

# Método Para Segmentação de Vídeo em Tempo-Real para Aplicações em Teleimersão

Silvio Ricardo Rodrigues Sanches<sup>1</sup>, Antonio Carlos Sementille<sup>2</sup>, Romero Tori<sup>1</sup>

<sup>1</sup> Laboratório de Tecnologias Interativas - Interlab  
Escola Politécnica da Universidade de São Paulo

<sup>2</sup> Laboratório de Sistemas de Tempo Real - LSTR  
Universidade Estadual Paulista (UNESP)

*silviorrs@usp.br, semente@fc.unesp.br, tori@acm.org*

## Abstract

*An essential task in immersive telecommunication systems is segmentation, i. e., the extraction of the actor from its original context in order to integrate it in a 3D environment. Modern techniques for video segmentation are applied in virtual studios based on dedicated hardware. This paper presents a method for real-time video segmentation, based on chroma-key technique, for teleimmersion applications. This method provides quality in foreground silhouette cut and smoothness integration between real image and virtual environment using imperfect background and conventional hardware to capture the source image.*

## 1. Introdução

Sistemas de videoconferência são ferramentas importantes e largamente utilizadas para que grupos de pessoas ligadas a empresas ou instituições trabalhem de forma cooperativa, mesmo que estejam geograficamente separadas.

Ainda que o vídeo consiga quebrar a barreira da distância física de forma eficaz, torna-se um grande desafio desenvolver sistemas de telecomunicação que aumentem a sensação de co-presença entre os participantes. Uma das formas de se conseguir tal melhoria é fazer com que um único ambiente (que pode não ser real) pareça compartilhado entre eles. Recursos de computação gráfica, mais especificamente de Realidade Aumentada (RA), são bastante apropriados para atingir tal objetivo.

Técnicas de RA aplicadas a sistemas de

telecomunicação fizeram surgir uma linha de pesquisa denominada Telecomunicação Imersiva (*Immersive Telecommunication*). Por meio de recursos de comunicação mais sofisticados que os das videoconferências tradicionais pode-se compartilhar um mesmo espaço virtual e obter formas de interação baseadas em imagens realistas, direcionamento do som, e interação dos participantes com objetos virtuais e com os demais participantes remotos [6]. Visualização em CAVES [3] ou técnicas de estereoscopia podem ser utilizadas para que o nível de imersão seja aumentado.

Grande parte das pesquisas relacionadas à Telecomunicação Imersiva está empenhada em oferecer alta qualidade na representação humana dentro do ambiente compartilhado. Sistemas desse tipo são predominantemente virtuais, tendo as imagens dos participantes extraídas do seu contexto original e inseridas, de alguma forma, no ambiente tridimensional (avatar). Muitos são os métodos descritos na literatura para realizar tal tarefa [7][14][10], deixando a cargo da aplicação a escolha do mais adequado.

Aplicações de telecomunicação imersiva, em sua maioria, são apoiadas em estruturas que envolvem múltiplas câmeras de alta definição, computadores de elevado poder computacional e *hardwares* gráficos de alto desempenho. Esforços têm sido feitos no intuito de se encontrar alternativas para o barateamento dessa tecnologia.

Entre os passos envolvidos no processo, uma

tarefa fundamental é a necessidade de se extrair o ator (ou elemento real) do seu contexto original, para que se possa inseri-lo no ambiente 3D. Técnicas modernas de composição de imagens são aplicadas em estúdios e se baseiam em algoritmos complexos para que a imagem resultante pareça natural.

Nesse contexto, apresenta-se neste artigo um método para realizar *chroma-key* (segmentação de vídeo que exige um fundo de cor homogênea, em geral verde ou azul, para extração do elemento de primeiro plano) baseado no algoritmo de Bergh e Lalioti [18] e que utiliza a técnica da diferença de imagens [13] como auxiliar no processo. O objetivo é proporcionar qualidade no recorte e suavidade na integração com o ambiente tridimensional compartilhado pelos usuários do sistema de telecomunicação, mesmo que as imagens de origem sejam obtidas em ambientes não dedicados e capturadas por equipamentos não profissionais.

Para que favoreça seu pleno entendimento, este artigo é organizado da forma como segue. Na Seção 2 são apresentadas as principais abordagens para segmentação de vídeo, as particularidades a respeito do posicionamento das câmeras em sistemas imersivos de telecomunicação e os trabalhos relacionados, no que se refere aos algoritmos mais utilizados.

Os experimentos preliminares que serviram como base para a implementação do protótipo final são mostrados na Seção 3. Esses algoritmos foram implementados e avaliados. A abordagem proposta é descrita em detalhes na Seção 4 e os resultados obtidos são discutidos na Seção 5. Finalmente, na Seção 6 são apresentadas as conclusões e perspectivas de trabalhos futuros.

## 2. Principais Abordagens para Segmentação de Vídeo

As principais abordagens de segmentação de vídeo para extração de elementos de primeiro plano são diferenciadas principalmente por características relacionadas ao plano de fundo. Entre essas técnicas podem-se destacar o caso em que o plano de fundo é conhecido e o caso em que o plano de fundo é completamente arbitrário.

