UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA

OSVALDO RAMOS TSAN HU

Contribuições ao desenvolvimento de um sistema de
telepresença por meio da aquisição, transmissão e projeção em
ambientes imersivos de vídeos panorâmicos

São Paulo
2006

OSVALDO RAMOS TSAN HU

Contribuições ao desenvolvimento de um sistema de
telepresença por meio da aquisição, transmissão e projeção em
ambientes imersivos de vídeos panorâmicos

Tese apresentada à Escola Politécnica da
Universidade de São Paulo para obtenção do Título
de Doutor em Engenharia Elétrica.

Área de Concentração: Sistemas Eletrônicos

Orientadora: Profa. Dra. Roseli de Deus Lopes

São Paulo
2006

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTES
TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA
FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Este exemplar foi revisado e alterado em relação à versão original,
sob responsabilidade única do autor e com a anuência de seu
orientador.

São Paulo, 01 de agosto de 2006.

Assinatura do autor

Assinatura do orientador

FICHA CATALOGRÁFICA

Hu, Osvaldo Ramos Tsan
Contribuições ao desenvolvimento de um sistema de
telepresença por meio da aquisição, transmissão e projeção
em ambientes imersivos de vídeos panorâmicos / O.R.T.
Hu. -- São Paulo, 2006.
128 p.

Tese (Doutorado) - Escola Politécnica da Universidade de
São Paulo. Departamento de Engenharia de Sistemas
Eletrônicos.

1.Telepresença 2.Realidade virtual 3.Vídeo panorâmico 4.Ca-
verna digital. Universidade de São Paulo. Escola Politécnica.
Departamento de Engenharia de Sistemas Eletrônicos II.t.

À Maria do Carmo, que me ensinou o significado da palavra amor.
Ao Yuri e à Natasha, que me ensinaram o significado da palavra carinho.
Aos meus pais, que me ensinaram o significado da palavra saudade.

AGRADECIMENTOS

Gostaria de agradecer à Profa. Dra. Roseli de Deus Lopes, minha orientadora, pela confiança em mim depositada, pela paciência com que me orientou, pelos conselhos que ofereceu durante esta fase de minha vida e principalmente pela amizade, que sempre permeou o nosso contato acadêmico.

Ao Prof. Dr. Marcelo Knorich Zuffo, coordenador dos projetos da CAVERNA Digital®* da Escola Politécnica da USP, que foi um co-orientador de fato, apoiando-me e fornecendo recursos para a elaboração de meu projeto.

Ao Dr. Celso Setsuo Kurashima, com quem desenvolvi parte do projeto do anel de câmeras, que sempre me incentivou a melhorá-lo, fornecendo críticas e sugestões.

Aos pesquisadores Marcio Calixto Cabral, Olavo da Rosa Belloc e Victor H. P. Gomes, que ofereceram suporte técnico no manejo dos equipamentos da CAVERNA Digital® da EPUSP e também auxiliaram na programação dos módulos que compõem este projeto. Ao pesquisador Dr. Luciano Pereira Soares, pelo apoio no início do desenvolvimento do projeto. Ao pesquisador Hilton Garcia Fernandes, pela sugestão de biblioteca de compressão JPEG e pelo apoio na implementação das rotinas da biblioteca ezJ.

Aos demais colegas do Laboratório de Sistemas Integráveis da EPUSP, em especial aos da CAVERNA Digital® e do NATE (Núcleo de Aprendizagem, Trabalho e Entretenimento), pela ajuda em muitas das etapas da elaboração do projeto.

Por fim, mas não menos importante, agradeço à Maria do Carmo, ao Yuri e à Natasha, pelo apoio e pela paciência com que me agraciaram durante toda esta empreitada.

* O LSI-EPUSP projetou, construiu e mantém o primeiro sistema de Realidade Virtual (RV) com múltiplas projeções estereoscópicas, da América Latina, denominada CAVERNA Digital®.

FINANCIAMENTO

O desenvolvimento deste projeto de pesquisa foi possível graças à sinergia com outras atividades desenvolvidas, bem como à infra-estrutura disponibilizada, nos seguintes projetos,:

- SISCOMPRO: Sistemas de Comunicação e Processamento Avançados e Aplicações Interativas, FINEP Convênio 01.02.0288.00. Ref. 2430/02;
- GigaVR: Plataformas para o Desenvolvimento de Aplicações de Realidade Virtual Imersiva e Distribuída sobre Redes de Altíssima Velocidade, Editais RNP (RNP, 2004A) e (RNP, 2004B);
- Programa Educacional da INTEL do Brasil, que doou os computadores utilizados no aglomerado de computadores utilizado no projeto.

