

6 Conclusão e Trabalhos Futuros

6.1. Principais Contribuições

As principais contribuições desta dissertação foram o aprimoramento da técnica proposta inicialmente por Viola e Jones (2001) e o desenvolvimento de um protótipo que permite o treinamento e a detecção de objetos em imagens ou vídeos digitais.

Enquanto a maior parte dos trabalhos existentes na literatura que lidam com o reconhecimento de objetos não consegue atingir o requisito de tempo real ou apenas apresenta soluções para detecção de objetos relacionados a um problema específico (e.g. reconhecimento de pedestres, sinalização de trânsito), este trabalho apresenta um protótipo capaz de detectar objetos em tempo real e apresentar uma solução genérica o suficiente para a detecção de qualquer tipo de objeto, inclusive a detecção de vários objetos paralelamente.

O aprimoramento da técnica de busca por objetos demonstrado neste trabalho é de suma importância para se obter altas taxas de processamento de imagens ou quadros de um vídeo, suficientes para atender ao requisito de tempo real, inclusive em vídeos de alta resolução. Tal aprimoramento foi possível através da combinação da utilização de: (1) técnicas de segmentação de vídeo, (2) a utilização do conceito de Imagem Integral, (3) uma nova heurística desenvolvida neste trabalho para o descarte de regiões da imagem e (4) a paralelização da busca caso seja necessária a detecção de mais de um objeto simultaneamente.

Com a disponibilização de um ambiente onde o usuário consiga facilmente treinar novos objetos, verificar a performance que o treinamento realizado atinge em relação ao número de detecções corretas, falsos positivos e não detecções de objetos, e selecionar quais cascatas de classificadores devem ser usadas em um vídeo ou imagem para a detecção dos objetos, a questão do problema de reconhecimento de objetos em vídeos em tempo real pode ser levado a um passo adiante.

6.2. Trabalhos Futuros

Uma primeira fase de trabalhos futuros decorrentes desta dissertação refere-se à pesquisa sobre como incrementar o nível de desempenho dos mecanismos de treinamento. O algoritmo para a geração da cascata de classificadores é o mesmo utilizado por Viola e Jones (2001). Apesar de o algoritmo ser genérico e apresentar excelentes resultados quanto à seleção das principais características de um objeto e o rápido descarte de regiões, ele é demasiadamente lento. Um único treinamento pode levar desde algumas horas até dias para ser completado, dependendo do objeto e a quantidade de imagens que compõe os conjuntos positivos e negativos. Este é um ponto vital para que o uso desta técnica para detecção de objetos seja maior e incentive o desenvolvimento de novas pesquisas relacionadas.

Uma nova técnica que permita o reconhecimento automático de qualquer objeto previamente treinado sem a necessidade de informar ao algoritmo que objetos devem ser procurados em um vídeo seria de extrema importância, pois o usuário deixaria de ser obrigado a especificar tais objetos tornando desta forma o algoritmo de detecção mais “inteligente” e independente de informações adicionais do usuário. Uma possível solução para este problema seria a utilização de cascatas de classificadores que formassem uma hierarquia de padrões e formas geométricas, de maneira similar ao trabalho de Epshtein e Ullman (2005). Para tanto, formas geométricas mais básicas como círculos e quadrados estariam nos níveis mais próximos à raiz das cascatas, enquanto formas mais particulares do objeto estariam nos níveis mais próximos às folhas.

Um estudo mais avançado quanto à paralelização do rastreamento por objetos se faz necessário quando o número de objetos desejados numa cena é demasiadamente grande. A utilização do processamento distribuído em uma rede de computadores (*cluster*) tende a ser uma boa alternativa, mas estudos mostram-se necessários para comprovar o desempenho através da distribuição do processamento sobre a rede e os custos envolvidos durante a transmissão de uma grande quantidade de dados entre os processadores, como os quadros de um vídeo em alta resolução e as cascatas de classificadores de cada objeto.

Sob o enfoque de imagens e vídeos em alta resolução, cumpre verificar se a redução da resolução da imagem a ser processada com fins de reconhecimento de objetos pode ser realizada sem perdas de detecção dos objetos na cena por parte do algoritmo detector. Imagens com resolução de

1.920 x 1.080 poderiam ter a sua resolução reduzida, por exemplo, por amostragem a fim de reduzir a quantidade de regiões a serem analisadas pelo detector de objetos, pois tal redução não alteraria de forma significativa as formas dos objetos que compõem a cena. Portanto, é necessário um estudo que constate o quanto uma imagem poderia ser reduzida sem prejudicar o reconhecimento de objetos, desta forma mantendo o algoritmo rápido o suficiente para o requisito de tempo real.

Outro ponto a ser desenvolvido é a maneira como as informações dos objetos treinados podem ser armazenadas em um banco de dados de forma otimizada. Esta questão é importante porque é desejado algum nível de interatividade com o objeto detectado pelo algoritmo. Desta forma o usuário poderia obter maiores informações sobre o objeto de interesse, por exemplo, visualizando mais imagens relacionadas ao objeto ou obtendo a sua descrição completa, fabricante, dados técnicos, etc. Maiores trabalhos nesta questão do armazenamento das informações complementares dos objetos seriam extremamente úteis em relação ao desenvolvimento de novas aplicações interativas, entre elas novas aplicações para a TV Digital Interativa ou *Internet TV*, onde as novas possibilidades de propaganda e marketing de produtos poderiam ser investigadas.

Por fim, seria interessante o desenvolvimento de uma técnica que permita a detecção de objetos rotacionados ou mesmo deformados pela cascata de classificadores. Desta forma, objetos que estejam em posições distintas das que foram originalmente treinados possam ser detectados. Um passo nesta direção foi dado no trabalho realizado por Messom e Barczak (2006), em que os autores apresentam a utilização de “características Haar” rotacionadas entre 0^0 e 90^0 para a detecção de objetos com sucesso, porém com o aumento do custo computacional do algoritmo de detecção, acarretando a perda de parte do desempenho do algoritmo original.