O primeiro caso é aplicado sobre um ambiente totalmente controlado e o princípio consiste em fotografar o ator (ou objeto) à frente de um plano de fundo de cor constante e extrair

a imagem de primeiro plano a cada quadro de vídeo [4]. Uma abordagem mais recente para isolamento do elemento de primeiro plano é a aplicação de técnicas que agem sobre planos de fundo arbitrários, processo conhecido como “matting de imagem natural” [2][1]. O método proposto, a ser aplicado no sistema de vídeo-avatar [14][6] do projeto TIDIA-Ae [16], está no contexto do primeiro caso.

### 2.1. Cenários para Captura do Vídeo

Sistemas imersivos de telecomunicação podem ser utilizados em situações como reuniões comerciais (entre funcionários de uma empresa ou entre empresas), ensino a distância ou entretenimento. Em aplicações desse tipo a infra-estrutura exigida pode variar consideravelmente e a abordagem de remoção de fundo mais adequada a cada uma delas deve ser utilizada [6].

Sistemas que reconstróem o avatar baseado em imagens obtidas de câmeras dispostas ao redor do ator posicionam essas câmeras de modo que não ocupem o campo de visão umas das outras, ou não se baseiam apenas na informação da cor do fundo para isolar o elemento de primeiro plano [7][12][5].

Avatares planos [14] ou 2,5D [15][6] normalmente são gerados a partir de imagens capturadas por câmeras posicionadas à frente do ator (ou elemento de interesse) e podem ser recortados por meio de técnicas de *chroma-key*. No caso desta pesquisa, cujos resultados serão aplicados em um sistema de vídeo-avatar 3D no qual a captura é feita por uma câmera (ou duas posicionadas lado a lado), uma abordagem que limite o plano de fundo a uma cor constante e se baseie na identificação dessa mesma cor para remover fundo é perfeitamente aceitável.

### 2.2. Trabalhos Relacionados

Um modo bastante sofisticado de realizar *chroma-key* é por meio do equipamento Ultimatte [17]. Rápido o bastante para ser executado em transmissões ao vivo, o Ultimatte é amplamente utilizado em produções de vídeos há algum tempo. Esse equipamento, que possui um algoritmo baseado no sistema de cores HLS [8] possibilita a criação de composições com elementos como fumaça, objetos transparentes, diferentes tons de azul e sombras. Bergh e Lalioti [18] apresentaram um método eficiente de realizar *chroma-key* em software, baseado

no sistema de cores RGB para eliminação do azul (fundo). O algoritmo torna fácil o ajuste do threshold para compensar variações de iluminação.

Yamashita et. al. [19] buscaram soluções para a restrição da cor azul no elemento de interesse utilizando-se de um plano de fundo composto por listras de duas cores. O algoritmo analisa condições de adjacência, além da informação de cor. No trabalho de Qian e Sezan [13] não há restrição a uma determinada cor para o plano de fundo. O método compara uma imagem capturada previamente (somente do fundo) com o quadro atual resultando em uma imagem que contém apenas o elemento em primeiro plano. Uma visão geral das técnicas modernas de *chroma-key* aplicadas em estúdios virtuais é apresentada em Gibbs et.al. [4].

### 3. Experimentos Preliminares

Uma das grandes dificuldades em se realizar *chroma-key* em ambientes não profissionais é o difícil controle da iluminação do local, o que causa variações no tom de azul (ou verde) adotado como plano de fundo. A qualidade do dispositivo que realiza a captura da imagem também influencia de modo determinante o resultado final da composição. Para contornar esse problema é importante que o algoritmo se adapte a essas situações ou possibilite um ajuste simples, que possa ser realizado por um usuário sem conhecimentos técnicos. Alguns algoritmos descritos na literatura foram avaliados em relação a sua adaptabilidade a condições do ambiente, facilidade de ajuste, qualidade no recorte e tempo de execução.

Uma abordagem clássica no que se refere à segmentação de vídeo é a utilização de um plano de fundo de cor constante (azul ou verde). O modo mais intuitivo de isolar a cor a ser eliminada (cor chave) é ajustar cada canal R, G e B de forma que a combinação de valores que representa essa tonalidade seja identificada, e assim, impedir que a cor seja exibida na imagem final. Esse tipo de ajuste, no entanto, é trabalhoso principalmente quando se considera que a iluminação muitas vezes não é uniforme e que as cores da mesma imagem podem variar consideravelmente quando obtidas em diferentes dispositivos, fazendo-se necessárias configurações particulares a cada um deles.

Algumas soluções baseadas em *chroma-key*

relatadas na Subseção 2.2 mostram algoritmos eficientes e que possibilitam um controle mais simples por parte do usuário. Um trabalho a ser ressaltado é o de Bergh e Lalioti [18], que exibe uma abordagem baseada no esquema de cores HLS e apresenta uma solução inovadora implementada utilizando-se o esquema RGB.

Soluções alternativas ao *chroma-key* e que apresentam resultados satisfatórios também podem ser encontradas. Cabe destacar o método de remoção de fundo baseado na diferença entre duas imagens desenvolvido por Qian e Sezan [13]. Os métodos citados foram implementados e analisados para que fosse possível estabelecer uma comparação entre eles.

