



Pós-Graduação em Ciência da Computação

**Técnica de Ghosting Baseada na Detecção de
Movimentos para Tratamento de Oclusão em
Realidade Aumentada**

Por

ARTHUR DE LIMA PADILHA

Dissertação de Mestrado



Universidade Federal de Pernambuco
posgraduacao@cin.ufpe.br
www.cin.ufpe.br/~posgraduacao

RECIFE, 2015



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO



ARTHUR DE LIMA PADILHA

TÉCNICA DE *GHOSTING* BASEADA NA DETECÇÃO DE MOVIMENTOS PARA TRATAMENTO DE OCLUSÃO EM REALIDADE AUMENTADA

DISSERTAÇÃO SUBMETIDA AO CENTRO DE INFORMÁTICA DA
UNIVERSIDADE FEDERAL DE PERNAMBUCO, COMO REQUISITO PARCIAL
PARA OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIA DA COMPUTAÇÃO.
ORIENTADORA: VERONICA TEICHRIEB (VT@CIN.UFPE.BR)

RECIFE
2015

Catálogo na fonte
Bibliotecária Jane Souto Maior, CRB4-571

P123t Padilha, Arthur de Lima

Técnica de ghosting baseada na detecção de movimentos para tratamento de oclusão em realidade aumentada / Arthur de Lima Padilha. – Recife: O Autor, 2015.

73 f.: il., fig.

Orientador: Veronica Teichrieb.

Dissertação (Mestrado) – Universidade Federal de Pernambuco. CIn, Ciência da computação, 2015.

Inclui referências.

1. Ciência da computação. 2. Realidade aumentada. I. Teichrieb, Veronica (orientadora). II. Título.

004

CDD (23. ed.)

UFPE- MEI 2015-69

Dissertação de Mestrado apresentada por **Arthur de Lima Padilha** à Pós Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, sob o título “**Técnica de Ghosting Baseada na Detecção de Movimentos para Tratamento de Oclusão em Realidade Aumentada**” orientada pela **Profa. Veronica Teichrieb** e aprovada pela Banca Examinadora formada pelos professores:

Prof. Silvio de Barros Melo
Centro de Informática / UFPE

Prof. João Paulo da Silva do Monte Lima
Departamento de Estatística e Informática / UFRPE

Profa. Veronica Teichrieb
Centro de Informática / UFPE

Visto e permitida a impressão.
Recife, 13 de março de 2015.

Profa. Edna Natividade da Silva Barros
Coordenadora da Pós-Graduação em Ciência da Computação do
Centro de Informática da Universidade Federal de Pernambuco.

“Be like water making its way through cracks. Do not be assertive, but adjust to the object, and you shall find a way around or through it. If nothing within you stays rigid, outward things will disclose themselves. Empty your mind, be formless. Shapeless, like water. If you put water into a cup, it becomes the cup. You put water into a bottle and it becomes the bottle. You put it in a teapot, it becomes the teapot. Now, water can flow or it can crash. Be water, my friend.”

- Bruce Lee

AGRADECIMENTOS

Para realização desse trabalho foram necessários muitos esforços e dedicação que só foram possíveis graças à grande vigília do meu Deus, que me acompanha nas horas fáceis e difíceis. Da mesma forma, devo muito agradecimento ao apoio e suporte oferecido pela minha mãe, pai, irmã, amigos e amigas.

Tão importante, também, foi a paciência e preocupação com as quais minha orientadora, Veronica Teichrieb, teve em inúmeros momentos de dificuldade, tirando minhas dúvidas e guiando minhas pesquisas em prol de um resultado conciso e satisfatório que pudesse justificar meu grau de mestre em Ciência da Computação pelo Centro de Informática - UFPE.

Agradeço a toda a equipe que forma o grupo de pesquisa Voxar Labs no Centro de Informática da UFPE, pelo auxílio e orientação nos momentos de dificuldade. Em especial, gostaria de agradecer à Edvar, Cledja, Daniel, Francisco e João Marcelo pela colaboração direta na construção deste trabalho.

Agradeço também aos grandes responsáveis pelos momentos de descontração e diversão que proporcionaram calma e tranquilidade para que eu me dedicasse a este trabalho, nas pessoas de Carlos Iglesias, Pietro Marchi, Marcelo Sobral e minha companheira Luciana Sampaio.

Agradeço principalmente aos meus pais Carlos Eduardo Padilha e Maria Valéria de Lima, pela construção emocional e moral da pessoa que sou hoje e, também, pela grande paciência e amor que tiveram comigo no meu dia-a-dia. E à minha incrível irmã Débora de Lima Padilha pelo amor, companheirismo e compreensão que sempre teve por mim.

RESUMO

Em Realidade Aumentada funcional o objetivo principal é ajudar os observadores com a execução de tarefas através de uma percepção melhorada da cena aumentada. Este trabalho aplica uma técnica de visualização bastante utilizada, conhecida como *Ghosting*, para melhorar a percepção de profundidade em cenas de Realidade Aumentada. Técnicas de visão computacional e processamento de imagens são usadas para extrair as características naturais em cenas reais para, em seguida, decidir a quantidade de transparência que será aplicada em cada pixel do objeto virtual, e assim, gerar o efeito de *ghosting* quando misturar o objeto virtual na cena real.

Um objeto em movimento em uma cena real chama bastante atenção do observador. Assim, espera-se que, quando um objeto em movimento passar na frente de um objeto aumentado, a informação visual do objeto real se mantenha, para que a visualização de informações naturais, e importantes da cena, não sejam perdidas. Por isso, a principal contribuição deste trabalho é a inclusão de uma técnica de detecção de movimento na etapa de análise de características da cena do *pipeline* da técnica de *Ghosting*. A avaliação qualitativa realizada com diversos usuários mostrou que os estudos de caso, em ambientes internos e externos, abordados neste trabalho, proporcionaram uma melhor percepção de profundidade da cena de Realidade Aumentada, pois preserva as informações mais relevantes para a correta percepção do ambiente pelo usuário.

Palavras-chave: *Ghosting*. Realidade Aumentada. Oclusão. Movimento. Transparência. Visualização. Mapa de Saliência.

ABSTRACT

Functional Augmented Reality realism focuses on helping users with tasks execution through an enhanced perception of the augmented scene. This work applies a common visualization technique, Ghosting, to improve depth perception in Augmented Reality scenes. Computer Vision and Image Processing techniques are used to extract natural features from a real scene, which will guide the assignment of transparency to each pixel of the virtual object, and provide the ghosting effect while blending the virtual object into the real scene.

A moving object in a real scene catches user attention. So, it's expected that natural and important visual information of the scene don't get occluded when the moving object passes over it. Because of that, the main contribution of this work is the inclusion of a motion detection technique to the scene feature analysis step, into the Ghosting technique pipeline. The qualitative evaluation showed that the user's studies of this work, in indoor and outdoor environments, using the proposed technique led to a better depth perception of the augmented scene, because it preserves the most relevant information for correct user's perception of the scene.

Keywords: Ghosting. Transparency. Occlusion. Motion. Saliency Map. Augmented Reality. Visualization.

LISTA DE FIGURAS

FIGURA 1 - EXEMPLOS DE CATEGORIAS DE PROBLEMAS QUE AFETAM A PERCEPÇÃO (KALKOFEN, ET AL., 2011).....	14
FIGURA 2 - TÉCNICAS PARA VISUALIZAÇÃO DE CENAS COM PARTES OCLUSAS. CUTAWAYS: INTERIOR DE UM MOTOR (A) (LI, ET AL., 2007); GHOSTING: MOTOR DENTRO DE UM CARRO (B) (KALKOFEN, ET AL., 2009); EXPLOSION DIAGRAMS: PARTES DE UM CARRO (C) (KALKOFEN, ET AL., 2011).....	16
FIGURA 3 - RESULTADOS OBTIDOS UTILIZANDO A TÉCNICA DE BANE E HÖELLERER (BANE, ET AL., 2004)...	20
FIGURA 4 - RESULTADOS DO EFEITO DE RAIOS-X OBTIDOS COM A TÉCNICA (A) AVERY ET AL. (SANDOR, ET AL., 2010) E (B) SANDOR ET AL. (AVERY, ET AL., 2009).....	21
FIGURA 5 - RESULTADOS DA TÉCNICA DE CHEN ET AL. (CHEN, ET AL., 2010).....	22
FIGURA 6 - TIPOS DE VISUALIZAÇÃO APRESENTADAS PARA OS USUÁRIOS: (A) DESENHO FINO, (B) DESENHO PREENCHIDO E (C) DESENHO FINO E PREENCHIDO, UTILIZANDO O DESIGN DO EXPERIMENTO (D). ADAPTADO DE (LIVINGSTON, ET AL., 2003).....	23
FIGURA 7 - RESULTADOS UTILIZANDO A TÉCNICA DE LU ET AL. (LU, ET AL., 2010).....	24
FIGURA 8 - RESULTADO OBTIDO UTILIZANDO A TÉCNICA DE ZOLLMANN ET AL. (ZOLLMANN, ET AL., 2010)...	24
FIGURA 9 - COMPARAÇÃO ENTRE OS RESULTADOS ENTRE OS RESULTADOS DOS TRABALHOS DE (A) ZOLLMANN ET AL. (ZOLLMANN, ET AL., 2010) E (B) PADILHA ET AL. (PADILHA, ET AL., 2013).....	25
FIGURA 10 - RESULTADO OBTIDO UTILIZANDO A TÉCNICA DE KALKOFEN ET AL. (KALKOFEN, ET AL., 2013). 26	
FIGURA 11 - CONFIGURAÇÃO DOS EXPERIMENTOS UTILIZADOS NO TRABALHO DE BERNING ET AL (BERNING, ET AL., 2014). OBJETO VIRTUAL VERDE DE MESMO TAMANHO DO REAL ROSA (A), CUBO COM MESMA PROFUNDIDADE DO OBJETO DE REFERENCIA (ROSA) E SOMBRA PROJETADA (B), CUBO COM MESMA PROFUNDIDADE E TEXTURA (C), OBJETO VIRTUAL (NOTA DE 10 EUROS) COM TAMANHO CONHECIDO PELO USUÁRIO (D) E ESFERA SEM INFORMAÇÕES DE PROFUNDIDADE.	27
FIGURA 12 - RESULTADO OBTIDO UTILIZANDO A TÉCNICA DE WITHER ET AL. (WITHER, ET AL., 2008). NO CANTO INFERIOR DIREITO O RESULTADO DA CENA DE RA COM ESTÁTUA VIRTUAL AUMENTADA.	28
FIGURA 13 - PIPELINE DA TÉCNICA DE GHOSTING UTILIZADO NESTE TRABALHO.....	32
FIGURA 14 - HISTOGRAMA DE COR REFERENTE À IMAGEM (A), E EM (B) A IMAGEM RESULTANTE APÓS APROXIMAÇÕES. IMAGEM ADAPTADA DE (CHENG, ET AL., 2011).....	33
FIGURA 15 - GERAÇÃO DO MAPA DE SALIÊNCIA PARA PERCEPÇÃO DE PROFUNDIDADE. (A) CENA ORIGINAL, (B) MAPA DE SALIÊNCIA HC RESULTANTE, (C) SOBREPOSIÇÃO SIMPLES DO OBJETO AUMENTADO, (D) CENA DE RA FINAL CONSIDERANDO OBJETO SALIENTE.	33
FIGURA 16 - MANTENDO A SILHUETA DOS OBJETOS REAIS ATRAVÉS DAS BORDAS. (A) SOBREPOSIÇÃO BRUTA DOS CANOS VERDES NA CENA ORIGINAL, (B) CENA DE RA FINAL COM A DETECÇÃO DE BORDAS.....	34
FIGURA 17 - TELA DE CONFIGURAÇÃO DOS PARÂMETROS DE THRESHOLD BAIXO E ALTO PARA A TÉCNICA DE DETECÇÃO DE BORDAS DE CANNY.....	35
FIGURA 18 - EXEMPLOS DE CLASSIFICAÇÃO DOS PADRÕES DE MOVIMENTO. NORMAL MOVING É APRESENTADO NA IMAGEM DO CANTO SUPERIOR DIREITO, TEMPORAL ALIASED MOVING NA IMAGEM CENTRAL NO LADO DIREITO E HYSTERETIC MOVING NA IMAGEM DO CANTO INFERIOR DIREITO (OH, ET AL., 2013).	37

FIGURA 19 - ESTUDO DE CASO DE ARQUITETURA, EM AMBIENTE <i>INDOOR</i>. QUADRO DA CENA REAL (A), CANOS VERDES VIRTUAIS A SEREM AUMENTADOS NA CENA (B), CENA DE RA UTILIZANDO SOBREPOSIÇÃO DO OBJETO VIRTUAL SEM TRATAMENTO DE OCLUSÃO (C) E CENA DE RA UTILIZANDO A TÉCNICA PROPOSTA NESTE TRABALHO (D).	42
FIGURA 20 - QUADROS CONSECUTIVOS DA SEQUÊNCIA DE VÍDEO DO CENÁRIO DA RESIDÊNCIA. PAREDE BRANCA E CÔMODA MARROM, FIXOS NO AMBIENTE, E PESSOA SE LOCOMOVENDO SÃO OBJETOS REAIS, E, CANOS VERDES SÃO OBJETOS AUMENTADOS UTILIZANDO A TÉCNICA PROPOSTA NESTE TRABALHO.	43
FIGURA 21 - COMPARAÇÃO DE RESULTADOS ENTRE DIFERENTES TÉCNICAS PARA CENA DE RA COM ALTO CONTRASTE DE COR ENTRE OBJETO EM MOVIMENTO E <i>BACKGROUND</i>. SOBREPOSIÇÃO DO OBJETO VIRTUAL SEM TRATAMENTO DE OCLUSÃO (A), RESULTADO DA TÉCNICA DE <i>GHOSTING</i> APRESENTADA POR (ZOLLMANN, ET AL., 2010) (B) E RESULTADO DA TÉCNICA PROPOSTA NESTE TRABALHO (C).	44
FIGURA 22 - COMPARAÇÃO DE RESULTADOS ENTRE DIFERENTES TÉCNICAS PARA CENA DE RA COM BAIXO CONTRASTE DE COR ENTRE OBJETO EM MOVIMENTO E <i>BACKGROUND</i>. RESULTADO DA TÉCNICA DE <i>GHOSTING</i> APRESENTADA POR (PADILHA, ET AL., 2013) (A) E RESULTADO DA TÉCNICA PROPOSTA NESTE TRABALHO (B).	45
FIGURA 23 - CENÁRIO <i>INDOOR</i> COM BAIXO E ALTO CONTRASTE DE COR ENTRE ALGUNS OBJETOS E O <i>BACKGROUND</i>. QUADRO DA CENA REAL (A), CANOS VERDES VIRTUAIS A SEREM AUMENTADOS NA CENA (B), RESULTADO DA DETECÇÃO DE MOVIMENTO (C), MAPA DE <i>GHOSTING</i> RESULTANTE (D) E CENA DE RA UTILIZANDO A TÉCNICA PROPOSTA NESTE TRABALHO (E).	47
FIGURA 24 - COMPARAÇÃO DE RESULTADOS ENTRE DIFERENTES TÉCNICAS PARA CENA DE RA COM BAIXO CONTRASTE DE COR ENTRE OBJETO EM MOVIMENTO E <i>BACKGROUND</i>. RESULTADO DA TÉCNICA DE <i>GHOSTING</i> APRESENTADA POR (PADILHA, ET AL., 2013) (A) E RESULTADO DA TÉCNICA PROPOSTA NESTE TRABALHO (B).	48
FIGURA 25 - ESTUDO DE CASO DE CONSTRUÇÃO CIVIL, EM AMBIENTE <i>OUTDOOR</i>. QUADRO DA CENA REAL (A), CANOS VERDES VIRTUAIS A SEREM AUMENTADOS NA CENA (B), RESULTADO DA DETECÇÃO DE MOVIMENTO (C) E CENA DE RA UTILIZANDO A TÉCNICA PROPOSTA NESTE TRABALHO (D).	49
FIGURA 26 - QUADROS CONSECUTIVOS DE RUA COM VEÍCULOS TRAFEGANDO (OBJETOS REAIS) E CANOS VERDES (OBJETOS AUMENTADOS) UTILIZANDO A TÉCNICA PROPOSTA NESTE TRABALHO.	50
FIGURA 27 - ESTUDO DE CASO DE CONSTRUÇÃO CIVIL, EM AMBIENTE <i>OUTDOOR</i>. QUADRO DA CENA REAL (A), CANOS VIRTUAIS A SEREM AUMENTADOS NA CENA (B), RESULTADO DA TÉCNICA DE <i>GHOSTING</i> APRESENTADA POR (ZOLLMANN, ET AL., 2010) (C) E RESULTADO DA TÉCNICA PROPOSTA NESTE TRABALHO (D).	50
FIGURA 28 - ESTUDO DE CASO DE <i>SAFE DRIVING</i>, EM AMBIENTE <i>OUTDOOR</i>. QUADRO DA CENA REAL (A), OBJETO VIRTUAL AUMENTADO (B), CENA DE RA UTILIZANDO SOBREPOSIÇÃO DO OBJETO VIRTUAL SEM TRATAMENTO DE OCLUSÃO (C) E CENA DE RA UTILIZANDO A TÉCNICA PROPOSTA NESTE TRABALHO (D).	52
FIGURA 29 - QUADROS CONSECUTIVOS DA SEQUÊNCIA DE VÍDEO DO CENÁRIO DA SITUAÇÃO DE TRÂNSITO. A CENA COMPLETA É REAL, COM EXCEÇÃO DA SETA INDICANDO DIREÇÃO DO TRAJETO E INDICADOR DE DISTÂNCIA, QUE SÃO AUMENTADOS UTILIZANDO A TÉCNICA PROPOSTA NESTE TRABALHO.	53
FIGURA 30 - QUADROS CONSECUTIVOS DA SEQUÊNCIA DE VÍDEO DA CENA DE RA COM SOBREPOSIÇÃO SEM TRATAMENTO DE OCLUSÃO DO OBJETO VIRTUAL (SETA INDICANDO DIREÇÃO DO TRAJETO E INDICAÇÃO DA	

