Português ou Inglês. Na computação, temos uma área chamada de linguagens formais e autômatos, que trata da construção de algoritmos e programas de computador, capazes, por exemplo, de reconhecer se uma determinada sentença pertence ou não a uma língua representada através de uma gramática. Esses algoritmos já são essenciais na construção dos compiladores de linguagens de programação e também tem aplicações em sistemas capazes de traduzir textos entre diferentes línguas ou nos intensamente utilizados corretores ortográficos e corretores gramaticais disponíveis em qualquer editor de textos moderno.

A adaptação para a área da visão computacional não é trivial, pois temos que sair do universo das letras e palavras para o universo das imagens. Existem no entanto vários avanços e estratégias para permitir que conceitos como gramáticas e análise sintática possem ser aplicados também ao universo das imagens. A análise sintática permite identificar se uma determinada sequência de letras ou palavras faz sentido dentro de uma determinada língua e além disse permite identificar,

classificar e representar as estruturas e sub-estruturas que compõem determinada sentença (como sujeitos, predicados, artigos, pronomes, adjetivos, etc). Ao transpor esse tipo de ideia para o universo das imagens, pretende-se melhorar o desempenho de sistemas de reconhecimento de objetos ou cena, por exemplo, analisando-se estruturas hierarquizadas e criando-se gramáticas que possam representar de forma resumida como determinada classe de objetos pode ocorrer em imagens. Em um sistema de visão computacional para classificar imagens de leveduras entre viáveis e inviáveis, na abordagem sintática teríamos algo como a língua visual das leveduras viáveis e a língua visual das leveduras inviáveis (ou seja, duas língua). Para cada língua teríamos que construir uma gramática ou autômato que a representasse. Essas gramáticas poderiam ser geradas automaticamente utilizando-se técnicas derivadas de uma área chamada inferência gramatical, que oferece formas de se chegar a uma gramática através de exemplos de sentenças que pertençam a língua que se pretende representar com a gramática (no caso da visão computacional, as sentenças seriam imagens ou sequências de caracteres que de alguma forma foram extraídos da imagem). Com as duas gramáticas disponíveis, novas imagens de levedura seriam classificadas utilizando-se algum tipo de análise sintática, que iria ajudar a determinar a qual das duas línguas a nova imagem de levedura pertence, e assim decidir se trata-se de uma levedura viável ou inviável. O vídeo indicado na Figura 42 apresenta com mais detalhes como funcionam duas estratégias de reconhecimento sintático de padrões.

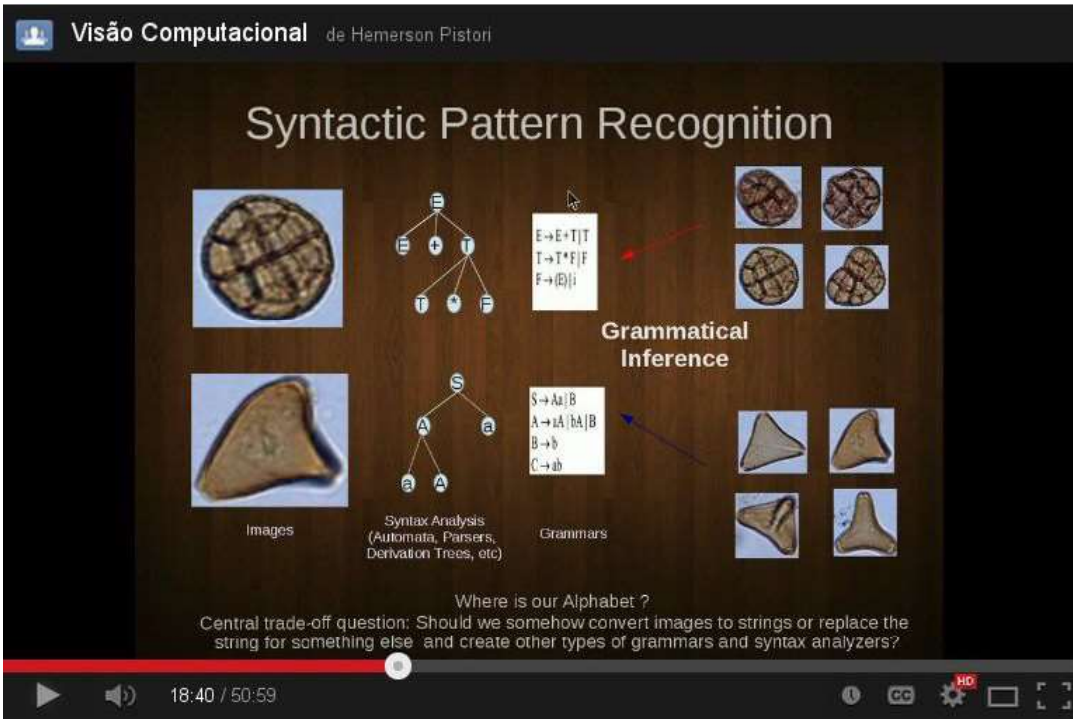

Figura 42: Vídeo sobre reconhecimento sintático de padrões disponível em http://youtu.be/rznfZaSguCU