O primeiro protótipo implementado é baseado no sistema de cores HLS [18]. A representação geométrica desse espaço de cores é um hexágono na base de dois cones que visto de cima tem a forma exibida na Figura 1.

Faz-se a representação da cor pelo ângulo ao redor da borda, que inicia do vermelho  $0^\circ$ , enquanto a posição do azul está em  $240^\circ$ . Para eliminar canais de azul, basta isolar o seu espaço de cores, que possui seus limites próximos a  $230^\circ$  e  $250^\circ$  [18]. A escolha da abertura do ângulo é arbitrária e deve ser feita de modo a melhor adequar-se à tonalidade do fundo azul da aplicação.

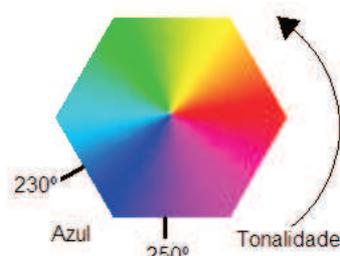
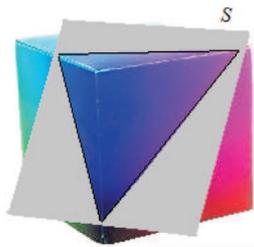


Figura 1. Espaço de cores HLS

A solução para isolamento do azul, baseada no espaço de cores RGB, proposta por Bergh e Lalioti [18] considera um plano S que corta a extremidade azul do cubo RGB. As cores localizadas acima deste plano são consideradas parte do fundo (azul) da imagem e o restante pertence ao elemento de primeiro plano. Uma explicação mais detalhada do método é apresentada na Seção 4. O cubo RGB cortado pelo plano S é mostrado na Figura 2.

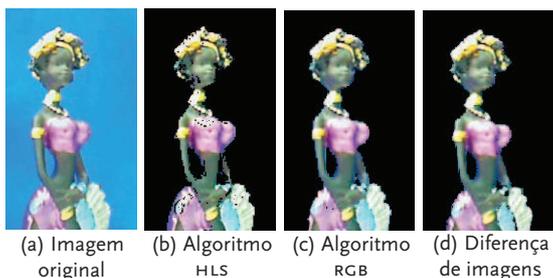
A remoção do fundo utilizando-se o método de diferença de imagens [13] é feita capturando-

se previamente uma imagem do fundo, que é usada como referência. O quadro atual originado da câmera é comparado *pixel a pixel* com a imagem armazenada. Se não houver coincidência de cores, o *pixel* é considerado pertencente ao primeiro plano. Uma pequena faixa de tolerância pode ser admitida para compensar variações na iluminação. O algoritmo de diferença de imagens implementado é baseado no módulo de subtração de fundo do sistema AVMTX [9].



**Figura 2.** Espaço de cores RGB cortado pelo plano S [18]

Na Figura 3 são exibidas as imagens obtidas da execução dos protótipos que implementam os algoritmos citados. Deve ser observado que nos três casos, um *threshold* binário foi utilizado. Isso significa que a imagem é segmentada de forma que a cor dos *pixels* seja a do plano de fundo ou a do elemento de primeiro plano, sem aplicação de nenhum recurso para evitar o “serrilhamento” nas fronteiras entre as duas imagens. Um fator a ser notado é que o elemento recortado contém uma tonalidade de azul, o que dificulta a eliminação dessa mesma cor em tonalidade diferente da escolhida para representar o fundo constante.



**Figura 3.** Comparação entre os métodos de remoção de fundo implementados

Com base nesta análise preliminar, que serviu como ponto de partida para o desenvolvimento do método proposto, foi possível observar algumas características dos algoritmos implementados, quando aplicados em ambientes com condições de iluminação

instáveis. Notou-se que a solução proposta por Bergh e Lalioti [18] baseada no espaço de cores RGB se mostrou eficiente e mais estável em relação à variação da iluminação que os métodos da diferença de imagens [13] e do algoritmo HLS implementado.

Em relação à simplicidade de ajuste do *threshold*, tanto o algoritmo RGB quanto o de diferença de imagens podem ser regulados modificando-se um único valor, conforme será detalhado na Seção 4. Por outro lado, no modelo HLS, valores mínimo e máximo de *threshold* devem ser configurados para que se possa isolar a faixa de azul no hexágono que representa o espaço de cores (Figura 1).

Com exceção do algoritmo HLS implementado, que apresentou algumas imperfeições na imagem final, o que pode ser observado na Figura 3(b), os demais algoritmos, cujas imagens geradas pelos protótipos podem ser visualizadas na Figura 3(c) e 3(d) se mostraram satisfatórios. Os tempos de execução de todos os protótipos implementados foram considerados dentro do esperado, como será discutido na Seção 5.

É importante ressaltar que os protótipos desenvolvidos nessa fase foram executados em um ambiente sem controle rígido de iluminação, foram baseados nos algoritmos descritos nos trabalhos citados e utilizaram equipamento convencional, especificado na Seção 5 para captura do vídeo. Resultados diferentes podem ser obtidos quando se aplicam os mesmos algoritmos em estúdios, utilizando-se equipamentos mais sofisticados.