DISTÂNCIA).....	54
FIGURA 31 - QUADROS CONSECUTIVOS DA SEQUÊNCIA DE VÍDEO DA CENA DE RA COM UM CARRO, COM POUCO CONTRASTE DE COR ENTRE O OBJETO EM MOVIMENTO E O BACKGROUND, PASSANDO. RESULTADO DA TÉCNICA PROPOSTA NESTE TRABALHO.....	55
FIGURA 32 - COMPARAÇÃO DE RESULTADOS ENTRE A TÉCNICA DE <i>GHOSTING</i> APRESENTADA POR (ZOLLMANN, ET AL., 2010) (A), POR (PADILHA, ET AL., 2013)(B) E A TÉCNICA PROPOSTA NESTE TRABALHO (C).....	56
FIGURA 33 - CENA DE RA COM UM CARRO EM MOVIMENTO. CENA REAL (A), OBJETO VIRTUAL AUMENTADO (B), CENA DE RA UTILIZANDO SOBREPOSIÇÃO DO OBJETO VIRTUAL SEM TRATAMENTO DE OCLUSÃO (C) E CENA DE RA UTILIZANDO A TÉCNICA PROPOSTA NESTE TRABALHO (D).....	57
FIGURA 34 - CENA DE RA COM CARRO E UMA BOLA EM MOVIMENTO. RESULTADO DA TÉCNICA DE <i>GHOSTING</i> APRESENTADA POR (ZOLLMANN, ET AL., 2010): CARRO 1 (A), BOLA (D), CARRO 2 (G), CARRO 3 (J); RESULTADOS DA TÉCNICA DE <i>GHOSTING</i> APRESENTADA POR (PADILHA, ET AL., 2013): CARRO 1 (B), BOLA (E), CARRO 2 (H), CARRO 3 (K); E, RESULTADO DA TÉCNICA PROPOSTA NESTE TRABALHO: CARRO 1 (C), BOLA (F), CARRO 2 (I), CARRO 3 (L).....	58
FIGURA 35 - QUADROS DA CENA DE VÍDEO <i>INDOOR</i> PARA APLICAÇÃO DE ARQUITETURA E CONSTRUÇÃO CIVIL COM OBJETOS NÃO SALIENTES E OBJETOS SALIENTES ESTÁTICOS E EM MOVIMENTO UTILIZADA NO QUESTIONÁRIO DE AVALIAÇÃO QUALITATIVA. A ORDEM APRESENTADA NO QUESTIONÁRIO CORRESPONDE À ORDEM APRESENTADA NA FIGURA. RESULTADO OBTIDO UTILIZANDO (ZOLLMANN, ET AL., 2010) (A), RESULTADO OBTIDO UTILIZANDO (PADILHA, ET AL., 2013) (B), RESULTADO OBTIDO UTILIZANDO A TÉCNICA PROPOSTA NESTE TRABALHO (C) E RESULTADO OBTIDO UTILIZANDO RA CONVENCIONAL (D).....	62
FIGURA 36 - QUADROS DA CENA DE VÍDEO <i>OUTDOOR</i> PARA APLICAÇÃO DE ARQUITETURA E CONSTRUÇÃO CIVIL COM OBJETOS SALIENTES EM MOVIMENTO, COM DIFERENTES TAMANHOS E VELOCIDADES, UTILIZADA NO QUESTIONÁRIO DE AVALIAÇÃO QUALITATIVA. A ORDEM APRESENTADA NO QUESTIONÁRIO CORRESPONDE COM A ORDEM APRESENTADA NA FIGURA. RESULTADO OBTIDO UTILIZANDO (ZOLLMANN, ET AL., 2010) (A), RESULTADO OBTIDO UTILIZANDO (PADILHA, ET AL., 2013) (B), RESULTADO OBTIDO UTILIZANDO A TÉCNICA PROPOSTA NESTE TRABALHO (C) E RESULTADO OBTIDO UTILIZANDO RA CONVENCIONAL (D).....	63
FIGURA 37 - QUADROS DA CENA DE VÍDEO <i>OUTDOOR</i> PARA APLICAÇÃO DE <i>SAFE DRIVING</i> COM OBJETOS NÃO SALIENTES E OBJETOS SALIENTES ESTÁTICOS E EM MOVIMENTO UTILIZADA NO QUESTIONÁRIO DE AVALIAÇÃO QUALITATIVA. A ORDEM APRESENTADA NO QUESTIONÁRIO CORRESPONDE À ORDEM APRESENTADA NA FIGURA. RESULTADO OBTIDO UTILIZANDO A TÉCNICA PROPOSTA NESTE TRABALHO (A), RESULTADO OBTIDO UTILIZANDO (PADILHA, ET AL., 2013) (B), RESULTADO OBTIDO UTILIZANDO (ZOLLMANN, ET AL., 2010) (C), RESULTADO OBTIDO UTILIZANDO RA CONVENCIONAL (D).....	64
FIGURA 38 - QUADROS DA CENA DE VÍDEO <i>OUTDOOR</i> PARA APLICAÇÃO DE <i>SAFE DRIVING</i> COM OBJETOS NÃO SALIENTES E OBJETOS SALIENTES ESTÁTICOS E EM MOVIMENTO UTILIZADA NO QUESTIONÁRIO DE AVALIAÇÃO QUALITATIVA. A ORDEM APRESENTADA NO QUESTIONÁRIO CORRESPONDE COM A ORDEM APRESENTADA NA FIGURA. RESULTADO OBTIDO UTILIZANDO RA CONVENCIONAL (A), RESULTADO OBTIDO UTILIZANDO (ZOLLMANN, ET AL., 2010) (B), RESULTADO OBTIDO UTILIZANDO A TÉCNICA PROPOSTA NESTE TRABALHO (C), RESULTADO OBTIDO UTILIZANDO (PADILHA, ET AL., 2013) (D).....	65

FIGURA 39 - GRÁFICO COM MÉDIA DAS AVALIAÇÕES DE CADA TÉCNICA PARA CADA CENA DE VÍDEO APRESENTADA NA FIGURA 35, NA FIGURA 36, NA FIGURA 37, E NA FIGURA 38. AVALIAÇÃO PARA TÉCNICA PROPOSTA NESTE TRABALHO EM AZUL, AVALIAÇÃO PARA TÉCNICA PROPOSTA POR ZOLLMANN ET AL.	66
FIGURA 40 - FALHA NA APROXIMAÇÃO DA ÁREA CLASSIFICADA COMO EM MOVIMENTO PARA A SILHUETA DO OBJETO. CENA DA APLICAÇÃO DE SAFE DRIVING À ESQUERDA E CENA DA APLICAÇÃO DE ARQUITETURA E CONSTRUÇÃO CIVIL À DIREITA.	67

LISTA DE ABREVIações E ACRÔNIMOS

AP	<i>Aliased Pattern</i>
FT_SEM	<i>Técnica proposta em trabalho concorrente Zollmann et al. 2010</i>
HC	<i>Histogram Contrast Saliency</i>
HC_COM	<i>Técnica proposta neste trabalho</i>
HC_SEM	<i>Técnica proposta em nosso trabalho anterior</i>
HP	<i>Hysteretic Pattern</i>
NP	<i>Normal Pattern</i>
OpenCV	<i>Open Computer Vision Library</i>
RA	Realidade Aumentada
SO	<i>RA convencional</i>
TAM	<i>Tonal Art Map</i>

SUMÁRIO

1 INTRODUÇÃO	14
1.1 OBJETIVOS	17
1.2 ESTRUTURA DO DOCUMENTO	18
2 TRATAMENTO DE OCLUSÃO EM RA	20
3 TÉCNICA DE <i>GHOSTING</i> COM DETECÇÃO DE MOVIMENTO	29
3.1 <i>GHOSTING</i>	30
3.2 DETECÇÃO DE MOVIMENTOS	35
4 ESTUDOS DE CASO	41
4.1 ARQUITETURA E CONSTRUÇÃO CIVIL	41
4.2 <i>SAFE DRIVING</i>	51
4.3 AVALIAÇÃO QUALITATIVA	59
5 CONCLUSÃO	68
5.1 TRABALHOS FUTUROS	69
5.2 CONTRIBUIÇÕES E PUBLICAÇÕES DIRETAMENTE RELACIONADAS À ESTE TRABALHO	69

1 INTRODUÇÃO

A percepção humana é definida por (Cutting, 2003), como o complexo reconhecimento e interpretação dos estímulos sensoriais. Nós, seres humanos, baseamos nossas interpretações e decisões nos diferentes tipos de informações obtidas de cada modalidade sensorial; nesta pesquisa o foco é em percepção visual (Kruijff, et al., 2010). No contexto de aplicações de Realidade Aumentada (RA), é comum existirem problemas de percepção, pela sua natureza complexa que envolve a sobreposição de informações virtuais na cena real. Esta sobreposição, muitas vezes, ocorre de uma forma que informações visuais necessárias para proporcionar um reconhecimento e interpretação adequados da cena não estão disponíveis para o usuário. Cenas aumentadas incorretas, do ponto de vista da percepção, são frequentemente o resultado de dicas conflituosas fornecidas por elas mesmas (Kruijff, et al., 2010).

Uma série de problemas é responsável por prejudicar a percepção em cenas de RA. Em (Kruijff, et al., 2010), foram listados e exemplificados vários desses problemas. Segundo (Kruijff, et al., 2010), existem três categorias de problemas que afetam a percepção: distorção e abstração da cena, distorção de profundidade e ordenação de objetos, e visibilidade. Essas categorias de problemas são bastante frequentes na visualização de cenas de RA (Kalkofen, et al., 2011) e alguns exemplos podem ser vistos na Figura 1.

Figura 1 - Exemplos de categorias de problemas que afetam a percepção (Kalkofen, et al., 2011).



Na Figura 1-a, é possível visualizar um exemplo das categorias de problemas

de distorção e abstração da cena e de distorção de profundidade e ordenação de objetos, pois o boneco de Lego com as descrições STB, que foi aumentado na cena, possui um tamanho menor em relação aos bonecos reais, indicando que estaria localizado em um plano de maior profundidade. Entretanto, ele é apresentado sobre o boneco real e, por isso, a relação de profundidade e a abstração da cena são prejudicadas.

Na Figura 1-b, não existe a categoria de problemas associadas à distorção de profundidade e ordenação, pois o boneco virtual aparenta estar em um plano de maior profundidade em relação aos outros, diferente da Figura 1-a. Entretanto, estão presentes as categorias de problemas de distorção e abstração da cena, pois a forma de disposição do objeto virtual dificulta o reconhecimento do objeto real (boneco de vermelho na Figura 1-b), e a categoria de problemas de visibilidade, pois os detalhes dos objetos virtuais e reais não podem ser claramente percebidos, principalmente, devido à forma como foram apresentados, dificultando assim, a visualização da cena de RA.

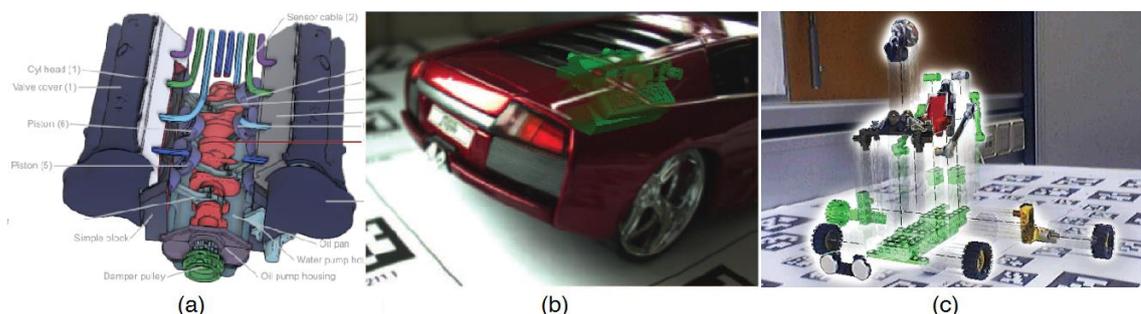
Na Figura 1-c, tem-se a categoria de problemas de visibilidade, pois o objeto aumentado (rodas traseiras do carro) e o objeto real (carro) estão dispostos com baixo contraste de cor, dificultando a visualização da cena de RA.

Neste trabalho, serão abordadas as categorias de problemas de distorção de profundidade e ordenação, e de visibilidade, segundo classificação de (Kruijff, et al., 2010), e exemplificadas nas Figura 1-a e Figura 1-c. Em especial, este trabalho terá como foco principal os problemas de oclusão e relação de profundidade entre objetos virtuais e reais na cena de RA, levando em consideração que objetos naturalmente salientes à visão humana devem ser considerados durante a construção da cena de RA. Dessa forma, a visualização da relação de profundidade entre o objeto virtual e os objetos reais importantes da cena ficará mais agradável para o observador. Os estudos de caso abordados neste trabalho foram realizados em ambientes *indoor* e *outdoor* e possuem situações de oclusão entre objeto virtual e real na cena.

Para atacar o problema de oclusão em cenas de RA, as técnicas mais utilizadas na literatura são: *Cutaways* (Feiner, et al., 1992), *Ghosting* (Feiner, et al., 1992) e *Explosion Diagrams* (Kalkofen, et al., 2009). As três foram consideradas neste trabalho por buscarem melhorar a visualização da cena de RA através do tratamento de oclusão entre objetos reais e virtuais.

A técnica de *Cutaways* utiliza uma geometria de corte nos volumes dos objetos (Figura 2-a) para introduzir pistas sobre as profundidades e relacionamentos dos objetos da cena (Feiner, et al., 1992). A técnica de *Ghosting* é a que menos modifica a cena original, pois exibe as imagens real e virtual juntas, entretanto submetidas a um grau de transparência que é definido para cada *pixel* da cena (Figura 2-b). O termo usado em visualização para indicar a contribuição ou importância de cada *pixel* para a cena é denominado função de transferência, ou mapa de *ghosting* (Zollmann, et al., 2010). E, por fim, a técnica de *Explosion Diagrams* consiste em separar as partes dos objetos a fim de permitir a visualização detalhada da maior quantidade de partes possível (Figura 2-c). *Explosion Diagrams* é, principalmente, utilizada em situações com cenas complexas, densamente desordenadas e nas quais se deseja visualizar todas as informações de objetos relevantes em um espaço pequeno, como em manuais técnicos (Kalkofen, et al., 2009).

Figura 2 - Técnicas para visualização de cenas com partes oclusas. Cutaways: interior de um motor (a) (Li, et al., 2007); Ghosting: motor dentro de um carro (b) (Kalkofen, et al., 2009); Explosion Diagrams: partes de um carro (c) (Kalkofen, et al., 2011).



A técnica de visualização escolhida para este trabalho foi a *Ghosting* (Feiner, et al., 1992) por necessitar de conhecimento mínimo sobre a cena real para inserir o

objeto virtual e manter bastante fidelidade ao contexto da cena original, já que transmite mais informação sobre a cena com o menor grau de modificação da cena original.

O processo humano de atenção visual, segundo (Treisman, et al., 1992), inicia-se com o cérebro identificando características visuais da cena a fim de determinar quais regiões vão necessitar que o observador preste mais atenção de forma inconsciente. Essas regiões da cena que possuem maiores chances de capturar a atenção do observador são chamadas de regiões de saliência (Treisman, et al., 1992). As características visuais, analisadas inconscientemente durante o processo de atenção, possuem propriedades de contraste de cor, forma, intensidade de cor e movimento (Treisman, et al., 1992). Dessa forma, sistemas computacionais de RA procuram representar esse processo de atenção humana identificando informações com essas propriedades a fim de identificar quais elementos da cena real chamam, naturalmente, a atenção do observador. Esse conceito de regiões de saliência é bastante importante para melhorar a visualização de uma cena de RA, já que essas regiões possuem, naturalmente, informações importantes para o observador e, portanto, não devem ser totalmente oclusas pelo objeto virtual. Neste trabalho, serão consideradas as propriedades de contraste de cor, forma, intensidade de cor e movimento para composição da técnica de *Ghosting*.

1.1 OBJETIVOS

Com este trabalho, busca-se melhorar a visualização de cenas de RA que possuem situações de oclusão, levando em consideração as características da cena real que são analisadas pelo cérebro humano durante o processo de atenção visual. Este objetivo é alcançado utilizando técnicas de visão computacional e processamento de imagens para identificar as características da cena que são naturalmente importantes para o observador e assim preservá-las na cena final de RA.

A técnica de *Ghosting* proposta neste trabalho busca realizar uma análise mais precisa da propriedade de movimento dos objetos reais da cena. Este trabalho utiliza como premissa que objetos naturalmente importantes (salientes) da cena devem ser considerados no tratamento de oclusão com os objetos aumentados, pois, assim, a visualização da cena ficará mais agradável para o observador. Com base em trabalhos anteriores, considera-se que um objeto que naturalmente chama a atenção do observador não pode ser ocluído pelo objeto virtual, caso contrário informações importantes de objetos reais da cena estariam sendo omitidas após a constituição da cena de RA, e isso é considerado prejudicial à visualização da cena. Além disso, considera-se que, ao manter as bordas dos objetos reais da cena, as silhuetas são preservadas e assim o contexto da cena é mantido (Zollmann, et al., 2010).

Os estudos de caso apresentados (na área de construção civil e *safe driving*) ocorrem em ambientes *indoor* e *outdoor*, e apresentam situações de oclusão entre objetos reais e virtuais.

Trabalhos anteriores não utilizam uma técnica de detecção de movimentos precisa, e por isso, em algumas situações desses estudos de caso, têm dificuldades em preservar os objetos naturalmente salientes, como originalmente se propõem.

Devido à subjetividade envolvida com a área de visualização de cenas de RA, o método de avaliação utilizado foi essencialmente qualitativo e submetido a sujeitos de idades, *background* técnico e gêneros distintos.

1.2 ESTRUTURA DO DOCUMENTO

Esta dissertação foi organizada de forma a introduzir os conceitos necessários para melhorar a visualização de uma cena de RA tratando oclusões com a técnica de *Ghosting*. As áreas de estudo e os problemas associados à visualização de cenas de RA foram abordados para contextualizar a necessidade de se traduzir características da percepção humana em sistemas computacionais que utilizam RA. A dissertação está organizada como segue.

No segundo capítulo, são apresentadas e discutidas as técnicas utilizadas na área de visualização e na área de RA para tratamento de oclusões.

O terceiro capítulo define e explica os algoritmos de visão computacional e processamento de imagens que foram utilizados na composição da técnica de *Ghosting* proposta por este trabalho. Além disso, descreve os motivos de cada uma das técnicas terem sido escolhidas e seus efeitos quando utilizadas na composição da cena de RA.

O quarto capítulo apresenta e discute os estudos de caso utilizados para avaliar qualitativamente a técnica proposta neste trabalho.

O quinto capítulo apresenta as conclusões desta dissertação e aborda futuros trabalhos.

2 TRATAMENTO DE OCLUSÃO EM RA

Existem na literatura diversos trabalhos desenvolvidos com o objetivo de tratar problemas de oclusão em RA e, portanto, melhorar a percepção da cena aumentada pelo usuário. Comumente estes trabalhos focam no uso de técnicas de visualização baseadas em imagens RGB.

Bane e Höellerer (Bane, et al., 2004) introduziram uma técnica que permite a interação do usuário no processo de decisão de quais estruturas oclusas ele deseja visualizar em cada momento. Os resultados de (Bane, et al., 2004), como pode ser visto na Figura 3, são apresentados com total opacidade do objeto virtual sobre a cena real, pois a renderização da cena aumentada acontece sem nenhuma transparência. Com isso, a visualização de informações dos objetos reais da cena e do contexto é prejudicada. Em (Bane, et al., 2004), os autores mantiveram as bordas das paredes e transferiram o poder de decisão para o usuário, permitindo a ele interagir com a ferramenta e ajudar no processo de criação da cena de RA, mas nos locais onde o objeto virtual é adicionado não há preocupação em melhorar a visualização através da inserção de dicas de profundidade.

Figura 3 - Resultados obtidos utilizando a técnica de Bane e Höellerer (Bane, et al., 2004).



O trabalho de (Avery, et al., 2009), utilizou a técnica de *Ghosting* para produzir o efeito de raio-x na visualização de cenas de RA. Avery *et al.* (Avery, et al.,

2009) mostraram que o efeito de raio-x consiste em permitir que o observador visualize cenas as quais, originalmente, estavam oclusas por paredes e outros objetos. O efeito foi obtido misturando a cena visível com a cena oclusa (não está visível naturalmente, por conta de uma parede ou outro objeto ocluser) e mantendo as bordas da cena visíveis (Figura 4-b). Ao manter as bordas, as silhuetas dos objetos visíveis ficam evidentes mesmo após concluir a geração da cena de RA. Isso permitiu que o observador visualizasse a cena oclusa em conjunto com a visível, e que o contexto da cena visível fosse minimamente preservado, mantendo assim informações sobre os objetos reais na cena final. Essa abordagem melhorou a visualização da cena de RA, mas, como apenas as bordas são mantidas, muita informação sobre os objetos reais foi perdida. Além disso, a técnica foi voltada apenas para o contexto de visualização no tratamento de oclusão entre cenas reais aplicando efeito de raios-x, não sendo voltada para cenas de RA com objetos virtuais como é o caso deste trabalho.

Figura 4 - Resultados do efeito de raios-x obtidos com a técnica (a) Avery et al. (Sandor, et al., 2010) e (b) Sandor et al. (Avery, et al., 2009).



Para melhorar a análise da cena real proposta em seu trabalho anterior (Avery, et al., 2009), Sandor *et al.* (Sandor, et al., 2010), aplicaram uma técnica de mapa de saliência, bastante tradicional na área de visualização, proposta em (Walther, 2006), durante a etapa de análise da cena para geração da cena de RA. As técnicas de mapa de saliência permitem identificar, em cenas reais, quais os objetos que, naturalmente, ao observar a cena, possuem maior destaque para o

cérebro humano. Dessa forma, o trabalho (Sandor, et al., 2010) apresenta resultados da técnica de *Ghosting* que possibilitam uma visualização de raios-x mais contextualizada com a cena real (Figura 4-a). O resultado da visualização é mais agradável para o observador, pois o algoritmo mantém, na cena final, as bordas e os objetos que naturalmente são salientes para o ser humano. Portanto, as informações importantes da cena original são mantidas mesmo após a mistura com a parte não visível da mesma. Esta técnica restringe-se apenas ao contexto de visualização e permite efeitos de raios-x nas cenas de RA.

Chen *et al.* (Chen, et al., 2010) trabalharam com camadas simples de oclusão. Utilizaram a abordagem de manter as bordas dos objetos reais a fim de não perder as silhuetas na cena aumentada resultante, vide Figura 5. Entretanto, a técnica exige vários parâmetros do usuário e os autores deixam claro que conhecem bastante a geometria do ambiente real a fim de não necessitarem de uma reconstrução 3D. Essa necessidade não é exigida pela técnica apresentada no presente trabalho. Além disso, não consideraram o contraste entre os objetos durante a composição da cena, o que ajuda a melhorar a visualização de cenas de RA, como mostrou (Itti, 2004).

