1
Introdução

1.1.
Motivação

Com o advento da TV Digital no país, abre-se a possibilidade de novas formas de interatividade com os telespectadores através de aplicativos interativos executados em seus set-top boxes. Dentre as novas possibilidades, podemos citar como relevantes as aplicações que são executadas de forma sincronizada aos vídeos da programação transmitida. Podem ser citados como exemplo programas de TV que possuem propagandas "embutidas" onde o telespectador pode interagir com diversos objetos contidos na cena e obter mais informações sobre eles, como fotos, uma descrição mais detalhada, visualização do objeto em 3D e até mesmo poder comprá-lo *online*. Para que os telespectadores possam interagir com tais objetos é necessário que o aplicativo em questão seja capaz de reconhecer suas localizações no vídeo e que ele seja necessariamente executado em tempo real. Neste trabalho é apresentado um protótipo de ambiente que possibilita a detecção em tempo real de um ou mais objetos e rastreá-los durante a exibição de um vídeo de alta definição. A Figura 1 ilustra um protótipo de aplicação feito pelo Laboratório VisionLab/PUC-Rio, onde os objetos em cena, num programa de canal de vendas em HDTV, são detectados como áreas ativas que respondem à seleção do telespectador.

Figura 1. Área ativa para interação com o telespectador (extraído de VisionLab(2007))

PUC-Rio - Certificação Digital Nº 0621288/CA

Existem diversas técnicas propostas na literatura acadêmica para a detecção de objetos em imagens digitais. Porém nenhuma delas apresenta uma abordagem definitiva para resolver o problema de detecção de objetos de forma genérica, seja ela em tempo real ou não. As soluções existentes são para resolver este problema em condições específicas de acordo com o domínio da aplicação que se deseja criar.

Por exemplo, existem soluções que visam um índice de detecção de um objeto corretamente próximo de 100% (praticamente sem uma falsa detecção) em detrimento da velocidade, e outras soluções que focam justamente a detecção em tempo real, mas com índices de falsos positivos e não detecções elevadas. Existem ainda outras soluções que não levam em consideração alguns aspectos importantes como mudança de luminosidade na imagem, sombras, mudança no cenário e variação no tamanho de objetos nas cenas.

Para o problema exposto, é necessário um algoritmo que consiga mesclar o ponto forte de técnicas existentes na literatura a fim de identificar corretamente os objetos de interesse em um vídeo, e o mais importante, que esta identificação seja realizada em tempo real.

Uma solução que mais se aproxima do que se é desejado é o trabalho desenvolvido por Viola e Jones (2001) e posteriormente estendido por Lienhart e Maydt (2002). Em seu trabalho, Viola e Jones (2001) propõem um algoritmo confiável capaz de detectar objetos em imagens em tempo real, e que segundo eles, possui uma velocidade quinze vezes maior do que outras técnicas propostas anteriormente, como em Rowley et al. (1998), Friedman e Russel (1997) e Papageorgiou et al. (1998).

O grande diferencial da solução de Viola e Jones (2001) é que eles apresentam um algoritmo baseado em características ao invés de pixels, o que já garante um desempenho muito maior em relação aos demais algoritmos que se baseiam em análise de pixels. Também é proposta a utilização de uma cascata de classificadores para a análise da existência ou não de um objeto nas regiões da imagem. Entretanto, a solução apresentada possui limitações quanto às resoluções de imagens que podem ser processadas em tempo real e ao número de objetos que podem ser buscados simultaneamente.

PUC-Rio - Certificação Digital Nº 0621288/CA

1.2. Proposta

Neste trabalho é proposta a geração de um ambiente integrado que permita que um usuário consiga treinar o algoritmo de detecção a fim de reconhecer um objeto qualquer e permita a sua identificação em tempo real, tanto em imagens quanto em vídeos digitais de alta resolução. Para atingir tal objetivo, é necessário um estudo, projeto e construção de um algoritmo capaz de processar rapidamente uma imagem em alta resolução, descobrindo-se áreas que podem ser descartadas sem que seja necessário qualquer tipo de análise.

Para o protótipo a ser criado é necessário que ele atenda aos seguintes requisitos levantados:

- Permitir a realização de um treinamento do algoritmo de classificação para um objeto qualquer de interesse do usuário;
- Analisar o desempenho do treino efetuado, exibindo para um dado treinamento realizado a quantidade de detecções realizadas com sucesso, a quantidade de detecções não realizadas e a quantidade de falsos positivos encontrados, de tal forma que de acordo com os resultados obtidos, o usuário possa realizar um novo treinamento;
- Realizar a detecção de objetos em tempo real em uma imagem ou em um vídeo de alta definição, indicando ao usuário a localização de todos os objetos identificados;
- Efetuar a busca de um ou mais objetos simultaneamente sem perda de desempenho na quantidade de *frames* processados por segundo.

Conforme já mencionado, dentre os estudos existentes na literatura o trabalho apresentado por Viola e Jones (2001) é o que possui uma solução mais abrangente e genérica para a rápida detecção de qualquer tipo de objeto, onde é proposta a utilização de uma cascata de classificadores que possibilita um rápido descarte de áreas da imagem que não possuam o objeto procurado, atingindo o requisito de tempo real, mas apenas para vídeos de baixa resolução.

Nesta dissertação são propostas algumas melhorias a fim de suprir algumas das carências do trabalho de Viola e Jones (2001) [1], como a detecção de objetos em tempo real em vídeos de alta resolução e a possibilidade de se buscar vários objetos simultaneamente em um vídeo. Ao longo deste trabalho é

[1] As extensões propostas por Lienhart e Maydt (2002) (ver Figura 14) não afetam as análises e propostas feitas na presente dissertação.

PUC-Rio - Certificação Digital Nº 0621288/CA

mostrado que tais carências podem ser sanadas através do uso de técnicas de segmentação de vídeo, pela criação de um novo algoritmo capaz de reduzir a quantidade de regiões analisadas na imagem, e por fim, através do uso de paralelismo durante as etapas do algoritmo que podem ser executadas de forma independente.

1.3. Estrutura

A organização dos capítulos da dissertação é listada a seguir.

No capítulo 2, são apresentados alguns dos trabalhos mais relevantes na literatura relacionados à questão de treinamento e detecção de objetos em tempo real. São mostradas as principais características e embasamentos teóricos para cada uma das técnicas, assim como as suas respectivas carências.

No capítulo 3, é descrita de maneira minuciosa a técnica proposta por Viola e Jones (2001) sobre a qual se baseiam as propostas de melhorias deste trabalho. Nele é descrito o funcionamento do treinamento de novos objetos e como é feita a detecção de objetos de maneira rápida e robusta.

O capítulo 4 apresenta as propostas de melhorias no algoritmo de detecção de objetos em tempo real, através da aplicação de uma nova heurística para segmentação e descarte de regiões de quadros de um vídeo, assim como um método para paralelização da busca quando for necessária a detecção de vários objetos simultaneamente.

O capítulo 5 fornece a arquitetura do protótipo desenvolvido ao longo deste trabalho e os ganhos obtidos através de sua utilização. É fornecida a descrição de cada um dos módulos construídos, com exemplos do treinamento de novos objetos, teste de desempenho do treinamento e a detecção de objetos em imagens e vídeos de alta resolução. São mostrados também alguns gráficos comparativos do desempenho do módulo de detecção relacionados a outros trabalhos da área.

Por fim, o capítulo 6 apresenta as conclusões gerais. São relacionadas as principais contribuições desta dissertação e analisadas as perspectivas de trabalhos futuros.

PUC-Rio - Certificação Digital Nº 0621288/CA