#### 4. Abordagem Proposta

Com base nos experimentos da Seção 3 foi possível observar a sensibilidade dos protótipos em relação a variação da iluminação durante o período de execução. Notou-se também que os efeitos dessa variação ficam visíveis principalmente na região central da imagem, e que, na maioria das vezes, são sombras provocadas pela movimentação do elemento de primeiro plano. Outro problema verificado é a dificuldade de se obter um plano de fundo totalmente azul, quando não se dispõe de equipamentos que direcionem corretamente luzes sobre esse plano de fundo.

A abordagem proposta procura minimizar essas limitações utilizando-se o algoritmo da

diferença de imagens como primeiro passo para solucionar o problema do plano de fundo não constante. Dessa forma, informações da imagem de referência possibilitam identificar o que é fundo. A técnica, no entanto, é sensível às variações maiores, que ocorrem próximas ao elemento de primeiro plano.

Por meio do método de Bergh e Lalioti [18], as variações dessa região são compensadas com base nas informações das tonalidades de azul, que são menos sensíveis a essa variação, desde que a cor ainda seja identificada como azul pelo *threshold* de *chroma-key*. Uma função de alfa é aplicada, baseada nesse *threshold*. Essa sequência de operações é mostrada na Figura 4.

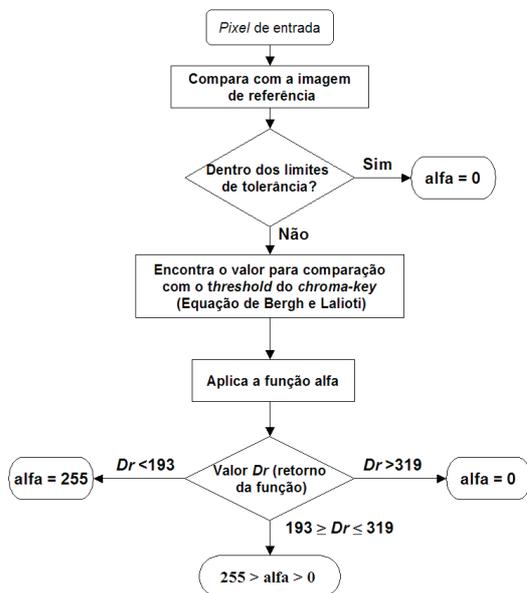


Figura 4. Operações realizadas sobre o pixel de entrada

O recurso do canal alfa [11], utilizado no método, torna possível a representação de transparência do pixel, diminuindo o “serrilhamento” provocado pelo *threshold* binário. Os valores de alfa podem variar em uma escala que vai de 0 (totalmente transparente) a 255 (totalmente opaco), sendo seus valores intermediários o nível de transparência do *pixel*. Tornar um *pixel* mais ou menos transparente é uma estratégia para suavizar os pontos de junção das duas imagens (elemento de primeiro plano e novo plano de fundo) aumentando-se a naturalidade da composição.

Inicialmente um processo de eliminação do plano de fundo da imagem original é aplicado,

baseado no método de diferença de imagens descrito na Seção 3. Dessa forma mesmo que o campo de visão da câmera seja maior que a área azul do fundo ou a tonalidade azul não seja constante devido a dificuldades em projetar luzes adequadamente, o plano de fundo ainda assim é eliminado pois a mesma iluminação presente durante a execução incidia sobre o fundo no momento da captura da imagem de referência.

A técnica da diferença de imagens apresenta bons resultados quando não existem grandes variações na iluminação do ambiente, como mostrado na Figura 3(d), mas possui algumas limitações. A principal delas é que partes do plano de fundo fotografado (tomado como imagem de referência) podem conter cores muito semelhantes às do elemento de interesse, pois, de acordo com a abordagem original, o fundo não necessariamente é formado de uma única cor. Esse problema não ocorre quando uma cor constante é utilizada como fundo, desde que o ator (ou elemento de interesse) não possua em seu vestuário essa mesma cor.

Um grande problema observado é que embora a técnica resolva o problema do azul não constante no plano de fundo, como mostrado na Figura 5(a) na qual se observa a dobra do tecido, o problema persiste quando a iluminação varia

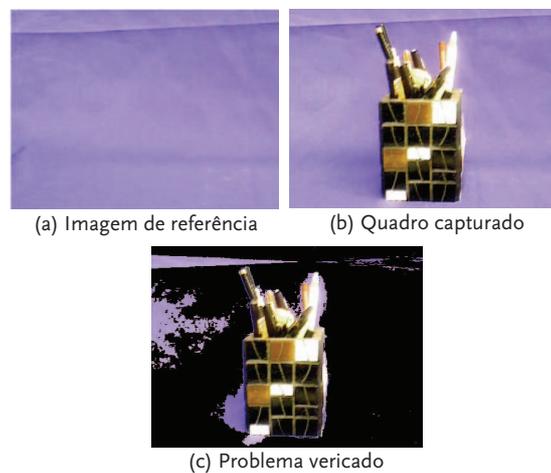


Figura 5. Método da diferença de imagens quando há variação da iluminação

durante a execução da aplicação (Figura 5(c)).