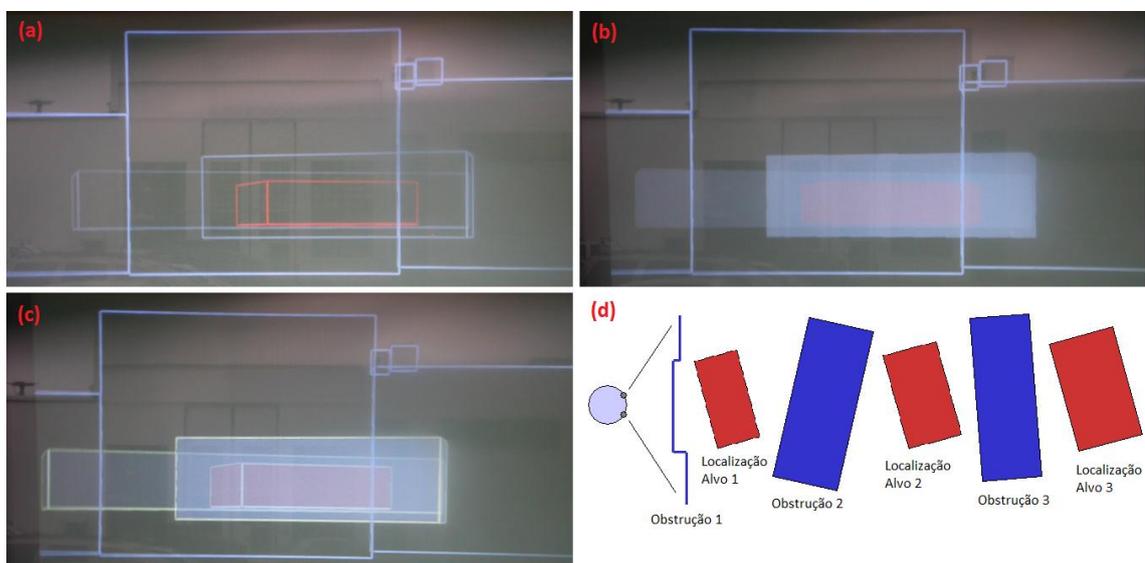
Figura 5 - Resultados da técnica de Chen et al. (Chen, et al., 2010).



Livingston *et al.* (Livingston, et al., 2003) desenvolveram uma técnica que trata múltiplas camadas de oclusão em cenas de RA. Sua principal contribuição foi ter mostrado que opacidade e transparência são discriminantes-chave na apresentação de objetos virtuais oclusos, pois transmite mais informação sobre a cena com o menor grau de modificação da cena original. Chegaram a esta conclusão após avaliarem cerca de 100 modelos, como os apresentados na Figura

6-a, Figura 6 -b e Figura 6-c, com variações no tamanho das linhas, estilo da linha, número de níveis de oclusão, sombreado e cor com um conjunto de indivíduos. Para esta dissertação, essa conclusão de Livingston *et al.* (Livingston, et al., 2003) foi muito importante, já que essa foi uma das premissas para a escolha da técnica de *Ghosting* e embasa nossos esforços na detecção de características naturais da cena a fim de construir um algoritmo para tratar problemas de oclusão e melhorar a visualização de cenas de RA.

Figura 6 - Tipos de visualização apresentadas para os usuários: (a) desenho fino, (b) desenho preenchido e (c) desenho fino e preenchido, utilizando o design do experimento (d). Adaptado de (Livingston, et al., 2003).



Lu *et al.* (Lu, et al., 2010) focaram em resolver problemas complexos de luminosidade em cenas de RA (ambientes externos) que possuem situações de oclusão. O trabalho utiliza segmentação de estruturas de *foreground* para sobrepô-las ao objeto virtual, vide Figura 7. Além disso, a técnica realiza detecção de sombras, o que melhora a relação de profundidade entre os objetos, como visto na Figura 7. Entretanto, a técnica falha em situações onde o observador não pode perder informações do objeto virtual, já que estas informações ficam sempre oclusas pelo objeto real, e não realiza detecção de bordas para manter as silhuetas dos objetos.

Figura 7 - Resultados utilizando a técnica de Lu et al. (Lu, et al., 2010).



O efeito de *ghosting* proporcionado pelo trabalho de Zollmann *et al.* (Zollmann, et al., 2010) reuniu técnicas de processamento de imagens e visão computacional clássicas, como detecção de borda (Canny, 1986) e geração de mapa de saliência (Achanta, et al., 2009), para analisar mais características, como forma e contraste de cor, da cena e decidir quais dicas de profundidade serão inseridas, vide Figura 8. O *pipeline* proposto por essa técnica é mais completo que os anteriores por tratar inclusive texturas e permitir a interação do usuário. A técnica não analisa as propriedades de movimento e por isso em situações com baixo contraste de cor entre os objetos salientes e o *background* ela não é capaz de melhorar a visualização da cena de RA.

Figura 8 - Resultado obtido utilizando a técnica de Zollmann et al. (Zollmann, et al., 2010).



Os resultados obtidos por Zollmann *et al.* (Zollmann, et al., 2010) foram

melhorados em um trabalho desenvolvido pelo autor dessa dissertação de mestrado (Padilha, et al., 2013), vide Figura 9. Nele foi proposta a utilização da técnica de contraste global apresentada em (Cheng, et al., 2011) para obtenção do mapa de saliência. Cheng *et al.* (Cheng, et al., 2011) mostraram a superioridade da sua técnica comparando seus resultados com técnicas clássicas de geração de mapa de saliência, inclusive com a técnica de Achanta *et al.* (Achanta, et al., 2009), utilizada no trabalho de Zollmann *et al.* (Zollmann, et al., 2010). Os resultados do trabalho (Padilha, et al., 2013) foram obtidos sem requisitos temporais.

Figura 9 - Comparação entre os resultados entre os resultados dos trabalhos de (a) Zollmann et al. (Zollmann, et al., 2010) e (b) Padilha et al. (Padilha, et al., 2013).



Kalkofen *et al.* (Kalkofen, et al., 2013) procuraram melhorar a visualização das cenas aumentadas utilizando a técnica de *Ghosting* proposta em (Zollmann, et al., 2010), e aumentando o contraste de iluminação dos objetos virtuais. Dessa forma, o contraste entre objeto real e virtual ficou mais evidente. Para obter este efeito, subtraíram os mapas de saliência, antes e depois da inserção do objeto virtual, o que permitiu identificar áreas salientes que deixaram de ser salientes, após a sobreposição do conteúdo aumentado na cena real, vide Figura 10. Com este trabalho, buscam apenas resolver o problema de visibilidade da cena de RA. Apenas as propriedades de saliência são consideradas.

O trabalho de Berning *et al.* (Berning, et al., 2014) reforça a importância do estudo da linha de pesquisa escolhida nesta dissertação. Eles apresentaram um estudo avaliando características de profundidade na observação de cenas de RA

utilizando *displays* estereoscópicos. O experimento constituiu em apresentar, a 20 participantes, uma série de objetos virtuais com as características apresentadas na Figura 11 como: forma equivalente (Figura 11-a), sombra (Figura 11-b), textura (Figura 11-c), tamanho conhecido (Figura 11-d) e até sem nenhuma pista de profundidade (Figura 11-e), com o objetivo de avaliar se o display estereoscópico ajuda o usuário a entender melhor a relação espacial entre os objetos virtual e real e comparar o efeito de algumas pistas de profundidade. Com essa avaliação qualitativa, eles concluíram que aplicações de RA devem incluir diferentes formas de indicadores de profundidade para facilitar a percepção de profundidade da cena por parte dos observadores. Apesar de o resultado ser focado em displays estereoscópicos e de as análises terem sido feitas apenas com relação ao objeto virtual, a conclusão de Berning *et al.* (Berning, et al., 2014) está completamente alinhada com as premissas utilizadas no presente trabalho de que várias características e pistas de profundidade devem ser inseridas na cena de RA para melhorar a percepção de profundidade entre objetos reais e virtuais.

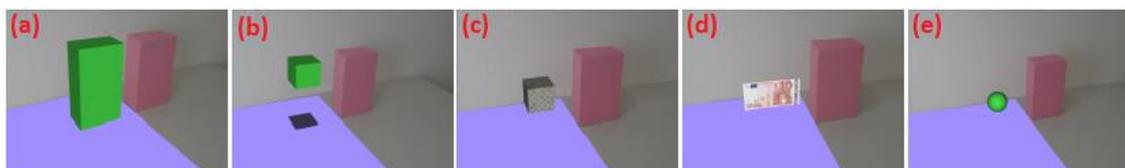
Figura 10 - Resultado obtido utilizando a técnica de Kalkofen et al. (Kalkofen, et al., 2013).



O trabalho de Berning *et al.* (Berning, et al., 2014) reforça a importância do estudo da linha de pesquisa escolhida nesta dissertação. Eles apresentaram um estudo avaliando características de profundidade na observação de cenas de RA utilizando *displays* estereoscópicos. O experimento constituiu em apresentar, a 20

participantes, uma série de objetos virtuais com as características apresentadas na Figura 11 como: forma equivalente (Figura 11-a), sombra (Figura 11-b), textura (Figura 11-c), tamanho conhecido (Figura 11-d) e até sem nenhuma pista de profundidade (Figura 11-e), com o objetivo de avaliar se o display estereoscópico ajuda o usuário a entender melhor a relação espacial entre os objetos virtual e real e comparar o efeito de algumas pistas de profundidade. Com essa avaliação qualitativa, eles concluíram que aplicações de RA devem incluir diferentes formas de indicadores de profundidade para facilitar a percepção de profundidade da cena por parte dos observadores. Apesar de o resultado ser focado em displays estereoscópicos e de as análises terem sido feitas apenas com relação ao objeto virtual, a conclusão de Berning *et al.* (Berning, et al., 2014) está completamente alinhada com as premissas utilizadas no presente trabalho de que várias características e pistas de profundidade devem ser inseridas na cena de RA para melhorar a percepção de profundidade entre objetos reais e virtuais.

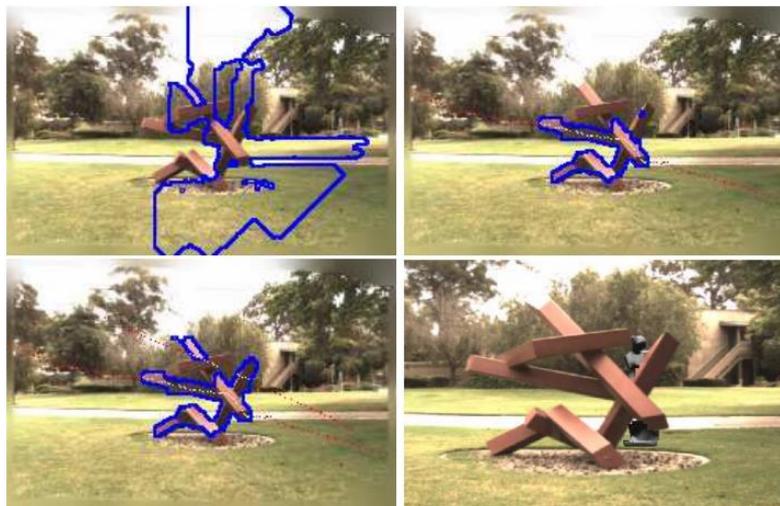
Figura 11 - Configuração dos experimentos utilizados no trabalho de Berning et al (Berning, et al., 2014). Objeto virtual verde de mesmo tamanho do real rosa (a), cubo com mesma profundidade do objeto de referencia (rosa) e sombra projetada (b), cubo com mesma profundidade e textura (c), objeto virtual (nota de 10 euros) com tamanho conhecido pelo usuário (d) e esfera sem informações de profundidade (e).



Alguns trabalhos vêm utilizando câmeras com sensores de profundidade para tratar problemas de visualização em cenas de RA com situações de oclusão. O trabalho de Wither et al. (Wither, et al., 2008) é um exemplo que usa um *hardware* com um laser de distância acoplado a uma câmera e um GPS para medir as distâncias do dispositivo até um ponto fixo no campo de visão do usuário. Através de interações, o usuário pode mapear o ambiente informando quais os objetos que fazem parte do *foreground* e do *background*. Com isso é possível obter resultados como o apresentado na Figura 12. Entretanto, a abordagem não considera várias características da cena a fim de prover uma visualização mais agradável para

percepção, como é o caso do presente trabalho.

Figura 12 - Resultado obtido utilizando a técnica de Wither et al. (Wither, et al., 2008). No canto inferior direito o resultado da cena de RA com estátua virtual aumentada.



O presente trabalho propõe adicionar uma técnica de detecção de movimento (Oh, et al., 2013) às abordagens de (Zollmann, et al., 2010) e (Padilha, et al., 2013), para em conjunto com outras características da cena, como objetos salientes e bordas, melhorar a percepção de profundidade da cena de RA. Não foram identificados outros trabalhos diretamente relacionados, que combinam importantes características para percepção humana, como detecção de saliência estática, detecção de movimento e detecção de bordas para melhorar a visualização de cenas de RA que apresentam problemas com uma camada de oclusão entre objeto virtual e real.

A técnica apresentada neste trabalho será descrita no capítulo seguinte e se baseia nas propriedades exploradas pelos trabalhos ora apresentados, os quais visam inserir pistas de profundidade para tratar o problema de oclusão e para melhorar a visualização de cenas de RA, tais como: a propriedade de forma dos objetos, explorada por (Zollmann, et al., 2010), (Avery, et al., 2009), (Sandor, et al., 2010), (Chen, et al., 2010) e (Padilha, et al., 2013); e, a propriedade de objetos salientes com alto contraste de cor e iluminação, explorada por (Zollmann, et al., 2010), (Sandor, et al., 2010) e (Padilha, et al., 2013).

3 TÉCNICA DE *GHOSTING* COM DETECÇÃO DE MOVIMENTO

A principal contribuição deste trabalho está na adição de uma técnica para detecção de regiões de saliência ao *pipeline* da técnica de *Ghosting*, desenvolvida em trabalho anterior do autor (Padilha, et al., 2013), com foco no reconhecimento de objetos em movimento (Oh, et al., 2013). Os resultados obtidos com esta técnica serão analisados no capítulo seguinte, e mostram melhorias na visualização da cena de RA utilizando esta nova abordagem.

Objetos em movimento naturalmente chamam a atenção do observador, como mostraram Treisman et al. (Treisman, et al., 1992). Abordagens da técnica de *Ghosting* apresentadas em outros trabalhos, como em (Zollmann, et al., 2010) e (Padilha, et al., 2013), não identificam as características de movimento devidamente, pois não são analisadas durante a composição do mapa de *ghosting*.

A técnica de *Ghosting* é capaz de decidir a quantidade de transparência que deve ser aplicada ao objeto aumentado para combiná-lo de forma consistente com o ambiente e gerar a cena de RA (Zollmann, et al., 2010).

Analisando características da cena real é possível obter informações necessárias para o processo de construção do mapa de *ghosting*. A partir dessas características, é possível inserir pistas de profundidade para que a visualização da cena de RA fique mais agradável para o observador, e o objeto aumentado fique contextualizado de forma consistente com o ambiente.

Trabalhos que objetivam melhorar a visualização de cenas de RA com problemas de oclusão, como por exemplo (Padilha, et al., 2013), (Lu, et al., 2010), (Kalkofen, et al., 2013) e (Sandor, et al., 2010), baseiam-se em alguns princípios para decidir como serão inseridas as pistas de profundidade na cena de RA. Os princípios são: objetos naturalmente salientes (alto contraste de cor, alto contraste de iluminação e objetos em movimento, baseado em (Itti, 2004)) à visão humana devem ser preservados na cena de RA (e não oclusos pelo objeto aumentado), pois carregam informações importantes para o entendimento da cena pelo observador; e,

manter a silhueta dos objetos garante a preservação de informações importantes dos objetos reais, como a sua forma original.

Várias técnicas, como Achanta et al. (Achanta, et al., 2009), Cheng et al. (Cheng, et al., 2011) e Itti et al. (Itti, 2004), têm sido utilizadas para construção de cenas de RA para detectar objetos salientes na cena real como em (Zollmann, et al., 2010), (Sandor, et al., 2010) e (Padilha, et al., 2013). E, para manter as silhuetas dos objetos reais, as técnicas de detecção de bordas (como Canny (Canny, 1986)) têm sido utilizadas, como pode ser visto em (Sandor, et al., 2010), (Zollmann, et al., 2010) e (Padilha, et al., 2013).

Com base nos princípios e trabalhos relacionados apresentados acima, este trabalho propõe combinar as técnicas propostas em (Padilha, et al., 2013), utilizadas na etapa de análise da cena real do *pipeline* da técnica de *Ghosting*, com uma técnica de detecção de saliência focada no movimento dos objetos e orientada à percepção visual humana (Oh, et al., 2013).

3.1 GHOSTING

A construção do mapa de *ghosting*, responsável por definir o nível de transparência que será utilizado em cada *pixel* da cena real, será discutida a seguir. O *pipeline* utilizado neste trabalho foi baseado em um trabalho anterior desenvolvido pelo autor (Padilha, et al., 2013) e está apresentado na Figura 13.

O *pipeline*, apresentado na Figura 13, possui três etapas: análise da imagem, criação do mapa de *ghosting* e renderização do mapa de *ghosting*. Na etapa de análise da imagem, técnicas de processamento de imagem e visão computacional são utilizadas para extrair as características da cena que serão consideradas na renderização da cena de RA. Cada uma das características extraídas será justificada no decorrer deste capítulo. A etapa de criação do mapa de *ghosting* consiste em reunir as informações, obtidas da etapa de análise da imagem, em um mapa contendo o nível de transparência que será aplicado a cada pixel ao compor a cena final. Por fim, a etapa de renderização do mapa de *ghosting* unirá a cena real com o

objeto virtual utilizando a técnica de *alpha blending* (Porter, et al., 1984) e os níveis de transparência indicados no mapa de *ghosting*.

Uma das técnicas utilizadas na etapa de análise da imagem, vide Figura 13, é a detecção do mapa de saliência. O mapa de saliência, no presente trabalho, foi gerado utilizando a técnica *Histogram Contrast* (HC), apresentada em (Cheng, et al., 2011), com o objetivo de identificar automaticamente quais os objetos que naturalmente chamam a atenção do usuário ao observar a cena. Esta técnica foi introduzida ao *pipeline* da técnica de *Ghosting* no trabalho anterior do autor (Padilha, et al., 2013). Essa técnica de detecção de mapa de saliência foi escolhida por ter resultado em altos valores de *precision-recall* quando submetida a uma base de 1000 imagens e comparada com as técnicas clássicas como (Itti, 2004) e (Achanta, et al., 2009), segundo análise de Oh et al. (Oh, et al., 2013).

A técnica HC consiste em gerar um histograma das cores presentes na imagem (ver Figura 14) e a partir dele calcular a diferença euclidiana entre as cores presentes, no espaço de cor CIELab (Hunter, 1948). Esse cálculo leva em consideração a probabilidade de ocorrência de cada cor, como pode ser visto em (1):

$$S(I_k) = S(c_l) = \sum_{j=0}^n f_j D(c_l, c_j), \quad (1)$$

onde S é a saliência do *pixel* k na imagem I , $D(c_l, c_j)$ é a distância euclidiana entre as cores l e j no espaço de cor CIELab, f_j é a probabilidade da cor j ocorrer na imagem I e n é a quantidade total de cores. Como para cada cor da imagem é calculada a sua distância em relação às outras, a computação dessa técnica se torna muito custosa. Então, para diminuir o tempo de execução, a técnica faz uma redução do espaço de cor, de 256^3 cores para 12^3 cores, o que não prejudica a qualidade da imagem, como pode ser visto comparando a Figura 14-a com a Figura 14-b.

Figura 13 - Pipeline da técnica de *Ghosting* utilizado neste trabalho.