E possível minimizar este efeito indesejado aumentando-se dinamicamente a faixa de tolerância mas esse ajuste pode provocar a ocorrência de imperfeições na imagem que

contém o elemento de primeiro plano. Pixels não muito semelhantes aos da imagem de



**Figura 6.** Problema observado quando se aumenta a faixa de tolerância.

referência podem ser considerados plano de fundo, como pode ser visualizado na Figura 6.

Para solução do problema citado, fez-se necessário utilizar a informação da cor conhecida do plano de fundo.

O *chroma-key* aplicado nesta etapa age apenas sobre o que é considerado elemento de primeiro plano pelo algoritmo anterior, evitando assim que a imagem inteira seja analisada novamente. Como descrito na Seção 3, o algoritmo de *chroma-key* de Bergh e Lalioti (1) [18] é representado por um plano  $S$  que corta a extremidade azul do cubo  $RGB$  (Figura 2). Para identificar-se *pixels* pertencentes ao plano de fundo, um valor  $d$ , calculado pela Equação 1 deve ser obtido. Esse valor representa a distância do ponto analisado ao canto azul do cubo.

$$d = 2 * B - R - G$$

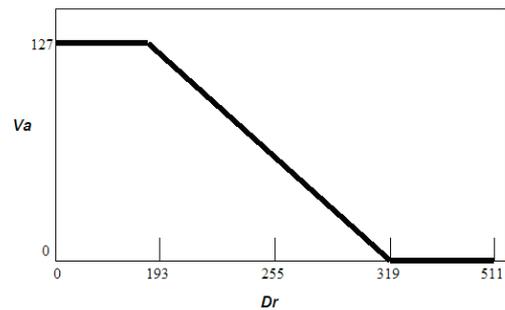
Verifica-se, dessa forma se o *pixel* está acima (fundo) ou abaixo do plano (elemento de interesse). Detalhes do algoritmo são apresentados no trabalho de Bergh e Lalioti [18] assim como uma comparação dos resultados obtidos pelos autores com imagens segmentadas pelo equipamento Ultimatte [17].

Baseado no valor obtido da Equação 1, estabelece-se um *threshold* binário para o *chroma-key*, o que provoca uma “costura” grosseira entre as imagens (elemento de interesse e novo plano de fundo), como pode ser observado na Figura 3(c). Para um resultado mais natural, Bergh e Lalioti sugerem que seja implementada uma função que estime valores de alfa intermediários. Com essa finalidade,

uma função que interpola esses valores baseados no resultado obtido da Equação 1 foi desenvolvida e aplicada.

Como pode ser observado na mesma equação, a distância  $d$  pode assumir 512 valores diferentes. Para que fosse possível trabalhar com dispositivos de baixo custo como webcams, em que as imagens podem variar consideravelmente de acordo com a marca e modelo do equipamento, foi necessário estabelecer um *threshold* inicial  $T_i$  que deve ser somado a distância  $d$ .

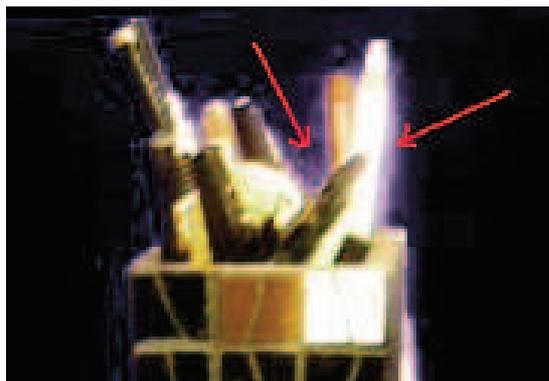
O modelo de câmera utilizado, Microsoft VX-6000 trabalha com um  $T_i$  inicial por volta de 80, já a câmera Avox/Vivicam 3780, utilizada em alguns testes deve ter um  $T_i$  definido em 60. A variação do alfa  $V_a$ , na função definida para este experimento, é feita em um intervalo de 128 valores (de 0 até 255, incrementados em 2 unidades). Uma distância relativa  $D_r$ , foi definida para compensar o valor de *threshold* inicial. A função que interpola os valores de alfa é mostrada na Figura 7, onde  $V_a = \text{alfa}/2$  e  $D_r = d - T_i$ .



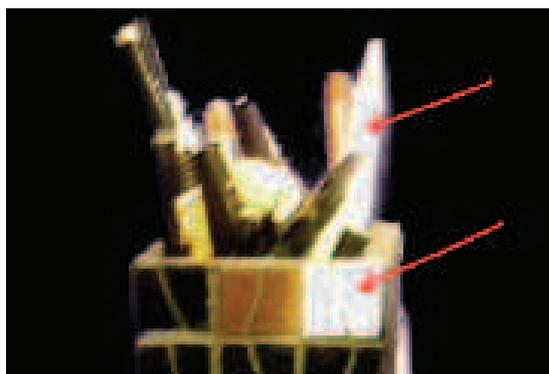
**Figura 7.** Função de alfa utilizada

A importância em se utilizar o algoritmo de diferença de imagens como primeiro passo para eliminação do fundo pode ser observada na Figura 8 em que a função alfa é aplicada diretamente na imagem de origem na qual o azul do plano de fundo não é constante, devido à dificuldade de se direcionar fontes de luz corretamente no ambiente.

Observa-se que interpolar valores de alfa em uma imagem desse tipo pode não remover completamente o fundo original permanecendo-se com um tom azulado em algumas regiões, como indicado na Figura 8(a). Regulando-se o *threshold* do *chroma-key* pode-se minimizar esse efeito, no entanto, quando se atinge determinados limites as cores da imagem de primeiro plano podem ser afetadas, conforme



(a) Eliminação parcial do fundo (tom azulado)



(b) Ajuste do *threshold*

**Figura 8.** Problema observado quando se aumenta o *threshold* da imagem de origem

indicado na Figura 8(b).

Como poderá ser observado nos resultados apresentados na Seção 5, o método proposto evita esse efeito indesejável pois oferece dois parâmetros para ajustes: o *threshold* do *chroma-key* e a faixa de tolerância do método de diferença de imagens.

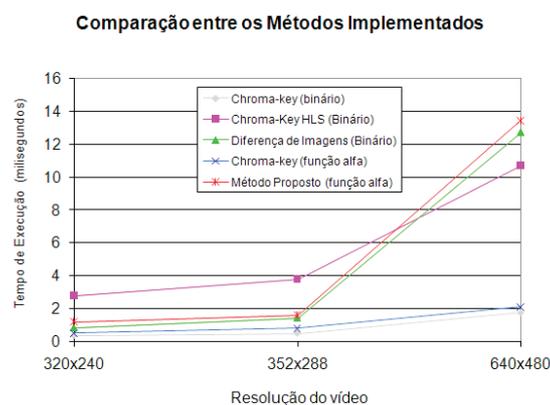
## 5. Resultados

Como forma de avaliação dos resultados foram considerados o tempo de execução, a qualidade no recorte e a adaptabilidade do algoritmo às condições do ambiente, características necessárias para que o método seja aplicável a um sistema que opere em um ambiente pouco controlado. Todos os protótipos foram executados em um computador pessoal Dual Core 1.9 GHZ, com 2 GB de RAM, placa de vídeo NVIDIA GeForce 6150SE (onboard) e equipado com uma webcam convencional modelo Microsoft VX-6000. Outros modelos de câmeras foram utilizadas em alguns testes.

O primeiro requisito a ser avaliado foi o tempo de execução dos protótipos. Um

sistema imersivo de telecomunicação exige atividades custosas computacionalmente e, por esse motivo, a remoção do fundo deve ocupar a menor fatia possível desse tempo de processamento, mantendo-se ainda a qualidade necessária para que o avatar pareça imerso. Cada método teve seu desempenho medido em 5 (cinco) execuções intercaladas. Cada uma delas fazia 100 (cem) acessos ao método, o que significa que foram feitas 500 (quinhentas) medições em cada protótipo implementado.

O gráfico da Figura 9 mostra a variação da média do tempo de execução dos protótipos, tanto nas avaliações preliminares quanto na fase de implementação do protótipo final.



**Figura 9.** Tempo de execução dos protótipos

A implementação do *chroma-key* que não utiliza a função alfa se mostrou a mais eficiente em relação ao tempo de execução. O custo do cálculo de valores intermediários de alfa pode ser notado no desempenho do mesmo algoritmo no qual a função está implementada. Este custo se manteve proporcional à medida que a resolução (quantidade de *pixels* analisados) aumentava. O algoritmo *chroma-key* que se baseia no espaço de cores HLS demandou maior tempo de processamento que os demais métodos baseados em planos de fundo de cor única, resultado do processo de conversão entre os sistemas HLS e o RGB, utilizado pelos dispositivos de captura.

Como era esperado, as abordagens que utilizam a técnica da diferença de imagens apresentaram maior tempo de processamento, provavelmente por necessitarem de informações de duas imagens para serem executadas. A abordagem proposta, que utiliza informações tanto de uma imagem de referência quanto da cor do fundo da imagem original, apresenta um desempenho ligeiramente menor que o método simples de

diferença de imagens, o que pode ser mais bem observado quando se utiliza vídeos com resolução de  $640 \times 480$  pixels. Essa variação, que não foi maior que 1 (um) milissegundo, faz com que a utilização do método seja viável.

A qualidade no recorte é de fundamental importância para que o elemento extraído de seu contexto original possa ser integrado ao ambiente virtual de maneira natural, sem “costuras” em sua silhueta. A composição exibida na Figura 10 é mostrada em detalhes na Figura 11, em que pode ser observada a importância de se permitir valores intermediários de alfa.

Comparando-se as Figuras 11(a) e 11(b) nota-se claramente que, na segunda, as bordas do elemento de primeiro plano são atenuadas, minimizando o problema do “serrilhamento” visualizado na primeira.