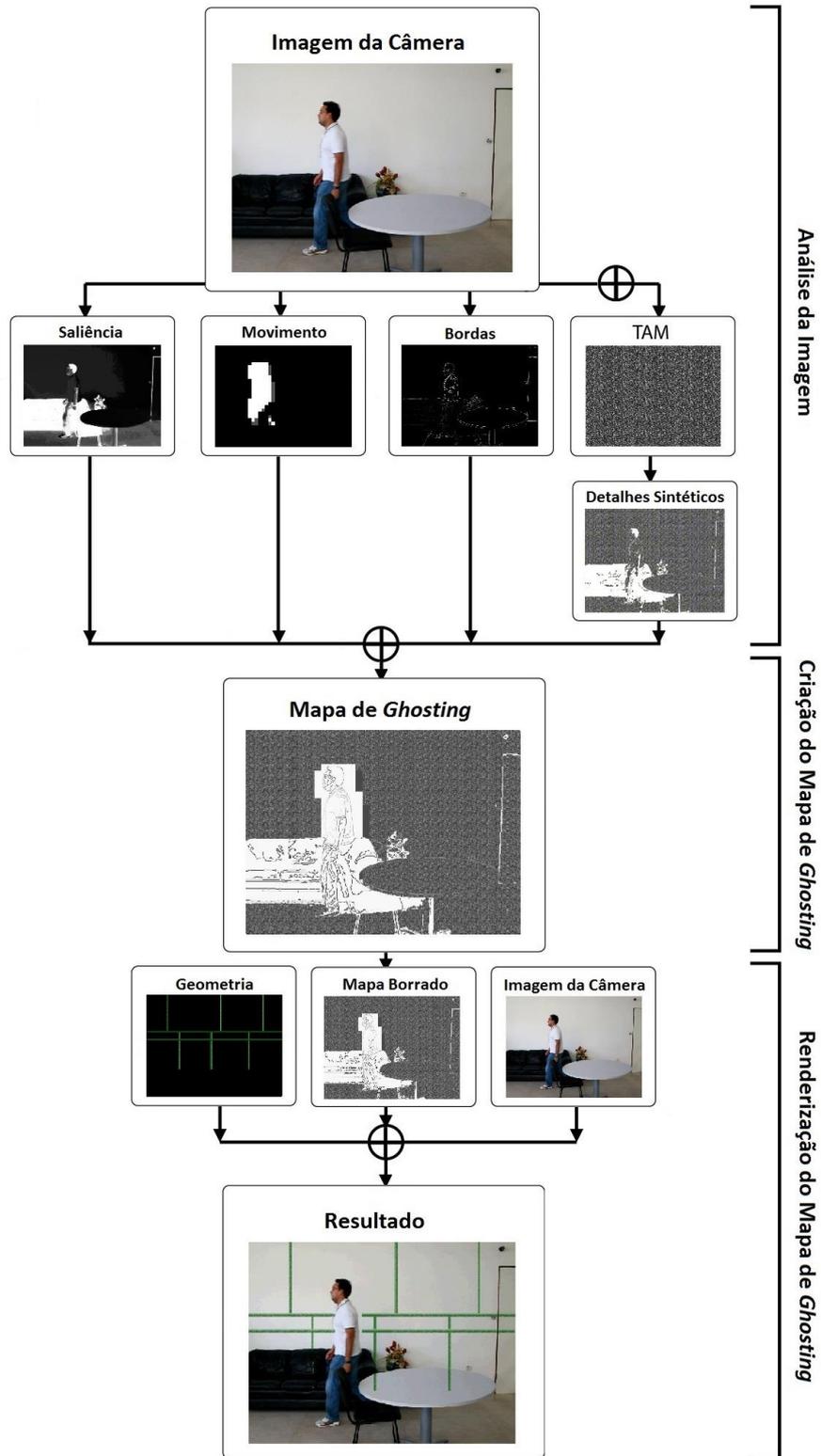
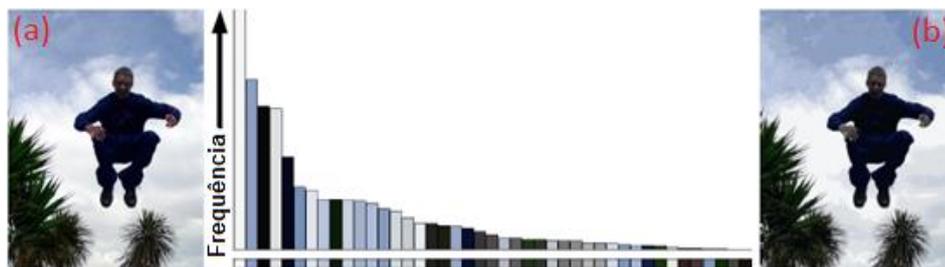


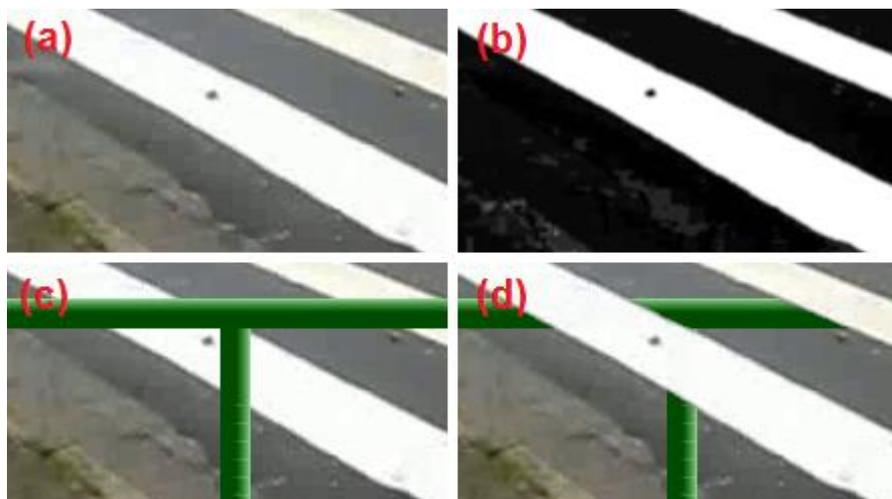
Figura 14 - Histograma de cor referente à imagem (a), e em (b) a imagem resultante após aproximações. Imagem adaptada de (Cheng, et al., 2011).



Neste processo, cada canal de cor é dividido, uniformemente, em 12 níveis e ainda são utilizadas apenas 95% das cores de maior frequência na imagem. As 5% restantes são alocadas para as cores mais próximas. Essas aproximações aceleram a computação da técnica viabilizando-a para diversas aplicações.

Para obter o mapa de saliência final, que será usado na constituição do mapa de *ghosting*, binariza-se o resultado da técnica HC (vide Figura 15-b) utilizando um ponto de corte de 128 (definido experimentalmente). Com isso, as regiões de saliência aparecem indicadas pelos *pixels* de valor não nulo após a binarização.

Figura 15 - Geração do mapa de saliência para percepção de profundidade. (a) Cena original, (b) mapa de saliência HC resultante, (c) sobreposição simples do objeto aumentado, (d) cena de RA final considerando objeto saliente.

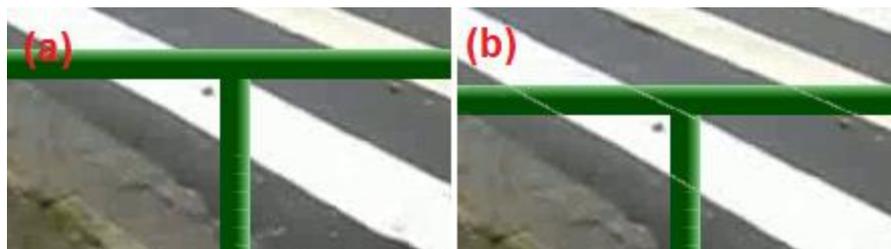


Na Figura 15, é possível analisar os resultados obtidos da técnica HC, ilustrados com um exemplo onde canos subterrâneos são aumentados em uma cena

de uma rua, preservando as partes salientes da cena. É possível perceber o ganho de percepção de profundidade quando o objeto saliente, neste caso a faixa de pedestres, sobrepõe-se ao objeto aumentado (Figura 15-d). A sobreposição simples, ilustrada na Figura 15-c, não apresenta, para o observador da cena, a sensação de que os canos estão no subterrâneo, prejudicando assim a visualização da cena. Essa comparação será mais detalhadamente discutida no capítulo seguinte.

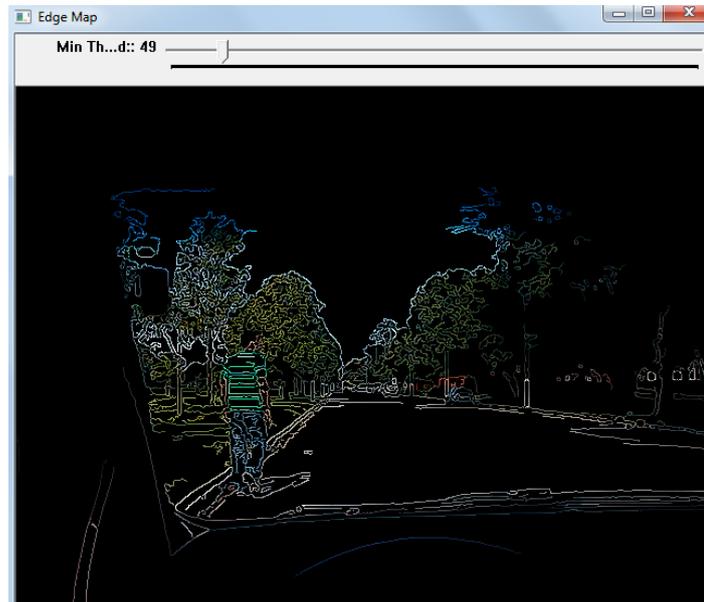
As bordas são, também, um importante elemento para proporcionar pistas de profundidade para melhorar a visualização de uma cena de RA, como pode ser visto na Figura 16, e como mostram Avery et al. (Avery, et al., 2009) e Zollmann et al. (Zollmann, et al., 2010). Ao se manter as bordas dos objetos reais na cena aumentada, a silhueta não é perdida e o contexto é preservado.

Figura 16 - Mantendo a silhueta dos objetos reais através das bordas. (a) Sobreposição bruta dos canos verdes na cena original, (b) cena de RA final com a detecção de bordas.



A detecção de bordas deste trabalho foi realizada utilizando o operador de Canny (Canny, 1986) por causa da sua abordagem ótima, além de ter sido sugerido no trabalho (Zollmann, et al., 2010) com um resultado adequado. Foi utilizado tamanho três para o filtro Gaussiano do Canny, segundo sugestão de Zollmann et al. (Zollmann, et al., 2010). Os parâmetros de *threshold* são definidos pelo usuário no início do uso da aplicação para que ele decida a quantidade de bordas que serão mantidas em cada contexto, como pode ser visto na Figura 17.

Figura 17 - Tela de configuração dos parâmetros de threshold baixo e alto para a técnica de detecção de bordas de Canny.



As regiões da cena que não possuem bordas ou objetos importantes recebem nível de transparência de acordo com uma textura sintética gerada de acordo com a técnica *Tonal Art Map* (TAM) (Praun, et al., 2001) e indicada na Figura 13. Essa abordagem foi sugerida por (Zollmann, et al., 2010) e permite que ao combinar a imagem real com a virtual as demais regiões possuam uma certa transparência que aumenta a sensação de profundidade.

Na seção seguinte, será apresentada a técnica de detecção de movimento escolhida para compor o *pipeline* deste trabalho.

3.2 DETECÇÃO DE MOVIMENTOS

Este trabalho adiciona uma técnica de detecção de movimento proposta por Oh et al. (Oh, et al., 2013) ao *pipeline* da técnica de *Ghosting*. Esta técnica foi escolhida, pois, como mostram os autores, ela reflete todos os efeitos da percepção visual humana durante o processo de detecção de objetos em movimento, os quais são salientes em uma cena real. Enquanto outros modelos de saliência como Itti (Itti, 2004) e Li et al. (Li, et al., 2011) visam, principalmente, os estímulos visuais,

essa técnica é baseada nos estímulos visuais, na capacidade limitada de resolução espaço-temporal e na memória sensorial visual como mostram Oh et al. (Oh, et al., 2013).

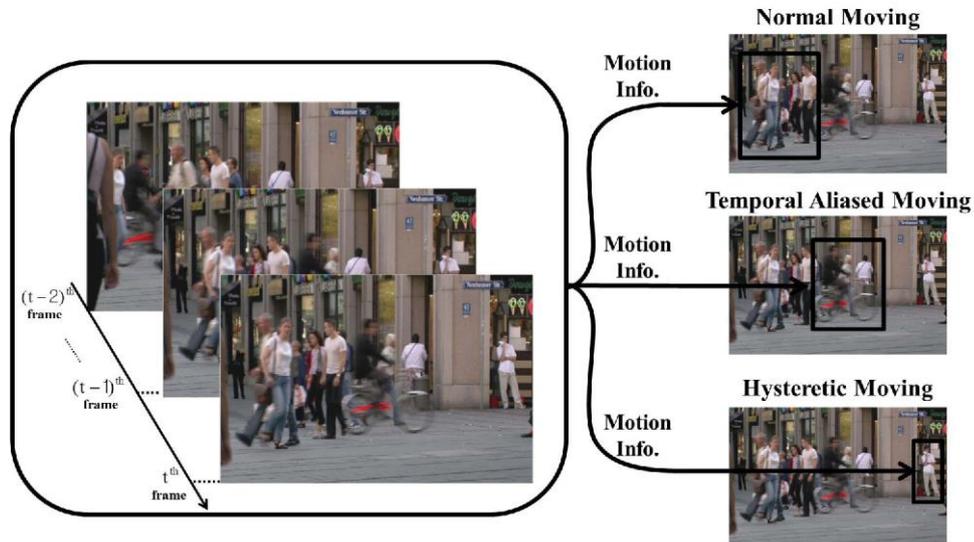
Recentemente, resultados parciais da presente dissertação de mestrado foram publicados em (Padilha, et al., 2014).

O algoritmo proposto em (Oh, et al., 2013) tem o objetivo de classificar blocos do quadro (*frame*), de uma sequência de vídeo, de acordo com seu padrão de movimento. Quatro padrões de movimento são definidos: *Normal Pattern* (NP), *Aliased Pattern* (AP), *Hysteresis Pattern* (HP) e *Background*.

O padrão de movimento NP é considerado como sendo uma movimentação natural e comum de objetos em uma cena. Um exemplo de padrão de movimento NP, na Figura 18, são pessoas caminhando na rua. O padrão AP é atribuído aos blocos do quadro em que os objetos presentes se movimentam com velocidade alta o suficiente para causar um efeito "borrado" na região. Um exemplo deste padrão de movimento pode ser visto na Figura 18, onde uma pessoa se locomove utilizando uma bicicleta. A técnica também propõe um padrão de movimento que reflete a capacidade do cérebro humano em manter a atenção em objetos que se moveram levemente por um momento e em seguida pararam de se mover, o HP. Esse padrão de movimento foi chamado de memória sensorial da visão humana. Um exemplo do padrão HP pode ser visto na Figura 18, onde uma pessoa se move para atender a uma ligação telefônica, e em seguida, para em uma posição durante a ligação.

A classificação de movimento para os *pixels* de um quadro de uma sequência de vídeo, proposta por Oh et al. (Oh, et al., 2013), foi baseada no classificador Hedge apresentado em (Freund, et al., 1995). O algoritmo é iniciado com o cálculo dos vetores de movimento para cada bloco 16x16 de cada quadro. O cálculo dos vetores de movimento utilizado não é detalhado no trabalho, portanto, o presente trabalho utilizou a clássica abordagem de Walther (Walther, 2006), que consiste em analisar os *pixels* comparando-os com os seus vizinhos em quadros consecutivos, pois assim é possível estimar para qual direção o *pixel* se moveu.

Figura 18 - Exemplos de classificação dos padrões de movimento. Normal Moving é apresentado na imagem do canto superior direito, Temporal Aliased Moving na imagem central no lado direito e Hysteretic Moving na imagem do canto inferior direito (Oh, et al., 2013).



Segundo (Walther, 2006), os vetores de movimento podem ser calculados utilizando as equações (2) e (3) para as direções horizontal e vertical, respectivamente:

$$M_{\leftrightarrow}(x, y, t) = M_I(x, y, t - \Delta t) \cdot M_I(x + \Delta x, y, t) - M_I(x + \Delta x, y, t - \Delta t) \cdot M_I(x, y, t) \quad (2)$$

$$M_{\updownarrow}(x, y, t) = M_I(x, y, t - \Delta t) \cdot M_I(x, y + \Delta y, t) - M_I(x, y + \Delta y, t - \Delta t) \cdot M_I(x, y, t), \quad (3)$$

onde (x, y) é a posição do *pixel* no mapa de intensidade (intensidades dos *pixels* da cena em escala de cinza), Δx e Δy são as variações de deslocamento em cada direção que o *pixel* sofreu de um quadro para outro, e Δt é a janela de tempo entre quadros consecutivos analisados. Para este trabalho foram utilizados os valores $\Delta x = \Delta y = 1 \text{ pixel}$ e $\Delta t = 16 \text{ quadros}$, definidos experimentalmente.

Em seguida, é realizado o cálculo da razão entre energia de textura e energia temporal residual de cada bloco, a fim de obter informações da cena para identificar os blocos com movimento embaçado, segundo a equação (4):

$$R_k^t = \frac{\sum_{q \in B_k^t} \{B_k^t(q) - B_k^t(q + M_k^t)\}^2}{\sum_{q \in B_k^t} \{B_k^t(q)\}^2 - \frac{1}{N \times N} \left\{ \sum_{q \in B_k^t} (B_k^t(q)) \right\}^2}, \quad (4)$$

onde R_k^t é a razão de energia do bloco k no quadro t , M_k^t é o vetor de movimentos calculado com as equações (2) e (3), $B_k^t(q)$ é a intensidade do *pixel* q do bloco k no quadro t , e N é o tamanho do bloco utilizado. Após encontrar a razão de energia, para cada bloco do quadro, é possível classificar o estado do movimento s_k^t de cada bloco com base na equação (5):

$$s_k^t = \begin{cases} \text{Normal state,} & \text{if } M_k^t \neq 0, R_k^t \leq 1 \\ \text{Aliased state,} & \text{if } M_k^t \neq 0, R_k^t \geq 1, \\ \text{Stationary state,} & \text{if } M_k^t = 0 \end{cases} \quad (5)$$

onde *Normal state* (NS) é atribuído aos blocos com movimento normal e que não caracteriza alta velocidade (efeito borrado na imagem), *Aliased state* (AS) é atribuído aos blocos com movimento de alta velocidade (efeito borrado na imagem) e *Stationary state* (SS) é atribuído a blocos que não possuem objetos em movimento.

Os estados obtidos na equação (5) são utilizados com o classificador Hedge, descrito em (Freund, et al., 1995), para definir qual o melhor padrão de movimento que se ajusta ao tipo de movimento apresentado em cada bloco do quadro analisado. Esta adaptação do classificador Hedge interpreta os estados de movimento (s_k^t) como os *experts*, chamados de fatores de aprendizado em (Freund, et al., 1995). Para cada bloco k do quadro t , é definido um vetor de não ocorrências I_k^t , que terá valor 0 na posição referente ao estado de movimento, e 1 caso contrário, como mostra a equação (6):

$$I_k^t = \begin{cases} [0,1,1]^T & , \text{if } s_k^t = NS \\ [1,0,1]^T & , \text{if } s_k^t = AS. \\ [1,1,0]^T & , \text{if } s_k^t = SS \end{cases} \quad (6)$$

A perda de *experts* do algoritmo Hedge (Freund, et al., 1995) é definida como vetor L_k^t , na equação (7), e é responsável por acumular a frequência dos estados de movimento obtidos durante uma janela de tempo $r = [t - T, t]$:

$$L_k^t = \sum_{r=t-T}^t I_k^r = [L_{0,k}^t, L_{1,k}^t, L_{2,k}^t]^T. \quad (7)$$

A probabilidade de um bloco ser classificado novamente com o mesmo estado de movimento anterior cresce exponencialmente a cada nova classificação repetida. Isso acontece pois o vetor L_k^t é usado para criar um vetor de pesos w_k^t , que é considerado nas probabilidades do padrão de movimento através da combinação das equações (8) e (9):

$$w_k^t = \begin{bmatrix} w_{0,k}^t \\ w_{1,k}^t \\ w_{2,k}^t \end{bmatrix} = \frac{1}{|w_k^t|} \begin{bmatrix} w_{0,k}^0 \cdot \exp(-nL_k^t(0)) \\ w_{1,k}^0 \cdot \exp(-nL_k^t(1)) \\ w_{2,k}^0 \cdot \exp(-nL_k^t(2)) \end{bmatrix} e \quad (8)$$

$$p_k^t = \begin{bmatrix} p_{0,k}^t \\ p_{1,k}^t \\ p_{2,k}^t \end{bmatrix} = \frac{1}{|w_k^t|} \begin{bmatrix} w_{0,k}^t \\ w_{1,k}^t \\ w_{2,k}^t \end{bmatrix}, \quad (9)$$

onde a norma do vetor de peso inicial w_k^0 deve ser 1, n é a taxa de aprendizado, $n > 0$ e $p_{0,k}^t, p_{1,k}^t$ e $p_{2,k}^t$ são, respectivamente, as probabilidades do bloco k no quadro t serem NS, AS e SS. Por fim, é possível definir o padrão de movimento de cada bloco k no quadro t usando a equação (10):

$$mp_k^t = \begin{cases} \text{Normal Pattern} & , \text{if } \max\{p_k^t\} = p_{0,k}^t, s_k^t \neq SS \\ \text{Aliased Pattern} & , \text{if } \max\{p_k^t\} = p_{1,k}^t, s_k^t \neq SS \\ \text{Hysteresis Pattern} & , \text{if } \max\{p_k^t\} \neq p_{2,k}^t, s_k^t = SS \\ \text{Background} & , \text{if } \max\{p_k^t\} = p_{2,k}^t, s_k^t = SS \end{cases} \quad (10)$$

Após classificar todos os blocos de um quadro, quanto ao padrão de movimento, é possível encontrar o padrão de movimento dos objetos da cena agrupando blocos com padrão de movimento igual e que sejam adjacentes.