A sensibilidade em relação a variação da iluminação, outra característica avaliada, pode ser notada na Figura 12, que mostra dois quadros da cena em que o método é aplicado em condições de iluminação significativamente diferentes. Essa alteração torna a tonalidade do azul escolhida como plano de fundo diferente nos dois vídeos. Apesar disso, o método é executado com qualidade aceitável.

Em se tratando de ambientes não profissionais, o espaço físico disponível muitas vezes pode não ter uma cor constante que ocupe todo o campo de visão da câmera. Na Figura 13 é mostrada a eliminação do fundo nessas condições. As partes brancas, localizadas nas bordas direita e esquerda são removidas pelo primeiro passo do algoritmo (comparação com a imagem de referência).

Variações significativas na iluminação podem fazer com que as partes não azuis não sejam removidas por completo, pois o algoritmo *chroma-key* aplicado no segundo passo age apenas sobre a tonalidade azul. No entanto, na maioria dos casos as variações se concentram na parte central da cena e o fundo é retirado por completo.

A aplicação do método no sistema de vídeo-avator do TIDIA-Ae [16] é exibida nas Figuras 14 e 15. Nas Figuras 14(a) e 14(b), é exibida a imagem de origem, capturadas por uma webcam, e o ambiente virtual respectivamente. O avatar extraído de seu contexto original

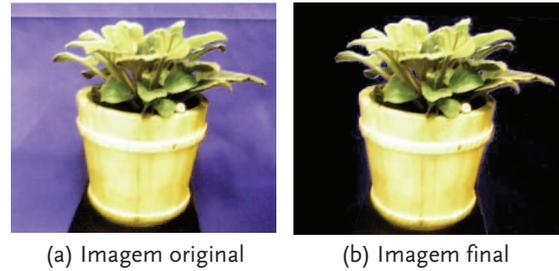


Figura 10. Resultados

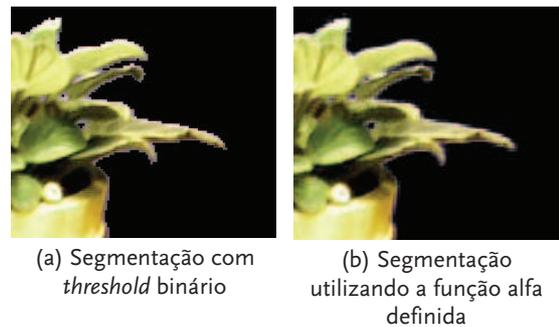


Figura 11. Atenuação das bordas do elemento de primeiro plano



Figura 12. Execução do protótipo em diferentes condições de iluminação

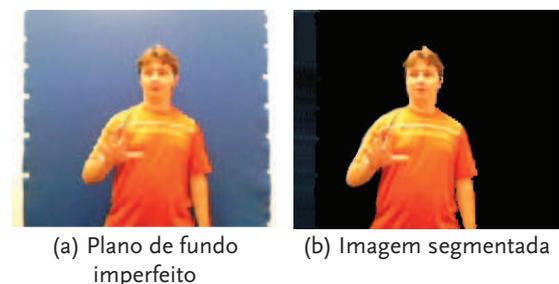


Figura 13. Remoção do plano de fundo não totalmente azul



(a) Imagem de origem

(b) Ambiente virtual



(c) Avatar integrado ao ambiente virtual

**Figura 14.** Integração do avatar no ambiente 3D



**Figura 15.** Sequência de quadros da aplicação em execução

utilizando-se o método proposto, e posteriormente, integrado ao ambiente pode ser visualizado na Figura 14(c). Essa integração é realizada por meio de um método desenvolvido no projeto TIDIA-Ae, que é uma variação da técnica proposta por Yamazaki et. al. [20]. Uma sequência de quadros da aplicação em execução é mostrada na Figura 15.

## 6. Conclusões e Trabalhos Futuros

Neste artigo apresentou-se um método de segmentação de imagens de vídeo que faz uso da técnica do *chroma-key* em conjunto com a abordagem da diferença de imagens. O objetivo principal é possibilitar a remoção do fundo da imagem para que esta seja inserida em um ambiente virtual compartilhado por usuários de um sistema imersivo de telecomunicação.

A abordagem proposta, que permite utilizar dispositivos convencionais para captura do vídeo, busca isolar o elemento de primeiro plano mesmo quando o plano de fundo apresenta variações em sua cor, que deveria ser constante, quando a iluminação se altera

no decorrer da aplicação ou quando o azul não ocupa todo o campo de visão da câmera. Todas essas características são comuns quando se utilizam ambientes não profissionais. Uma função que estima valores de alfa foi desenvolvida para que houvesse uma integração suave entre o elemento de primeiro plano e o novo plano de fundo.

O método se mostrou eficiente quanto ao tempo de execução, quanto à qualidade de recorte do elemento de primeiro plano e quanto à facilidade de ajuste de *threshold* para compensar variações de iluminação durante a execução, dentro de determinados limites.