Com isso, conclui-se a etapa de análise da imagem do *pipeline* e todas as informações necessárias para montar o mapa de *ghosting* já foram obtidas (vide Figura 13). O mapa de *ghosting* é obtido combinando todas as características

analisadas segundo a equação (11):

$$\alpha(x, y) = \begin{cases} E(x, y), & \text{if } E(x, y) > 0 \\ S(x, y), & \text{else if } S(x, y) > 0 \\ M(x, y), & \text{else if } M(x, y) > 0 \\ T(x, y), & \text{caso contrário} \end{cases} \quad (11)$$

onde $E(x, y)$ é o resultado da detecção de bordas com pixels de valor 1 se pertencer à uma borda e 0 caso contrário, $S(x, y)$ é o resultado da detecção de mapa de saliência, $M(x, y)$ resultado da detecção de movimentos e $T(x, y)$ é a textura artificial (TAM).

Por fim, para obter a cena aumentada, utiliza-se a técnica de *alpha-blending* (Porter, et al., 1984) combinando a imagem virtual e o quadro do vídeo, conforme a equação (12):

$$I = \alpha * r + (1 - \alpha) * v, \quad (12)$$

onde I é a intensidade do *pixel* na cena final aumentada, r é a intensidade do *pixel* na imagem real, v é a intensidade do *pixel* no objeto virtual e α é o valor *alpha* (a intensidade de cor do *pixel* no mapa de *ghosting*).

No capítulo seguinte, serão discutidos os estudos de caso, para ambientes *indoor* e *outdoor*, desenvolvidos nesta pesquisa para validar a técnica proposta e descrita aqui.

4 ESTUDOS DE CASO

Para validar a técnica proposta e descrita neste trabalho, foram realizados experimentos em ambientes *indoor* e *outdoor* com objetos salientes em repouso e em movimento. As cenas foram capturadas em vídeo com a resolução 640x480 a 25 quadros por segundo.

Os experimentos descritos neste capítulo foram realizados em um computador com configuração Intel(R) Core(TM) i5-3210 M CPU 2.50GHz, 8GB RAM, com placa gráfica Intel HD4000 e NVIDIA GeForce GT630M.

No decorrer deste capítulo serão apresentados estudos de caso nas áreas de arquitetura e construção civil e *safe driving* para validar a técnica proposta e para compará-la com os resultados das técnicas relacionadas da literatura descritas em (Padilha, et al., 2013) e (Zollmann, et al., 2010).

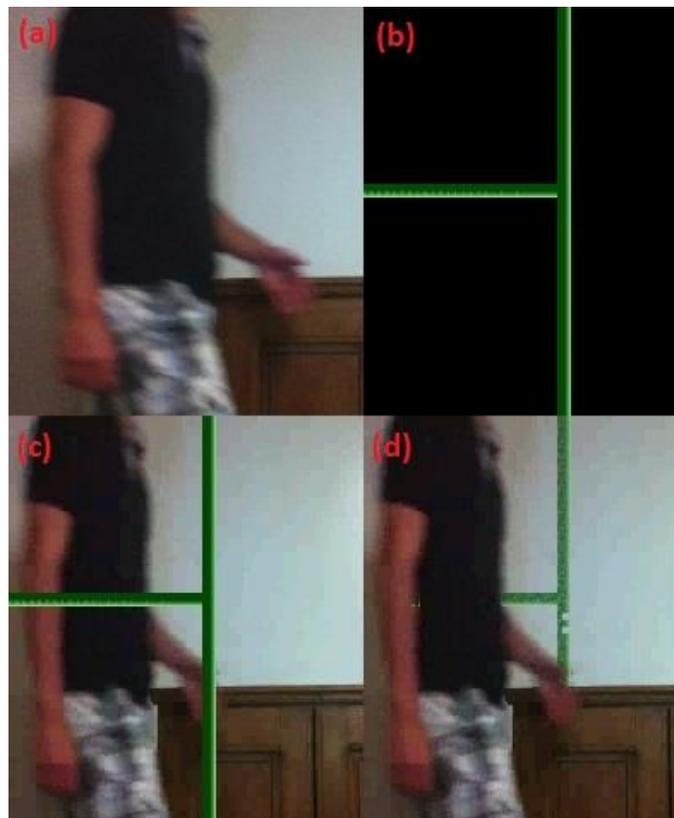
4.1 ARQUITETURA E CONSTRUÇÃO CIVIL

Como primeiro estudo de caso foi definido um cenário voltado para aplicações de arquitetura e construção civil que consiste em exibir a tubulação de canos presentes em áreas do subsolo ou no interior das paredes. Dessa forma, por exemplo, o cliente de uma construtora, interessado em um imóvel, poderia visualizar a configuração da tubulação do mesmo utilizando uma aplicação de RA com suporte da técnica proposta neste trabalho. Com isso, ao visualizar a cena, o usuário poderá perceber corretamente o conteúdo aumentado pela correta interpretação da profundidade dos objetos que a compõem. Além disso, os objetos reais que compõem a cena e que são importantes para o usuário devem ser vistos para garantir uma correta interpretação da cena, por isso não serão oclusos pelo conteúdo aumentado.

O primeiro cenário consiste de um ambiente *indoor*, em uma residência, com objetos em repouso e em movimento. A Figura 19-a apresenta um quadro extraído do vídeo da cena onde uma pessoa se locomove em frente a uma parede branca e

uma cômoda marrom, fixos no ambiente. A cena de RA é montada adicionando canos virtuais verdes (Figura 19-b) simulando a tubulação real, não visível, existente no interior da parede. Na Figura 19-c, é apresentado o resultado da cena de RA, sobrepondo o objeto virtual à cena real sem nenhum tratamento de oclusão. A cena de RA, nesse caso, não fornece ao usuário a correta percepção da cena, gerando uma interpretação incorreta da profundidade e impedindo o usuário de corretamente combinar a informação virtual sobreposta com a cena real. Portanto, o usuário não tem a sensação clara de que a tubulação está no interior da parede.

Figura 19 - Estudo de caso de arquitetura, em ambiente *indoor*. Quadro da cena real (a), canos verdes virtuais a serem aumentados na cena (b), cena de RA utilizando sobreposição do objeto virtual sem tratamento de oclusão (c) e cena de RA utilizando a técnica proposta neste trabalho (d).

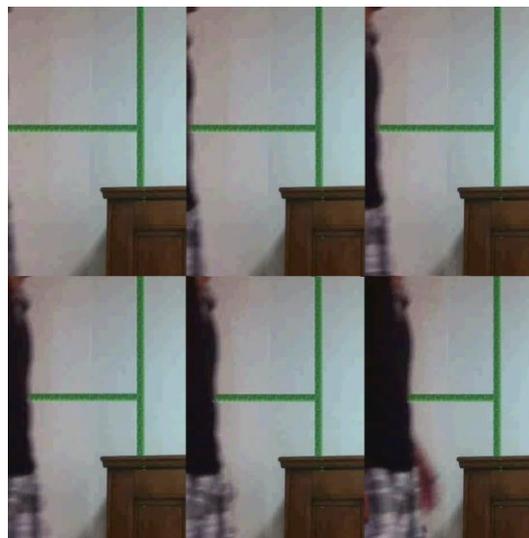


Na Figura 19-d, é possível perceber a relação de profundidade entre os objetos reais e virtuais na cena de RA. Isso ocorre porque a técnica de detecção de saliência HC (Cheng, et al., 2011) detectou o objeto marrom como saliente e, assim, os canos virtuais foram excluídos da cena nesta região. Com isso, é inserida uma

pista de profundidade na cena permitindo ao usuário ter a sensação de que os canos estão de fato dentro da parede, o que não ocorre se os canos forem simplesmente sobrepostos à cena, como apresentado na Figura 19-c. Além disso, a técnica HC, combinada com a técnica de detecção de movimento (Oh, et al., 2013), permitiu detectar a pessoa se locomovendo e, por isso, nesta região, os canos puderam ser, também, oclusos. A pessoa em movimento é algo que chama naturalmente a atenção do usuário, como foi explicado no capítulo anterior, e por isso este conteúdo deve ser preservado para uma adequada interpretação da cena.

Na Figura 20, é possível ver o efeito obtido, em quadros consecutivos do vídeo, após a geração da cena de RA utilizando a técnica proposta neste trabalho. A tubulação é oclusa nas regiões onde objetos salientes aparecem na cena e nas demais regiões a cena real fica transparente. Com isso, o usuário tem uma percepção visual correta da cena, já que passa a ser mais fácil identificar a relação de profundidade e oclusão entre os objetos reais e virtuais, o que não acontece no exemplo da Figura 19-c.

Figura 20 - Quadros consecutivos da sequência de vídeo do cenário da residência. Parede branca e cômoda marrom, fixos no ambiente, e pessoa se locomovendo são objetos reais, e, canos verdes são objetos aumentados utilizando a técnica proposta neste trabalho.



Apesar da visualização coerente da cena utilizando a técnica proposta neste

trabalho, em relação às dicas de profundidade dos objetos e ao tratamento da oclusão entre eles (Figura 21-c), o resultado obtido com a técnica (Zollmann, et al., 2010), apresentado na Figura 21-b, é muito semelhante. Isso acontece, pois, apesar de não se ter uma detecção de movimentos eficiente na técnica de (Zollmann, et al., 2010), a técnica de detecção de mapa de saliência, utilizada no trabalho, é suficiente para identificar o objeto em movimento (a pessoa) como saliente e corrigir o registro do objeto virtual nesta região. Isto ocorre porque a intensidade do contraste de cor entre a pessoa se movendo e o *background* (parede branca) é bastante alto, devido à característica escura da camisa da pessoa e a característica clara da parede.

Figura 21 - Comparação de resultados entre diferentes técnicas para cena de RA com alto contraste de cor entre objeto em movimento e *background*. Sobreposição do objeto virtual sem tratamento de oclusão (a), resultado da técnica de *Ghosting* apresentada por (Zollmann, et al., 2010) (b) e resultado da técnica proposta neste trabalho (c).

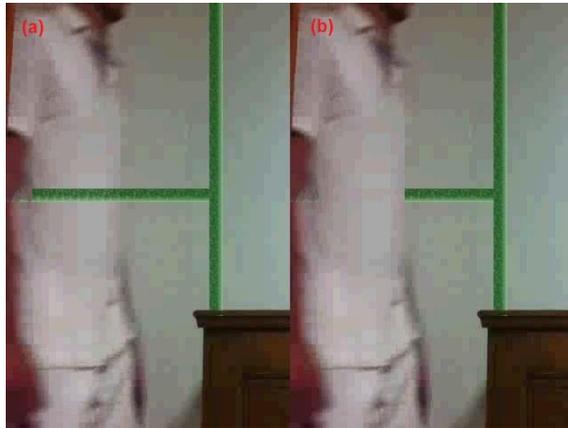


Nesse caso, a técnica proposta neste trabalho obtém resultado equivalente ao das técnicas existentes na literatura descritas em (Zollmann, et al., 2010) e (Padilha, et al., 2013).

Entretanto, em cenários onde o contraste de cor entre o objeto em movimento e o *background* tem baixa intensidade, as técnicas existentes descritas em (Zollmann, et al., 2010) e (Padilha, et al., 2013) têm dificuldades em corrigir o registro do objeto virtual. Este resultado pode ser visto na Figura 22, onde são apresentados os resultados para a técnica de (Padilha, et al., 2013) (Figura 22-a) em

comparação com a técnica proposta neste trabalho (Figura 22-b).

Figura 22 - Comparação de resultados entre diferentes técnicas para cena de RA com baixo contraste de cor entre objeto em movimento e *background*. Resultado da técnica de *Ghosting* apresentada por (Padilha, et al., 2013) (a) e resultado da técnica proposta neste trabalho (b).



É possível perceber que o cenário apresentado na Figura 22 tem substancial diferença com o apresentado na Figura 21, devido à diferença no contraste de cor entre o objeto em movimento e o *background*. A técnica de detecção de mapa de saliência, utilizada em (Padilha, et al., 2013), não foi suficiente para identificar o objeto em movimento como saliente e por isso o registro do objeto virtual não foi corrigido no resultado apresentado na Figura 22-a. Com isso, a percepção do relacionamento interobjetos fica difícil na Figura 22-a, pois não se tem a sensação clara de que a tubulação está no interior da parede, já que ela aparece sobre a pessoa que está caminhando em frente à parede. Na Figura 22-b, que apresenta o resultado da técnica proposta neste trabalho, o objeto em movimento é corretamente detectado. Com isso, o registro do objeto virtual é corrigido nesta região e a visualização se torna mais agradável no que diz respeito à relação de profundidade e ao tratamento da oclusão entre os objetos reais e virtuais.

Esse comportamento se repete em outros cenários similares. Na Figura 23, é possível ver o resultado utilizando a técnica proposta neste trabalho de outra cena *indoor* aumentada com uma tubulação virtual embutida na parede, e uma pessoa passando em frente à parede. Apesar de similar, esta cena tem uma característica

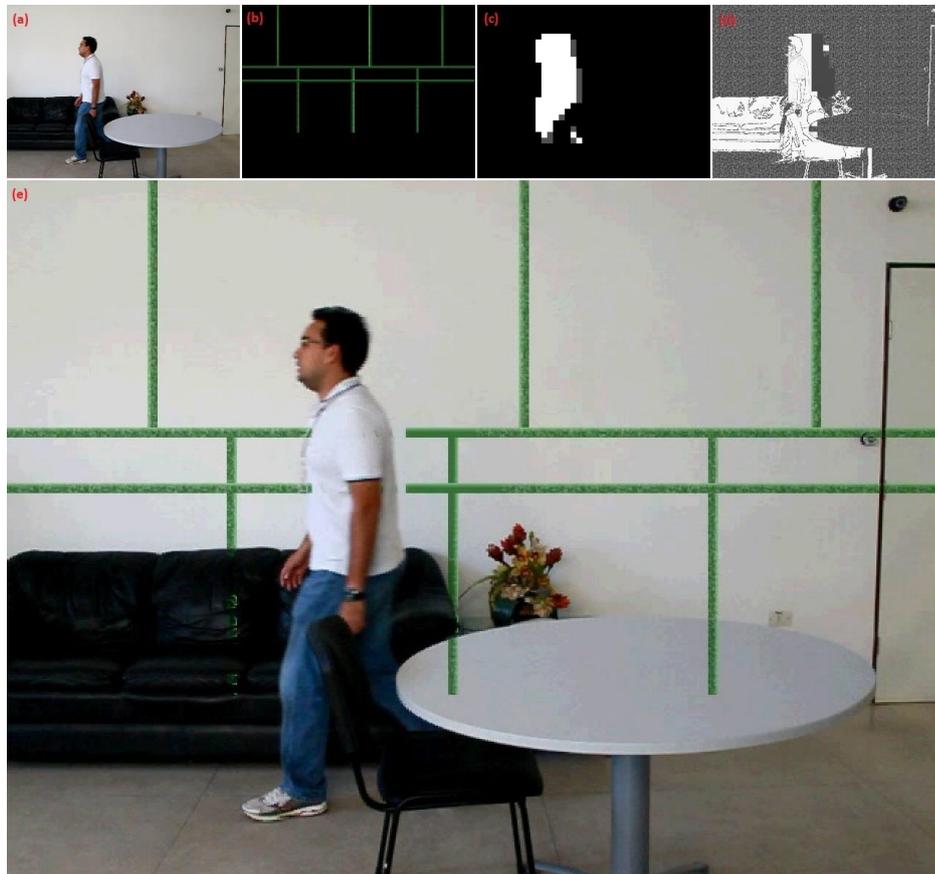
interessante que é a presença de objetos reais com alto e baixo contraste de cor com a parede, exigindo que a técnica leve em consideração os princípios de atenção visual humana para decidir como realizar o aumento do conteúdo virtual na cena real. Assim como no resultado da Figura 22-b, é possível perceber que o objeto em movimento é corretamente detectado como saliente e por isso, os canos são oclusos na região percorrida pela pessoa. Além disso, a cadeira e o sofá pretos são identificados como salientes devido ao seu alto contraste de cor com o *background*. A mesa cinza, no entanto, possui baixo contraste de cor com o *background* da cena e por isso não é considerada saliente pela técnica (Figura 23-e). Dessa forma, os canos são sobrepostos apenas com um certo nível de transparência nessa região e não são oclusos pela mesa (Figura 23-e). Como explicado anteriormente neste trabalho, segundo modelagens de saliência propostas por (Walther, 2006), (Itti, 2004) e (Achanta, et al., 2009) a mesa não é um objeto naturalmente saliente à visão humana por que não possui alto contraste de cor com o *background* e, portanto, a percepção da cena não fica prejudicada pelo fato do cano aparecer sobreposto à mesa. Se a informação visual da mesa não é tão importante então faz sentido que mais informação do objeto virtual esteja visível nessa região (Figura 23-e).

Isso não acontece na região da cena onde está o sofá preto (Figura 23-e), pois, por ter alto contraste com o *background*, ele se torna um objeto que chama, naturalmente, a atenção do usuário ao visualizar a cena. Por isso, é importante que a informação visual do sofá não seja perdida pela oclusão do objeto virtual. Dessa forma, a técnica, coerentemente, decide por remover os canos virtuais nessa região, mantendo visível a informação que é naturalmente saliente à visão humana.

Na Figura 23-c, é possível ver a área detectada como em movimento pela técnica que foi adicionada ao *pipeline* do *Ghosting* neste trabalho, descrita em (Oh, et al., 2013). O resultado condiz com a região de deslocamento da pessoa em movimento na cena, apesar de não possuir um enquadramento perfeito do objeto em movimento. Isso acontece porque a técnica de detecção de movimento classifica os blocos de 16x16 dos quadros e assim a região detectada como em movimento

não aproxima perfeitamente o contorno do objeto.

Figura 23 - Cenário *indoor* com baixo e alto contraste de cor entre alguns objetos e o *background*. Quadro da cena real (a), canos verdes virtuais a serem aumentados na cena (b), resultado da detecção de movimento (c), mapa de *ghosting* resultante (d) e cena de RA utilizando a técnica proposta neste trabalho (e).



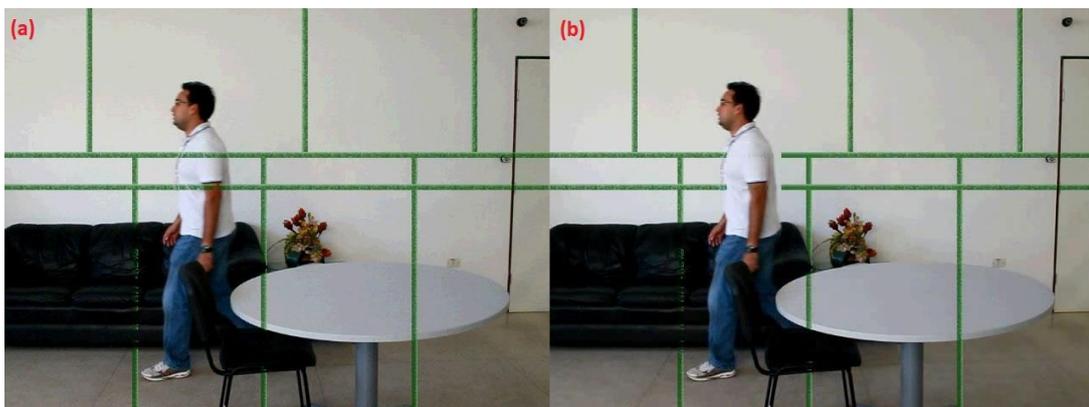
A região que não está com cor preta (Figura 23-c) é considerada como saliente, e com isso, o efeito visual de profundidade da Figura 23-e é obtido. É possível visualizar blocos com cores diferentes na Figura 23-c. Os blocos com cor preta indicam regiões onde não existem objetos em movimento na cena, os blocos com a cor branca indicam regiões onde existem objetos com movimento normal na cena e os blocos com a cor cinza indicam as regiões onde existem objetos com movimento rápido na cena, segundo classificação da técnica (Oh, et al., 2013) apresentada no capítulo anterior.

O mapa de *ghosting* resultante (após análise do quadro do vídeo da cena

mostrado na Figura 23-a) pode ser observado na Figura 23-d. Como explicado no capítulo anterior, a renderização da cena de RA a partir da imagem real e virtual é feita utilizando a técnica de *alpha-blending* (Porter, et al., 1984). Dessa forma, a Figura 23-d mostra uma representação gráfica do mapa de *ghosting* que traduz os valores de *alpha* em cores RGB (valores entre 0 e 1 que são aplicados em cada *pixel* da imagem usando a equação 12), durante o processo de renderização do *Ghosting* apresentado na Figura 23. Dessa forma, quanto mais próxima do branco RGB (255,255,255) for a cor do *pixel*, mais opaca (objeto virtual menos visível) será a cena real e quanto mais próxima do preto RGB (0,0,0), mais transparente (objeto virtual mais visível) será a cena real.

O mesmo não ocorre com a aplicação de outras técnicas neste tipo de cena, como foi discutido em relação à Figura 22-a. O efeito visual para facilitar a percepção da profundidade entre os objetos virtuais e reais não é atingido utilizando as técnicas existentes de (Padilha, et al., 2013) e (Zollmann, et al., 2010) (Figura 23-a). É possível perceber que no resultado de (Padilha, et al., 2013), a tubulação sobrepõe a pessoa em movimento (objeto saliente) prejudicando a percepção de profundidade da cena, enquanto que, no resultado deste trabalho, ilustrado na Figura 24-b, a visualização é coerente.

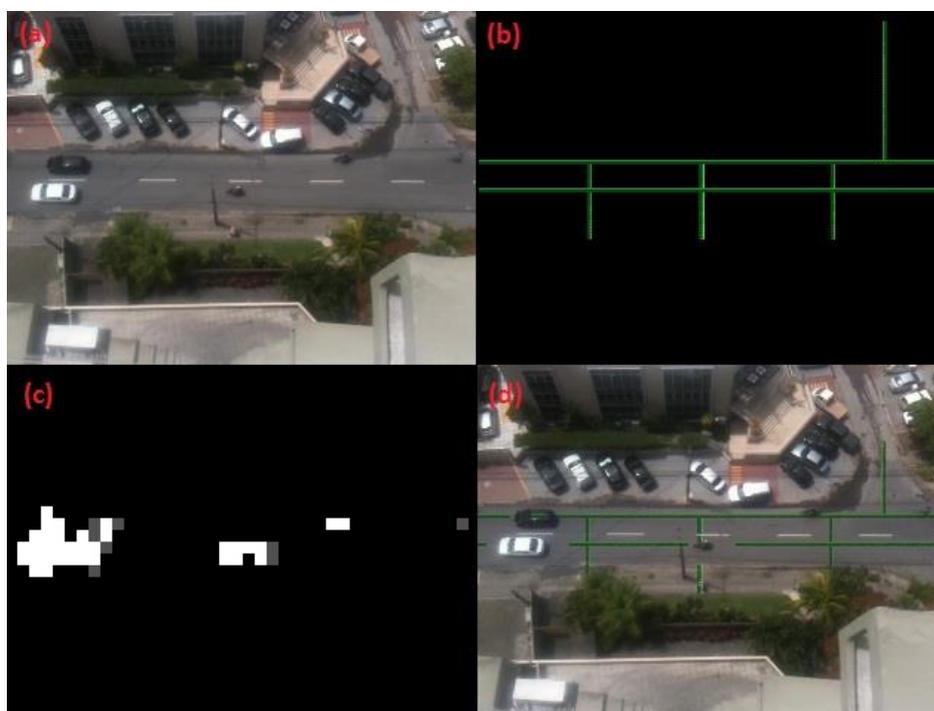
Figura 24 - Comparação de resultados entre diferentes técnicas para cena de RA com baixo contraste de cor entre objeto em movimento e *background*. Resultado da técnica de *Ghosting* apresentada por (Padilha, et al., 2013) (a) e resultado da técnica proposta neste trabalho (b).



Ainda no contexto de arquitetura e construção civil, a técnica foi testada em

ambientes externos. O cenário, neste caso, consiste de uma rua onde pessoas e veículos de duas e quatro rodas trafegam, conforme ilustrado na Figura 25-a. Na Figura 25-c, é possível visualizar as regiões onde os objetos em movimento foram detectados. A cena de RA é montada adicionando canos verdes virtuais (Figura 25-b) simulando a tubulação real, não visível, existente no subsolo da rua. Os resultados obtidos podem ser vistos na Figura 25-d e Figura 26, a qual apresenta vários quadros da cena de RA final.

Figura 25 - Estudo de caso de construção civil, em ambiente *outdoor*. Quadro da cena real (a), canos verdes virtuais a serem aumentados na cena (b), resultado da detecção de movimento (c) e cena de RA utilizando a técnica proposta neste trabalho (d).



Em outro cenário *outdoor*, com diferentes características, a técnica proposta neste trabalho pode não obter o resultado esperado, vide Figura 27. É possível perceber, na Figura 27-d, que em algumas regiões os canos aumentados não se integram de forma coerente com cena. Isso acontece, pois a região afetada possui alta diferença de iluminação e acaba induzindo a técnica de detecção de mapa de saliência (Cheng, et al., 2011), a identificar essa região como saliente. Apesar desse ponto negativo da técnica apresentada neste trabalho, o resultado obtido utilizando a

técnica de (Zollmann, et al., 2010), também não é capaz de proporcionar pistas de profundidade para melhorar a visualização da cena de RA (Figura 27-c).

Figura 26 - Quadros consecutivos de rua com veículos trafegando (objetos reais) e canos verdes (objetos aumentados) utilizando a técnica proposta neste trabalho.

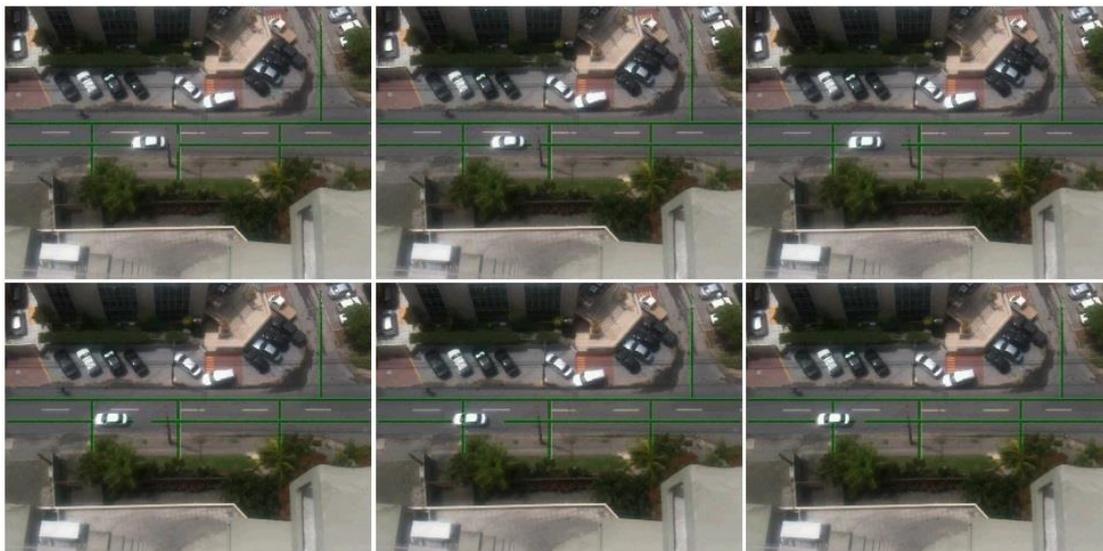
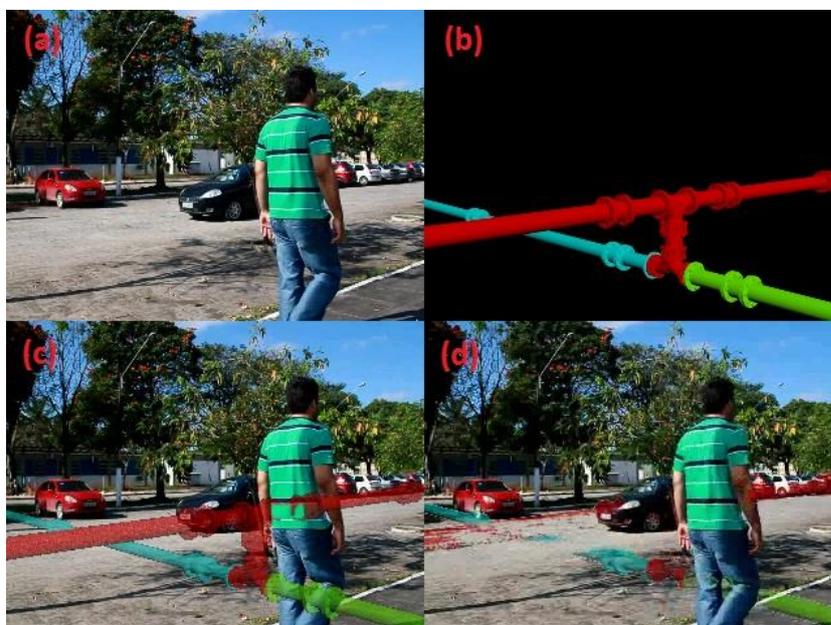


Figura 27 - Estudo de caso de construção civil, em ambiente *outdoor*. Quadro da cena real (a), canos virtuais a serem aumentados na cena (b), resultado da técnica de *Ghosting* apresentada por (Zollmann, et al., 2010) (c) e resultado da técnica proposta neste trabalho (d).



4.2 *SAFE DRIVING*

Cada vez mais os carros que dirigimos têm recebido sistemas computacionais para diversas funções que facilitam a utilização do veículo. Uma aplicação possível para a técnica desenvolvida por este trabalho é o uso de RA em automóveis, mais especificamente, projeções de imagens no para-brisa do carro. Este tipo de aplicação é interessante, pois permite exibir informações úteis para o condutor evitando distrair sua atenção ao volante.

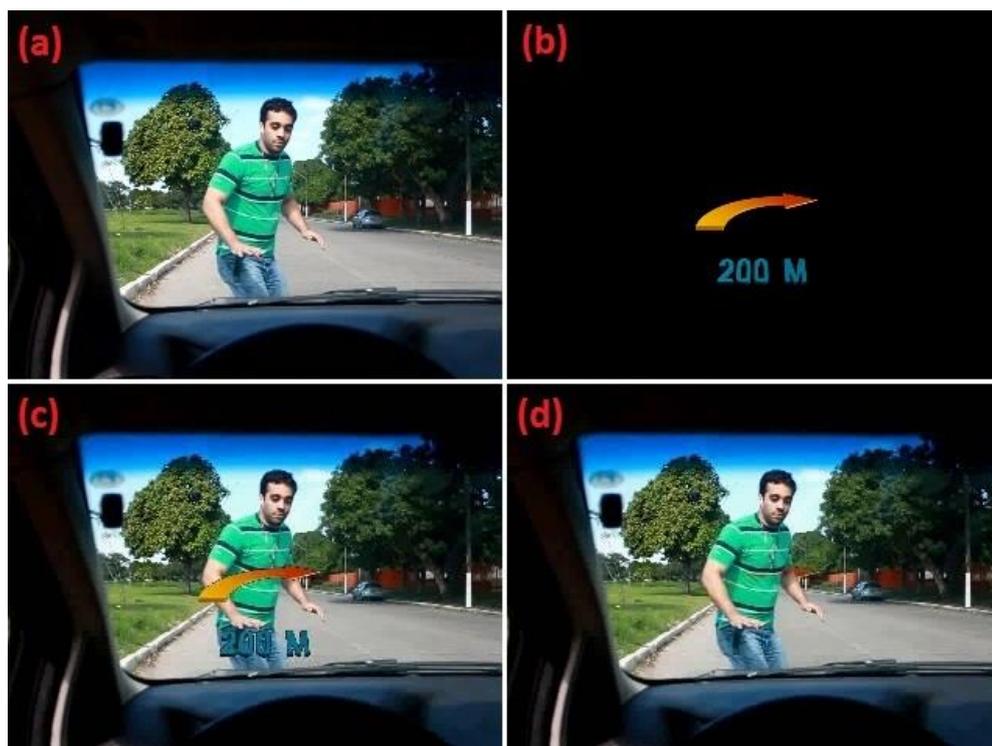
O objetivo é permitir que o sistema de RA decida o nível de transparência do objeto aumentado de acordo com as características naturais do ambiente que o motorista está vendo. Em um cenário real, uma câmera captaria as imagens da cena real que o motorista está vendo e o sistema de RA utilizaria a técnica, proposta neste trabalho, para corrigir o registro do objeto virtual. Deste modo, a cena de RA permitiria ao condutor visualizar informações importantes do ambiente mesmo com a existência do sistema de RA no para-brisa do carro.

Os cenários abordados para esse estudo de caso consistem de cenas de vídeo filmadas do ponto de vista do motorista em diferentes situações do trânsito, como pessoas, automóveis e objetos atravessando o campo de visão do motorista.

O primeiro cenário consiste de uma situação em que o motorista se surpreende com a passagem repentina de um pedestre na frente do carro. É importante que a projeção da cena aumentada se adapte para facilitar a visão do motorista do objeto importante na frente do carro, que é o pedestre. A Figura 28-a mostra um quadro do vídeo da cena analisada. A Figura 28-b mostra a seta virtual que é aumentada no para-brisa do carro referente à direção do GPS, o qual indica uma curva para a direita. A Figura 28-c mostra a cena de RA com a sobreposição da informação virtual sem a utilização de nenhuma técnica com foco na melhoria da percepção da cena. Finalmente, a Figura 28-d mostra o resultado utilizando a técnica proposta por este trabalho, onde a cena aumentada é gerada realizando o tratamento de oclusão na região onde a pessoa se move na cena.

Na Figura 28-d, é possível perceber que o número virtual aumentado na cena, o qual indica a distância até a próxima curva, não aparece. Isso acontece porque a pista está sendo identificada como um objeto saliente similar ao que foi discutido em relação à Figura 27.

Figura 28 - Estudo de caso de *safe driving*, em ambiente *outdoor*. Quadro da cena real (a), objeto virtual aumentado (b), cena de RA utilizando sobreposição do objeto virtual sem tratamento de oclusão (c) e cena de RA utilizando a técnica proposta neste trabalho (d).



Na Figura 29, é possível visualizar uma sequência da cena de RA utilizando a técnica proposta por este trabalho. A aplicação de RA de GPS está aumentando a cena com uma seta virtual e um número no para-brisa do carro, indicando, respectivamente, a direção da próxima curva e a distância até ela. Após a aplicação da técnica, a visualização da cena é corrigida quando uma pessoa atravessa a pista na frente do carro, vide Figura 29.

A aplicação de RA, utilizando a técnica proposta neste trabalho, apresenta o objeto virtual, em uma região do para-brisa do carro, e o esconde no momento em que um objeto importante (segundo as características já discutidas no capítulo

anterior) invade essa região. Dessa forma, é possível que o usuário usufrua dos benefícios da aplicação de RA e ainda consiga visualizar objetos importantes no seu campo de visão.

Figura 29 - Quadros consecutivos da sequência de vídeo do cenário da situação de trânsito. A cena completa é real, com exceção da seta indicando direção do trajeto e indicador de distância, que são aumentados utilizando a técnica proposta neste trabalho.



Nessa situação, a informação virtual aumentada na cena poderia prejudicar a visualização do ambiente caso o registro do objeto virtual não fosse corrigido, como é o caso da Figura 30. Esse cenário de atropelamento, ilustrado na Figura 29 e na Figura 30, pode ficar ainda mais crítico caso o objeto virtual ocupe uma área maior do campo de visão do motorista e atrapalhe a visualização da pessoa atravessando a pista. A técnica proposta neste trabalho evita esse problema por identificar regiões de saliência na cena, incluindo regiões de movimento.

Outros cenários foram testados, no contexto de *safe driving*, a fim de analisar o desempenho da técnica em diferentes situações e com diferentes objetos interagindo na cena de RA.

Figura 30 - Quadros consecutivos da sequência de vídeo da cena de RA com sobreposição sem tratamento de oclusão do objeto virtual (seta indicando direção do trajeto e indicação da distância).



Na Figura 31, é possível visualizar os quadros do vídeo de uma cena onde um carro com cor similar ao do *background* se desloca na cena vista pelo motorista, a qual possui, novamente, um objeto virtual de uma seta referente a uma aplicação de GPS aumentada. Mais uma vez, é importante que o motorista visualize com clareza o carro em deslocamento e que o objeto virtual não atrapalhe a visualização dos objetos importantes do ambiente. Os resultados apresentados na sequência de quadros da Figura 31 mostram que a técnica proposta neste trabalho foi capaz de detectar os objetos importantes e corrigir o registro do objeto virtual nessas regiões, entretanto, persiste o problema do enquadramento da região em movimento com o contorno do objeto, assim como explicado na Figura 23.

É possível perceber que o carro é identificado como em movimento e considerado como saliente pela técnica proposta neste trabalho, vide Figura 31. Dessa forma, o registro do objeto virtual é corrigido para que o motorista visualize o objeto naturalmente saliente da cena.

Figura 31 - Quadros consecutivos da sequência de vídeo da cena de RA com um carro, com pouco contraste de cor entre o objeto em movimento e o background, passando. Resultado da técnica proposta neste trabalho.



A Figura 32 apresenta a comparação visual entre os resultados obtidos com a técnica proposta neste trabalho, ilustrado na Figura 32-c, e com as técnicas existentes na literatura propostas em (Zollmann, et al., 2010), vide Figura 32-a, e em (Padilha, et al., 2013), vide Figura 32-b. É possível perceber que nenhuma das técnicas anteriores foi capaz de identificar o objeto saliente em movimento para tratar a oclusão entre objeto real e virtual. Isso acontece por causa da falta de

contraste entre este objeto e o *background*, similarmente ao que foi discutido em relação ao outro estudo de caso comparado na Figura 24.

Figura 32 - Comparação de resultados entre a técnica de *Ghosting* apresentada por (Zollmann, et al., 2010) (a), por (Padilha, et al., 2013)(b) e a técnica proposta neste trabalho (c).



Outros dois cenários de teste envolvem carros em movimento (Figura 33 e Figura 34) e uma bola (Figura 34) atravessando na frente do carro em movimento. Na Figura 33-d, é possível ver o resultado obtido para um quadro da cena de RA utilizando a técnica proposta por este trabalho, e é possível compará-lo com uma cena de RA obtida sobrepondo o objeto virtual na cena sem tratamento de oclusão, mostrado na Figura 33-c.

A visualização do objeto importante em movimento é prejudicada no caso ilustrado na Figura 33-c, enquanto que, na Figura 33-d, a cena de RA permite uma visualização clara de um objeto tão importante para o motorista, como um outro carro em movimento na pista.

Ainda se tratando do mesmo cenário analisado na Figura 33, a Figura 34 mostra a comparação visual entre os resultados obtidos com a técnica proposta neste trabalho (Figura 34-c, Figura 34-f, Figura 34-i, Figura 34-l) e com as técnicas anteriores propostas em (Zollmann, et al., 2010) (Figura 34-a, Figura 34-d, Figura 34-g, Figura 34-j) e em (Padilha, et al., 2013) (Figura 34-b, Figura 34-e, Figura 34-h, Figura 34-k). Aqui ilustramos exemplos com o carro em movimento atravessando na frente do carro com a projeção de RA no para-brisa, assim como um objeto (bola) atravessando na frente do carro.

Figura 33 - Cena de RA com um carro em movimento. Cena real (a), objeto virtual aumentado (b), cena de RA utilizando sobreposição do objeto virtual sem tratamento de oclusão (c) e cena de RA utilizando a técnica proposta neste trabalho (d).