Uma limitação implícita da técnica do *chroma-key* e que se apresenta na abordagem proposta é que o elemento de primeiro plano não deve conter azul da mesma tonalidade ou de tonalidades semelhantes a do plano de fundo. A câmera de vídeo também deve permanecer imóvel durante a execução da aplicação, limitação imposta pela técnica da diferença de imagens, que se baseia em uma imagem de referência capturada no momento em que a câmera de vídeo se encontrava em determinada posição.

Pretende-se, como trabalhos futuros, implementá-lo para que seja executado diretamente pela placa gráfica, além de tratar impurezas na imagem de primeiro plano provocadas pela contaminação da cor do plano de fundo (Color Spill).

## 7. Agradecimentos

À FAPESP pelo apoio financeiro (Processo nº 2008/10528-4) e aos colegas do Interlab Daniel Tokunaga, Victor Gomes, Lucas Trias, Valdinei Silva, Ricardo Nakamura e Alexandre Tomoyose, pelo apoio. Esta pesquisa é parte do projeto TIDIA-Ae [16], financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (Processo nº 2005/60618-1).

## 8. Referências

- [1] Y.-Y. Chuang. *New Models and Methods for Matting and Compositing*. PhD thesis, University of Washington, 2004.
- [2] Y.-Y. Chuang, B. Curless, D. H. Salesin, e R. Szeliski. A bayesian approach to digital matting. In *Proceedings of IEEE CVPR 2001*, volume 2, páginas 264-271. IEEE Computer Society, December 2001.
- [3] C. Cruz-Neira, D. J. Sandin, e T. A. Defanti.

- Surround-screen projection-based virtual reality: the design and implementation of the cave. In *SIGGRAPH '93: Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, páginas 135-142. ACM Press, 1993.
- [4] S. Gibbs, C. Arapis, C. Breiteneder, V. Lalioti, S. Mostafawy, e J. Speier. Virtual studios: an overview. *Multimedia, IEEE*, 5(1):18-35, Jan-Mar 1998.
- [5] T. Kanade e P. J. Narayanan. Virtualized reality: Perspectives on 4d digitization of dynamic events. *IEEE Comput. Graph. Appl.*, 27(3):32-40, 2007.
- [6] S.-Y. Lee, I.-J. Kim, S. C. Ahn, M.-T. Lim, e H.-G. Kim. Toward immersive telecommunication: 3d video avatar with physical interaction. In *ICAT '05: Proceedings of the 2005 international conference on Augmented tele-existence*, páginas 56-61, New York, NY, USA, 2005. ACM.
- [7] W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, e L. McMillan. Image-based visual hulls. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, páginas 369-374, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.
- [8] Y. Mishima. Soft edge chroma-key generation based upon hexoctahedral color space. United States Patent no 5,355,174, October 1994.
- [9] R. Nakamura e R. Tori. Improving collision detection for real-time video avatar interaction. In *SVR '08: Proceedings of X Symposium on Virtual and Augmented Reality*, páginas 105-114, 2008.
- [10] T. Ogi, T. Yamada, Y. Kurita, Y. Y. Hattori, e M. Hirose. Usage of video avatar technology for immersive communication. In *ACL 2003 Colocated Workshop: First International Workshop on Language Understanding and Agents for Real World Interaction*, 2003.
- [11] T. Porter e T. Du. Compositing digital images. In *SIGGRAPH '84: Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, páginas 253-259, New York, NY, USA, 1984. ACM Press.
- [12] S. Prince, A. Cheok, F. Farbiz, T. Williamson, N. Johnson, M. Billinghurst, e H. Kato. 3d live: real time captured content for mixed reality. *Mixed and Augmented Reality, 2002. ISMAR 2002. Proceedings. International Symposium on*, páginas 7-317, 2002.
- [13] R. Qian e M. Sezan. Video background replacement without a blue screen. *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, 4:143-146 vol.4, 1999.
- [14] R. Siscoutto e R. Tori. Augmented virtuality teleconferencing. In *SVR '04: Proceedings of VII Symposium on Virtual Reality*, páginas 124-136, 2004.
- [15] K. Tamagawa, T. Yamada, T. Ogi, e M. Hirose. Developing a 2.5-d video avatar. *Signal Processing Magazine, IEEE*, 18(3):35-42, May 2001.
- [16] TIDIA-ae. Tidia aprendizado eletrônico. Disponível em <<http://tidia-ae.incubadora.fapesp.br/portal>>, Acessado em Janeiro 2009.
- [17] Ultimatte. Ultimatte corporation. Disponível em <<http://www.ultimatte.com>>, Acessado em Dezembro 2008.
- [18] F. van den Bergh e V. Lalioti. Software chroma keying in an immersive virtual environment. *South African Computer Journal*, 24:155-162, Nov 1999.
- [19] A. Yamashita, T. Kaneko, S. Matsushita, e K. T. Miura. Region extraction with chromakey using stripe backgrounds. *IEICE Transactions on Information and Systems E Series D*, 87:66-73, 2004.
- [20] S. Yamazaki, R. Sagawa, H. Kawasaki, K. Ikeuchi, e M. Sakauchi. Microfacet billboarding. In *EGRW '02: Proceedings of the 13th Eurographics workshop on Rendering*, páginas 169-180, 2002.