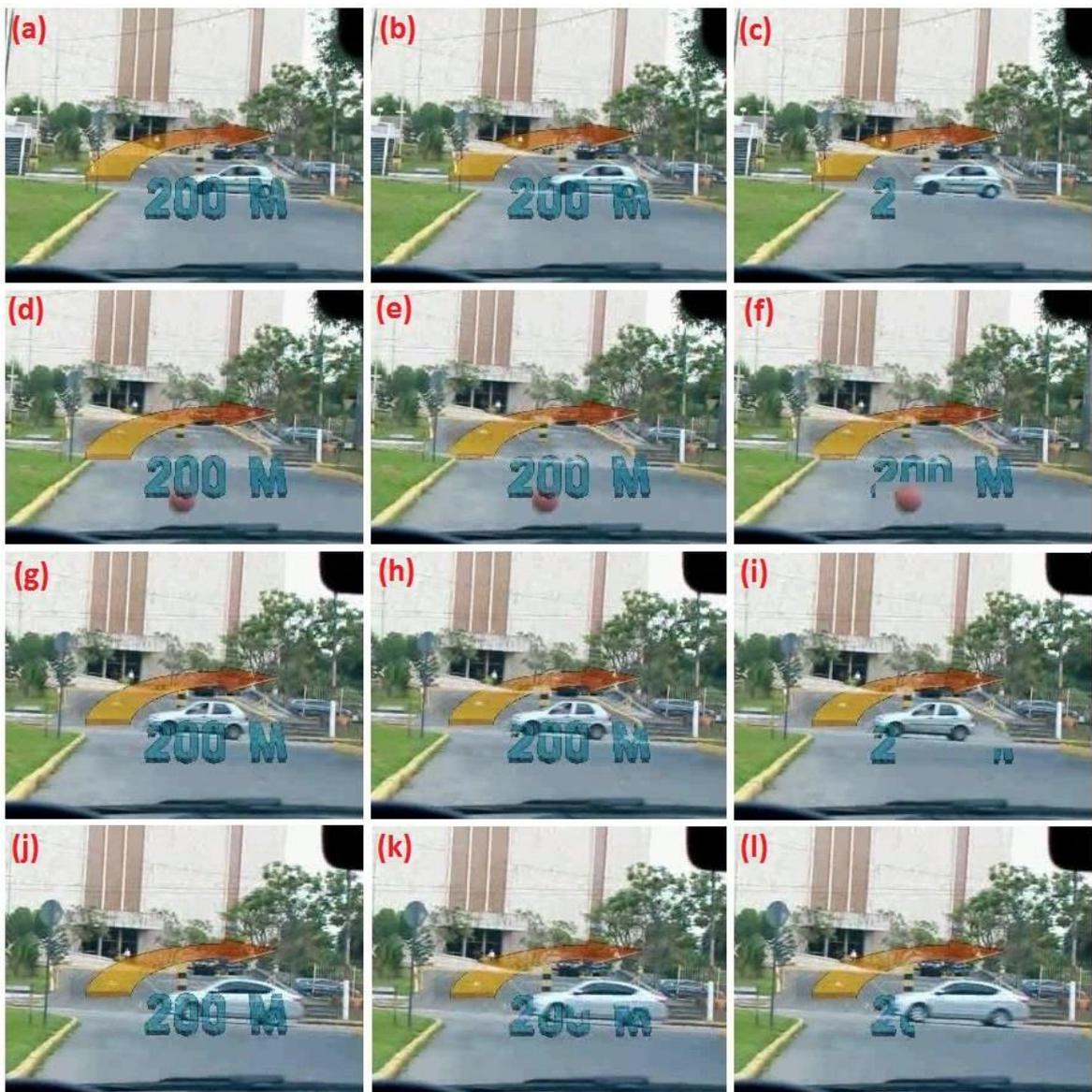


Cada linha da Figura 34 apresenta uma situação. A primeira linha (Figura 34-a, Figura 34-b e Figura 34-c) mostra um quadro do vídeo onde passa o primeiro carro, a segunda linha (Figura 34-d, Figura 34-e e Figura 34-f) mostra um quadro do vídeo onde passa uma bola arremessada no meio da pista, a terceira linha (Figura 34-g, Figura 34-h e Figura 34-i) mostra um quadro do vídeo onde passa o segundo carro, e, a quarta linha (Figura 34-j, Figura 34-k e Figura 34-l) mostra um quadro do vídeo onde passa o terceiro carro na pista. Nesse cenário, foram analisados vários objetos de tamanhos diferentes e com velocidades diferentes invadindo a região onde é apresentado o objeto virtual na aplicação de RA, a fim de estressar as técnicas e avaliar quais seriam capazes de identificar esses objetos importantes e corrigir o registro do objeto virtual de modo a evitar possíveis acidentes.

É possível perceber que os resultados obtidos utilizando as técnicas de (Zollmann, et al., 2010) (Figura 34-a, Figura 34-d, Figura 34-g e Figura 34-j) e de (Padilha, et al., 2013) (Figura 34-b, Figura 34-e, Figura 34-h e Figura 34-k) não foram capazes de tornar a visualização da aplicação de RA mais segura para o usuário, afinal, objetos importantes, como os carros e a bola foram oclusos pelo

objeto virtual, podendo assim causar um acidente. Por outro lado, a técnica proposta neste trabalho permitiu visualizar o objeto importante em todos os quadros destacados (Figura 34-c, Figura 34-f, Figura 34-i e Figura 34-l), garantindo assim mais segurança na direção enquanto o usuário utiliza o sistema de RA.

Figura 34 - Cena de RA com carro e uma bola em movimento. Resultado da técnica de *Ghosting* apresentada por (Zollmann, et al., 2010): carro 1 (a), bola (d), carro 2 (g), carro 3 (j); resultados da técnica de *Ghosting* apresentada por (Padilha, et al., 2013): carro 1 (b), bola (e), carro 2 (h), carro 3 (k); e, resultado da técnica proposta neste trabalho: carro 1 (c), bola (f), carro 2 (i), carro 3 (l).



Com isso, foi possível testar a técnica proposta neste trabalho em vários

contextos e compará-la visualmente com as técnicas anteriores propostas na literatura.

4.3 AVALIAÇÃO QUALITATIVA

Desenvolveu-se um questionário a fim de avaliar qualitativamente os resultados obtidos com a técnica proposta. O método utilizado, para avaliação, foi enviar o questionário para usuários com idades de 19 a 35 anos, homens e mulheres, de diferentes estados (Pernambuco e São Paulo) e com diferentes níveis de conhecimento técnico (graduados em computação, pós-graduados em computação e sem conhecimento técnico).

O questionário foi elaborado com dez perguntas sendo duas sobre o usuário e oito referentes à comparação de resultados utilizando técnicas de tratamento de oclusão relacionadas com este trabalho, vide (Ava15).

As primeiras três perguntas, referentes ao trabalho, foram apresentadas com caráter introdutório e com o objetivo de instruir o usuário sobre os problemas comuns abordados pelo presente trabalho.

Para a primeira pergunta, foi apresentada a imagem da Figura 1-a e perguntado: "Na sua opinião, o boneco STB deveria aparecer na frente do boneco vermelho?". Diante das opções "Sim" e "Não", 73% dos participantes optaram pela opção "Não", indicando que de fato o problema de distorção da profundidade apresentado na Figura 1-a é relevante para o estudo.

Para a segunda pergunta, foi apresentada a imagem da Figura 1-b e perguntado: "Na sua opinião, você acha que é possível ver os três bonecos com clareza?". Diante das opções "Sim" e "Não", 65% dos participantes optaram pela opção "Não", indicando que de fato o problema de visibilidade e abstração da cena apresentado na Figura 1-b é relevante para o estudo.

Para a terceira pergunta, foi apresentada a imagem da Figura 1-c e perguntado: "Na sua opinião, você acha que é possível ver, com clareza, as rodas

vermelhas traseiras do carro?". Diante das opções "Sim" e "Não", apenas 38% dos participantes optaram pela opção "Não", indicando que, provavelmente, a pergunta não evidenciou corretamente que se desejava visualizar claramente as duas rodas traseiras. A roda traseira aumentada que está totalmente ocluída pelo carro possui cor muito próxima da cor do carro e isso dificulta a visibilidade da cena de RA, vide Figura 1-c.

Após as perguntas introdutórias, foram apresentados vídeos de resultados obtidos utilizando diferentes técnicas de tratamento de oclusão. Foi utilizada uma codificação para cada resultado apresentado a fim de que os nomes dos autores não influenciassem na escolha das respostas. A codificação utilizada em todo o questionário foi: HC_COM para cenas de RA obtidas utilizando a técnica proposta neste trabalho, HC_SEM para cenas de RA obtidas utilizando a técnica proposta em nosso trabalho anterior (Padilha, et al., 2013), FT_SEM para cenas de RA obtidas utilizando a técnica proposta em (Zollmann, et al., 2010) e SO para cenas de RA convencional, sem tratamento de oclusão.

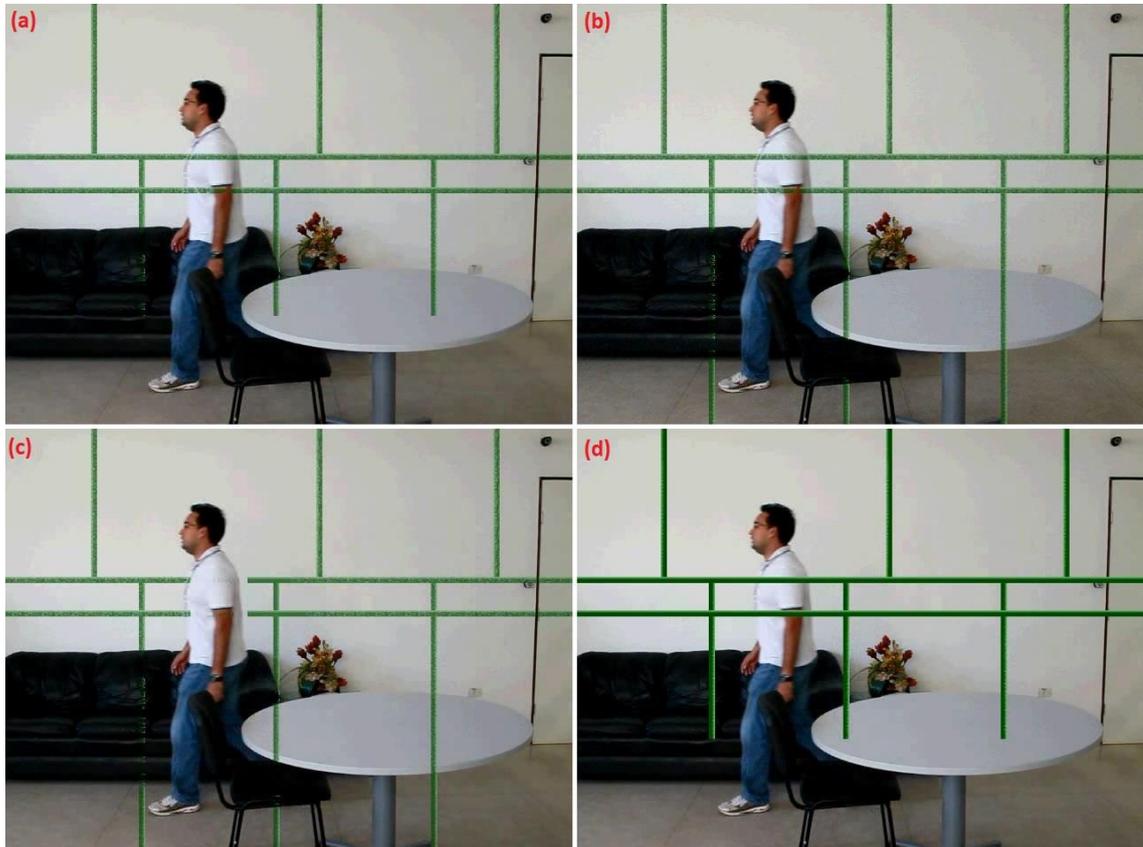
Inicialmente foram apresentados dois vídeos, um utilizando a técnica HC_COM proposta no presente trabalho, e outro utilizando RA convencional. O objetivo com essa pergunta era avaliar se os usuários consideram que resultados obtidos utilizando tratamento de oclusão são mais seguros ao utilizar um sistema de RA no carro (aplicação de *Safe Driving*) em comparação com RA convencional. Os vídeos utilizados estão ilustrados na Figura 29 (técnica proposta no presente trabalho) e Figura 30 (RA convencional). 88% dos usuários participantes escolheram o sistema de RA utilizando a técnica proposta no presente trabalho. Os 12% que optaram pela RA convencional justificaram sua resposta como: "A informação é mais precisa e clara", "Primeira opção é vista com mais clareza" e "A realidade aumentada não alterou minha percepção do mundo ao redor"; provavelmente, alguns deles desconsideraram o fator segurança, que foi levantado na pergunta, durante as suas avaliações.

As perguntas seguintes seguiram o formato da escala de Likert (Likert, 1932) para avaliar os resultados apresentados utilizando as técnicas HC_COM, HC_SEM, FT_SEM e SO.

Para cada cena abordada, foi apresentada uma introdução da seguinte forma: "Por favor, assista ATENTAMENTE aos 4 vídeos do youtube que estão apresentados abaixo e em seguida marque uma avaliação depois deles.". Para cada cena, foi descrito brevemente o conteúdo que seria exibido e o objeto virtual que estava sendo aumentado na cena, como por exemplo, "Os vídeos consistem de uma cena de uma casa onde uma pessoa se desloca. O sistema de Realidade Aumentada (ou apenas RA) está inserindo uma tubulação verde que não é visível a olho nu, pois está no interior da parede.". Essas explicações foram necessárias para evitar dupla interpretação quanto aos objetos aumentados e o conteúdo da cena. Após visualizar cada vídeo com os resultados das técnicas, foi solicitado que o usuário avaliasse o resultado segundo as características, como por exemplo, "Por favor, avalie cada um dos vídeos abaixo levando em consideração a percepção de profundidade. Considere 1 estrela sendo baixa coerência na relação de profundidade entre os objetos reais e virtuais e 5 sendo altíssima coerência na relação de profundidade entre os objetos reais e virtuais".

As primeiras cenas de vídeo (Vídeo 1 e Vídeo 2) fazem parte do contexto de Arquitetura e Construção Civil e o objetivo com o questionário foi avaliar se o tratamento de oclusão permite uma melhoria na percepção de profundidade para os usuários. A primeira cena possui objetos salientes em movimento e estáticos em um ambiente *indoor* e a segunda cena possui objetos salientes sem movimento com tamanhos diferentes e velocidades diferentes em um ambiente *outdoor*.

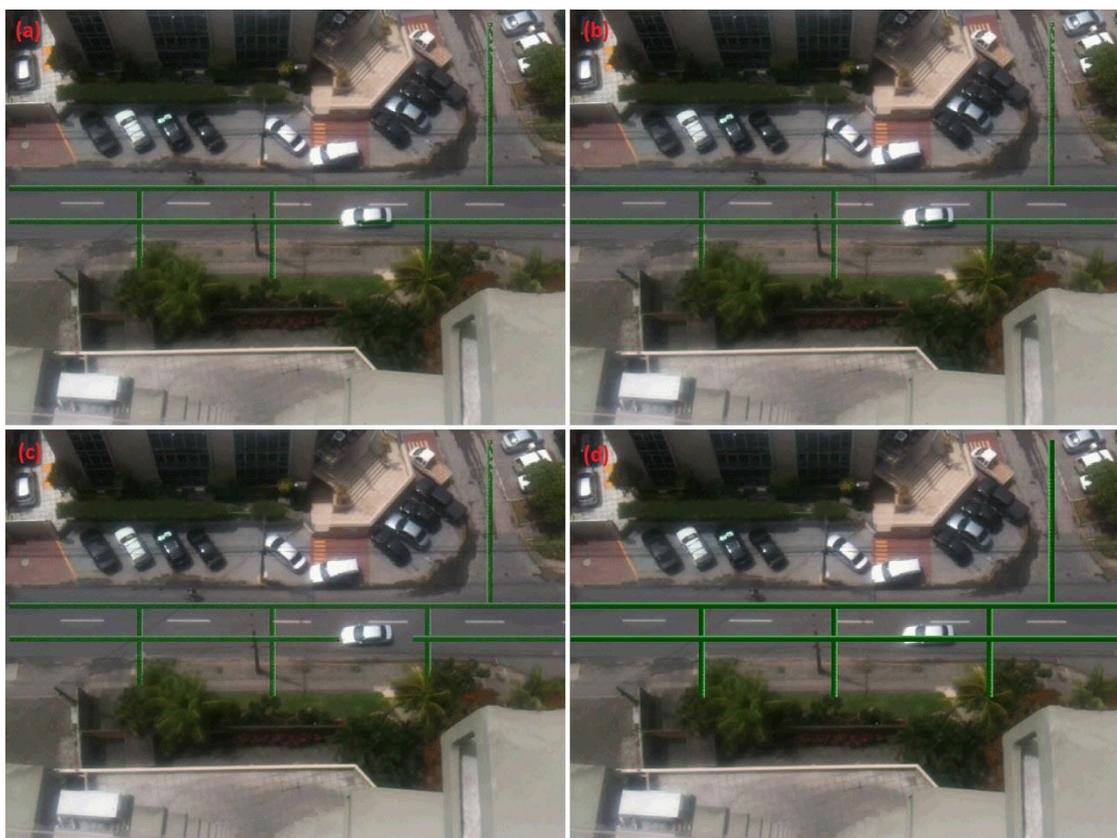
Figura 35 - Quadros da cena de vídeo *indoor* para aplicação de Arquitetura e Construção Civil com objetos não salientes e objetos salientes estáticos e em movimento utilizada no questionário de avaliação qualitativa. A ordem apresentada no questionário corresponde à ordem apresentada na figura. Resultado obtido utilizando (Zollmann, et al., 2010) (a), resultado obtido utilizando (Padilha, et al., 2013) (b), resultado obtido utilizando a técnica proposta neste trabalho (c) e resultado obtido utilizando RA convencional (d).



Para a primeira cena de vídeo voltada à comparação de resultados referentes à aplicação de Arquitetura e Construção Civil, a seguinte explicação foi fornecida para guiar a avaliação dos resultados: "Abaixo você verá vários resultados de técnicas visando melhorar a visualização de cenas de RA através do tratamento de situações de oclusão entre objetos reais (sofá, pessoa, vaso, mesa etc.) e virtuais (tubulação verde dentro da parede). Por favor, avalie cada um dos vídeos abaixo levando em consideração a percepção de profundidade. Considere 1 estrela sendo baixíssima coerência na relação de profundidade entre os objetos reais e virtuais e 5 sendo altíssima coerência na relação de profundidade entre os objetos reais e virtuais.". Os resultados em vídeo estão ilustrados a partir de um

quadro da cena na Figura 35.

Figura 36 - Quadros da cena de vídeo *outdoor* para aplicação de Arquitetura e Construção Civil com objetos salientes em movimento, com diferentes tamanhos e velocidades, utilizada no questionário de avaliação qualitativa. A ordem apresentada no questionário corresponde com a ordem apresentada na figura. Resultado obtido utilizando (Zollmann, et al., 2010) (a), resultado obtido utilizando (Padilha, et al., 2013) (b), resultado obtido utilizando a técnica proposta neste trabalho (c) e resultado obtido utilizando RA convencional (d).

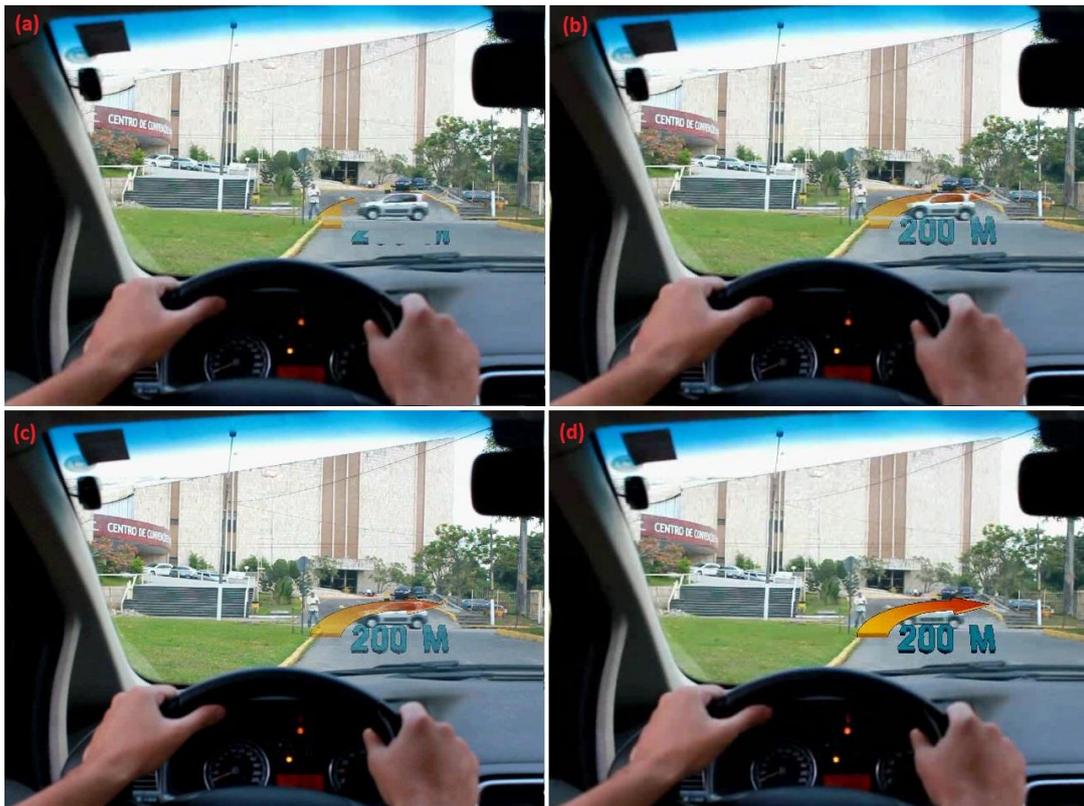


Para a segunda cena de vídeo voltada à comparação de resultados referentes à aplicação de Arquitetura e Construção Civil, a seguinte explicação foi fornecida para guiar a avaliação dos resultados: "Abaixo você verá vários resultados de técnicas visando melhorar a visualização de cenas de RA através do tratamento de situações de oclusão entre objetos reais (carros, motos, bicicletas, árvores etc.) e virtuais (tubulação verde no subsolo). Por favor, avalie cada um dos vídeos abaixo levando em consideração a percepção de profundidade. Considere 1 estrela sendo baixíssima coerência na relação de profundidade entre os objetos reais e virtuais e 5 sendo altíssima coerência na relação de profundidade

entre os objetos reais e virtuais.". Os resultados em vídeo estão ilustrados a partir de um quadro da cena na Figura 36.

As cenas seguintes foram voltadas para aplicações de *Safe Driving*, e o objetivo com o questionário foi avaliar se a direção se torna mais segura após o tratamento de oclusão ser utilizado na composição da cena de RA. A terceira cena de vídeo procura avaliar o comportamento das técnicas diante de um objeto saliente invadindo a área onde é projetada a aplicação de RA, enquanto que a quarta cena visa avaliar o comportamento das técnicas quando objetos com tamanhos diferentes e velocidades diferentes invadem a área onde é projetada a aplicação de RA.

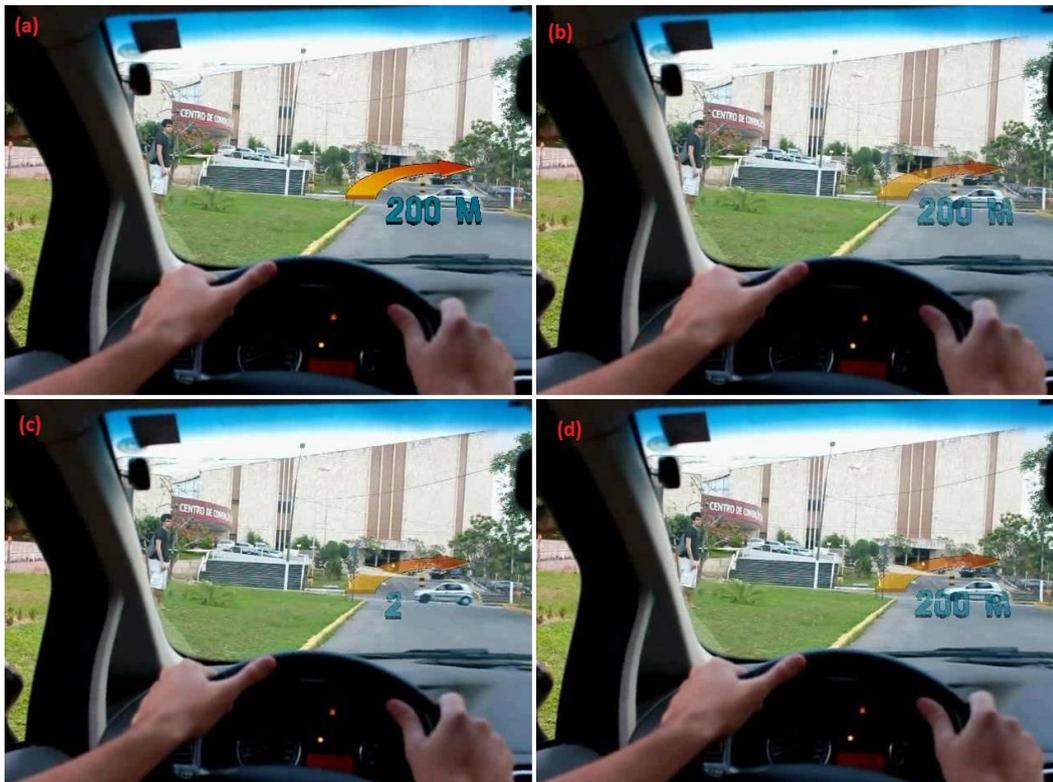
Figura 37 - Quadros da cena de vídeo *outdoor* para aplicação de *Safe Driving* com objetos não salientes e objetos salientes estáticos e em movimento utilizada no questionário de avaliação qualitativa. A ordem apresentada no questionário corresponde à ordem apresentada na figura. Resultado obtido utilizando a técnica proposta neste trabalho (a), resultado obtido utilizando (Padilha, et al., 2013) (b), resultado obtido utilizando (Zollmann, et al., 2010) (c), resultado obtido utilizando RA convencional (d).



Para a terceira cena de vídeo voltada à comparação de resultados

referentes à aplicação de *Safe Driving*, a seguinte explicação foi fornecida para guiar a avaliação dos resultados: "Abaixo você verá vários resultados de técnicas visando a utilização do sistema de RA, no carro, de modo que não atrapalhe a segurança do motorista na percepção de coisas importantes no trajeto. Objetos reais (pedestres, carros, bolas, bicicletas etc.) e virtuais (seta para direita e texto - 200 m). Por favor, avalie cada um dos vídeos abaixo levando em consideração que o motorista deve usufruir do sistema de realidade aumentada e ter segurança para dirigir. Considere 1 estrela sendo baixíssima segurança na direção e 5 sendo altíssima segurança na direção.". Os resultados em vídeo apresentados estão ilustrados a partir de um quadro da cena na Figura 37.

Figura 38 - Quadros da cena de vídeo *outdoor* para aplicação de *Safe Driving* com objetos não salientes e objetos salientes estáticos e em movimento utilizada no questionário de avaliação qualitativa. A ordem apresentada no questionário corresponde com a ordem apresentada na figura. Resultado obtido utilizando RA convencional (a), resultado obtido utilizando (Zollmann, et al., 2010) (b), resultado obtido utilizando a técnica proposta neste trabalho (c), resultado obtido utilizando (Padilha, et al., 2013) (d).

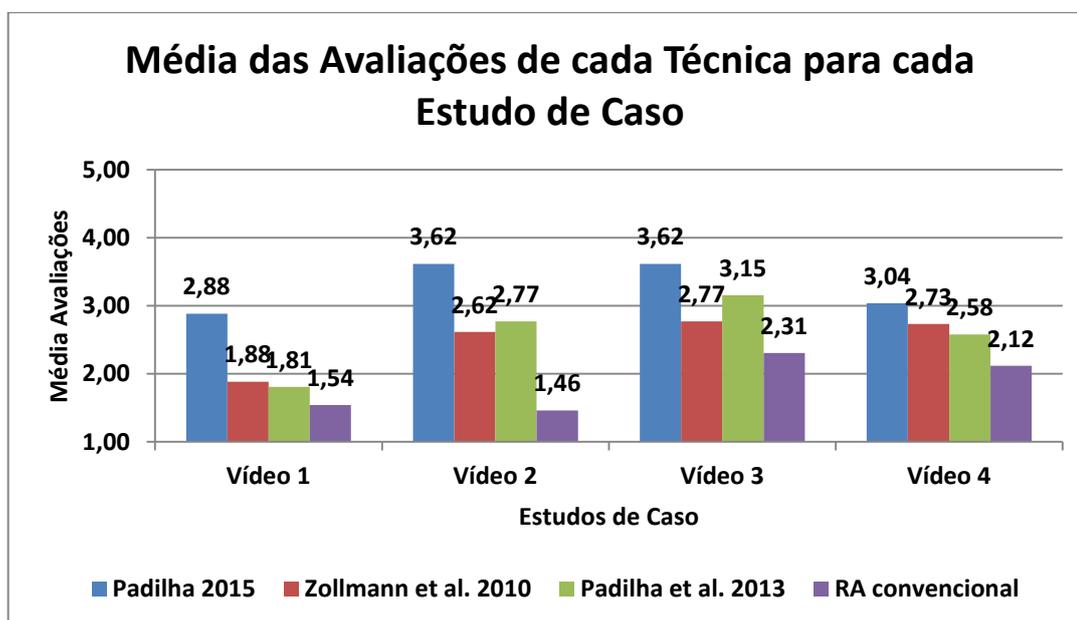


Para a quarta cena de vídeo voltada à comparação de resultados

referentes à aplicação de *Safe Driving*, a seguinte explicação foi fornecida para guiar a avaliação dos resultados: "Abaixo você verá vários resultados de técnicas visando a utilização do sistema de RA, no carro, de modo que não atrapalhe a segurança do motorista na percepção de coisas importantes no trajeto. Objetos reais (pedestres, carros, bolas, bicicletas etc.) e virtuais (seta para direita e texto - 200 m). Por favor, avalie cada um dos vídeos abaixo levando em consideração que o motorista deve usufruir do sistema de realidade aumentada e ter segurança para dirigir. Considere 1 estrela sendo baixíssima segurança na direção e 5 sendo altíssima segurança na direção.". Os resultados em vídeo apresentados estão ilustrados a partir de um quadro da cena na Figura 38.

Os resultados obtidos com as avaliações dos vídeos estão apresentados no gráfico da Figura 39.

Figura 39 - Gráfico com média das avaliações de cada técnica para cada cena de vídeo apresentada na Figura 35, na Figura 36, na Figura 37, e na Figura 38. Avaliação para técnica proposta neste trabalho em azul, avaliação para técnica proposta por Zollmann et al.



As cenas dos Vídeos 1 e 2 apresentaram, consideravelmente, avaliação melhor para a técnica proposta em relação às técnicas anteriores. Isso aconteceu, principalmente, porque a pergunta elaborada se refere à perspectiva de

profundidade. Visualizar o objeto aumentado ocludindo o objeto real saliente influenciou na perspectiva de profundidade dos usuários em algumas situações, o que comprova a necessidade de se utilizar técnicas de tratamento de oclusão em sistemas de RA, em especial, uma que considere diversos aspectos da atenção humana, como é o caso da técnica proposta no presente trabalho a qual adiciona a detecção de movimentos. Além disso, é possível perceber que, apesar de melhor avaliada, a técnica proposta no presente trabalho não obteve média próxima do valor máximo, isso ocorreu principalmente devido ao problema na segmentação do objeto em movimento, vide Figura 40. Como não houve aproximação com a silhueta dos objetos então muitos usuários consideraram que a visualização da cena de RA não está totalmente satisfatória.

Figura 40 - Falha na aproximação da área classificada como em movimento para a silhueta do objeto. Cena da aplicação de *Safe Driving* à esquerda e cena da aplicação de *Arquitetura e Construção Civil* à direita.



As cenas dos Vídeos 3 e 4 não apresentaram, consideravelmente, avaliação melhor para a técnica proposta em relação às técnicas anteriores. Apesar de a técnica proposta ter sido melhor avaliada em todos os casos de *Safe Driving*, muitas pessoas levaram em consideração que apenas o efeito de transparência permitiu uma direção mais segura e portanto as avaliações tiveram valores médios muito próximos.

5 CONCLUSÃO

Neste trabalho foi apresentada uma nova forma de utilizar a técnica de *Ghosting* com o objetivo de melhorar a percepção de profundidade durante a visualização de cenas de RA. Trabalhos anteriores com base na área de psicologia têm mostrado que a visão humana é sensível ao movimento em sequências de vídeo. Esta sensibilidade indica que o observador percebe o movimento como informação importante (saliente), e, por isso, este trabalho teve como objetivo adicionar ao *pipeline* da técnica de *Ghosting* uma técnica de detecção de saliência que reflete os efeitos da percepção visual humana em relação a movimentos e objetos naturalmente salientes.

A técnica de detecção de movimentos foi combinada com outras técnicas de visão computacional e processamento de imagens, como detecção de regiões de saliência e detecção de bordas, para obter o mapa de *ghosting*, responsável por descrever qual o nível de transparência que é usado em cada *pixel* durante o processo de composição da cena de RA.

Os resultados deste trabalho foram comparados com os resultados de trabalhos anteriores através de uma análise qualitativa realizada com diversos usuários, e mostraram uma melhoria na visualização da cena de RA com respeito à relação de profundidade entre os objetos. A técnica apresentada neste trabalho analisa várias características da cena real, principalmente os movimentos, para prover uma visualização mais agradável da cena segundo a tese de que as informações naturalmente salientes à visão humana, que estão presentes na cena real, não devem ser oclusas pelos objetos aumentados, pois carregam informação importante para o observador.

Foram definidos dois estudos de caso para avaliar este trabalho. O primeiro voltado para aplicação em construção civil ou arquitetura apresentando uma cena de RA com tubulações em ambientes internos (residência) e externos (ruas). O segundo foi voltado para aplicações de *safe driving*, onde a informação virtual é

adicionada ao para-brisa de um automóvel, e ela não deve obstruir a visualização do motorista em situações onde objetos salientes aparecem no seu campo de visão. Para tratar ambos os contextos, a técnica em questão apresentou resultados bastante satisfatórios segundo avaliação qualitativa realizada.

5.1 TRABALHOS FUTUROS

Trabalhos futuros vão melhorar a detecção dos objetos em movimento aplicando técnicas de segmentação para obter uma aproximação mais fiel à silhueta do objeto em movimento e assim evitar a existência de alguns artefatos na cena de RA como vistos em alguns estudos de caso que foram apresentados.

Melhorar a robustez a mudanças de iluminação e analisar o impacto das sombras dos objetos, para a percepção de profundidade da cena de RA, utilizando abordagens semelhantes à descrita em (Lu, et al., 2010).

Considerar a informação de profundidade na composição do mapa de *ghosting* utilizando técnicas de detecção de profundidades da cena a partir de sensores RGB como descrito em (Eigen, et al., 2014) e a partir de sensores RGB-D como descrito em (Tian, et al., 2015).

Além disso, todas as técnicas de visão computacional e processamento de imagem utilizadas neste *pipeline* de *Ghosting* serão paralelizadas usando a GPU.

5.2 CONTRIBUIÇÕES E PUBLICAÇÕES DIRETAMENTE RELACIONADAS À ESTE TRABALHO

A principal contribuição deste trabalho está na técnica de *Ghosting* baseada na detecção de movimentos que considera aspectos da percepção humana que foi proposta. Além disso, estudos de caso na área de Construção Civil e Arquitetura e *Safe Driving* também foram propostos.

Modificações realizadas na detecção de saliência estática no *pipeline* da

técnica de *Ghosting* resultaram na publicação (Padilha, et al., 2013). Além disso, resultados parciais utilizando a detecção de movimentos no *pipeline* da técnica de *Ghosting* foram publicados em (Padilha, et al., 2014). Este trabalho está em revisão para publicação (Padilha, et al., 2015).

REFERÊNCIAS

- Avaliação Qualitativa dos Resultados. 15 de março de 2015. Disponível em: <http://form.jotformz.com/form/50548551706659>.
- ACHANTA R. [et al.] **Frequency-tuned Salient Region Detection**. IEEE Conference on Computer Vision and Pattern Recognition. 2009. pp. 1597-1604.
- AVERY B., SANDOR C. e THOMAS B. **Improving spatial perception for augmented reality x-ray vision** In Proceedings of the IEEE Virtual Reality Conference. 2009. pp. 79–82.
- BANE R e HOELLERER T. **Interactive Tools for Virtual X-Ray Vision in Mobile Augmented Reality**. International Symposium on Mixed and Augmented Reality. 2004. pp. 231-239.
- BERNING M. [et al.] **A Study of Depth Perception in Hand-Held Augmented Reality using Autostereoscopic Displays**. IEEE International Symposium on Mixed and Augmented Reality (ISMAR). 2014.
- CANNY J. **A Computational Approach to Edge Detection**. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1986. v. PAMI-8. pp. 679-698.
- CHEN J. [et al.] **On-Line Visualization of Underground Structures using Context Features**. ACM Symposium on Virtual Reality Software and Technology. 2010. pp. 167-170.
- CHENG M. [et al.] **Global Contrast based Salient Region Detection**. IEEE Conference on Computer Vision and Pattern Recognition. 2011. pp. 409-416.
- CUTTING J. **Reconceiving Perceptual Space, in Perceiving Pictures: An Interdisciplinary Approach to Pictorial Space**. MITPress. 2003. pp. 215-238.
- EIGEN D., PUHRSCH C. e FERGUS R. **Depth Map Prediction from a Single Image using a Multi-Scale Deep Network**. CoRR. 2014. v. abs/1406.2283.
- FEINER S. K. e SELIGMANN D. D. **Cutaways and Ghosting: Satisfying Visibility Constraints in Dynamic 3D Illustrations**. Visual Computer. 1992. v. 8. pp. 292-302.
- FREUND Y. e SCHAPIRE R. E. **A Decision-Theoretic Generalization of On-Line Learning and an**

Application to Boosting. Journal of Computer and System Sciences. 1995. v. 55. pp. 119-139.

HUNTER R. S. **Photoelectric Color-Difference Meter.** Proceedings of the Winter Meeting of the Optical Society of America. 1948.

ITTI L. **Automatic foveation for video compression using a neurobiological model of visual attention.** IEEE Transaction Image Processing. 2004. v. 13. pp. 1304-1318.

KALKOFEN D. [et al.] **Adaptive Ghosted Views for Augmented Reality.** IEEE International Symposium on Mixed and Augmented Reality. Adelaide, Austrália. 2013.

KALKOFEN D. [et al.] **Visualization Techniques for Augmented Reality.** Handbook of Augmented Reality. Springer Science + Business Media, 2011.

KALKOFEN D., MENDEZ E. e SCHMALSTIEG D. **Comprehensible visualization for augmented reality.** IEEE Transactions on Visualization and Computer Graphics. 2009. pp. 193-204.

KALKOFEN D., TATZGERN M. e SCHMALSTIEG D. **Explosion Diagrams in Augmented Reality.** Virtual Reality. 2009. pp. 71-78.

KRUIJFF E., SWAN II J. E. e FEINER S. **Perceptual Issues in Augmented Reality Revisited.** IEEE International Symposium on Mixed and Augmented Reality. 2010. pp. 3-12.

LI W. [et al.] **Interactive cutaway illustrations of complex 3D models.** ACM Transactions on Graphics. 2007. p. 31.

LI Z., QIN S. e ITTI L. **Visual attention guided bit allocation in video compression.** Image and Vision Computing. 2011. v. 29. pp. 1-4.

LIKERT R. **A Technique for the Measurement of Attitudes.** Archives of Psychology. 1932. v. 140. pp. 1-55.

LIVINGSTON M. A. [et al.] **Resolving Multiple Occluded Layers in Augmented Reality.** International Symposium on Mixed and Augmented Reality. Tokyo, 2003.

LU B. V. [et al.] **Foreground and Shadow Occlusion Handling for Outdoor Augmented Reality.** International Symposium on Mixed and Augmented Reality. Seol, Korea, 2010.

- OH H. e KIM W. **Video Processing for Human Perceptual Visual Quality-Oriented Video Coding**. IEEE Transactions on Image Processing. 2013. v. 22. pp. 1526 - 1535.
- PADILHA A. e TEICHRIEB V. **Motion Detection Based Ghosted Views for Occlusion Handling in Augmented Reality**. In Review IEEE International Symposium on Mixed and Augmented Reality (ISMAR). Fukuoka, Japan, 2015.
- PADILHA A. e TEICHRIEB V. **[Poster] Motion Detection Based Ghosted Views for Occlusion Handling in Augmented Reality**. IEEE International Symposium on Mixed and Augmented Reality (ISMAR). Munich, Germany, 2014. pp. 291 - 292.
- PADILHA A., ROLIM C. e TEICHRIEB V. **The Ghosting Technique Applied to Augmented Reality Visualization**. XV Symposium on Virtual and Augmented Reality (SVR). Cuiaba, 2013. pp. 159 - 166.
- PORTER T. e DUFF T. **Compositing Digital Images**. Computer Graphics. 1984. pp. 253-259.
- PRAUN E. [et al.] **Real-Time Hatching**. Proceedings SIGGRAPH. 2001. pp. 579-584.
- SANDOR C. [et al.] **An Augmented Reality X-Ray System Based on Visual Saliency**. IEEE International Symposium on Mixed and Augmented Reality. 2010. pp. 27-36.
- TIAN Y. [et al.] **Handling occlusions in augmented reality based on 3D reconstruction method**. **Neurocomputing**. 2015. 10.1016/j.neucom.2014.12.081. pp. 96-104.
- TREISMAN A. M. e GELADE G. **A feature-integration theory of attention**. **Cognitive Psychology**. 1992. v. 12. pp. 127-132.
- WALTHER D. **Interactions of Visual Attention and Object Recognition: Computational Modeling, Algorithms, and Psychophysics**. California Institute of Technology, Ph.D. Thesis. 2006.
- WITHER J. [et al.] **Fast Annotation and Modeling with a Single-Point Laser Range Finder**. IEEE International Symposium on Mixed and Augmented Reality. Cambridge, UK, 2008.
- ZOLLMANN S. [et al.] **Image-Based Ghostings For Single Layer Occlusions In Augmented Reality**. IEEE International Symposium on Mixed and Augmented Reality. 2010. pp. 19-26.