RESUMO

Sistemas de telepresença têm sido pesquisados e desenvolvidos para inúmeras aplicações que exigem a presença física de pessoas em ambientes inacessíveis; tais situações são diversas, desde aquelas relacionadas com educação a distância até aquelas que envolvem alta periculosidade. Neste trabalho, a pesquisa e desenvolvimento se concentram na concepção e implementação de sistemas de telepresença voltados para ambientes imersivos em 360°, capazes de realizar a aquisição, transmissão e projeção de imagens em movimento em sistemas de multiprojeção imersivos, como é o caso da CAVERNA Digital®, desenvolvida pelo Laboratório de Sistemas Integráveis da Escola Politécnica da USP. Assim, o presente trabalho apresenta contribuições para o desenvolvimento de sistemas de telepresença, dentre os quais destacam-se: o detalhamento da arquitetura geral do sistema, a implementação de métodos para a calibração, correção das imagens e montagem de panoramas de 360°. Construiu-se um protótipo composto de: um Módulo de Aquisição, que adquire as imagens de oito câmeras (montadas num anel de câmeras sobre um robô), efetua as correções e monta uma imagem panorâmica; um Módulo de Composição que costura as várias imagens panorâmicas em um panorama final; e um Módulo de Exibição que ajusta e projeta o panorama nas telas da CAVERNA Digital®. Finalmente, apresentam-se considerações sobre o presente trabalho e perspectivas futuras.

Palavras-Chave: Telepresença. Realidade Virtual (RV). Vídeo Panorâmico. Ambiente Imersivo. Multiprojeção. Estereoscopia.

ABSTRACT

Research and development in telepresence systems have been done in several applications that require physical presence of people into non-accessible environments. These situations may vary, from those related to distance education to those related to very hazardous places for humans. In this work, the main research and development goal is the conception and implementation of a telepresence system for 360 degrees immersive environments, this system is able to perform acquisition, transmission and projection of moving images on immersive multi-projection environments, such as the CAVERNA Digital®, developed by the Laboratório de Sistemas Integráveis at the Escola Politécnica, Universidade de São Paulo. The main contributions to the development of telepresence systems are: a detailed specification of the general architecture of the system, the implementation of the calibration, the imagery correction and the 360 degrees panoramas composition methods. The prototype that was implemented includes: an Acquisition Module, that acquires image of eight cameras (mounted on a ring of cameras placed on top of a robot), executes corrections and prepares a panoramic image; a Composition Module, that stitches the images in a final panorama; and an Exhibition Module, that adjusts and projects the panorama into the screens of the CAVERNA Digital®. Final, remarks on this present work and future perspectives are presented at the end.

Key-words: Telepresence. Virtual Reality. Panoramic Video. Immersive Environments. Multi-projection. Stereoscopy.

SUMÁRIO

1	INTRODUÇÃO	1
1.1	Contexto	1
1.2	Objetivos	4
1.3	Motivação	5
1.4	Relevância	6
1.5	Atividades	7
1.6	Trabalhos correlatos	7
1.7	Estrutura da tese	9
2	CONCEITOS E TECNOLOGIAS EM IMAGENS PANORÂMICAS	11
2.1	Introdução	11
2.2	Fotos panorâmicas	11
2.3	Fotos estereoscópicas	19
2.4	Fotos panorâmicas e estereoscópicas	22
2.5	Imagens analógicas e digitais	24
2.5.1	Imagens analógicas	24
2.5.2	Imagens digitais	25
2.6	Correção e projeção de imagens	26
2.6.1	Características de lentes	27
2.6.2	Distorções de lentes de câmeras	28
2.6.3	Correção de distorções de lentes de câmeras	31
2.6.4	Projeção cilíndrica	34
2.6.5	Conversão de imagem projetada cilindricamente em imagem plana	37
2.6.6	Alinhamento de imagens e transformação de perspectiva	38
2.7	Compressão de imagens	40
2.7.1	Compressão JPEG	40
2.7.2	Compressão MJPEG	46

2.7.3	Compressão MPEG	46
2.8	Ambientes avançados	48
2.8.1	Ambientes de realidade virtual (RV)	48
2.8.2	Comunicação entre microcomputador e periféricos	51
2.8.3	Comunicação de alta velocidade - Gigabit Ethernet	52
2.8.4	Comunicação de alta velocidade - Internet 2	53
2.8.5	Processamento de alto desempenho – Aglomerado de Computadores	55
2.9	Conclusão	56

3 PROPOSTA DE UM SISTEMA DE TELEPRESENÇA EM AMBIENTES

IMERSIVOS DE VÍDEOS PANORÂMICOS

3.1	Introdução	57
3.2	Requisitos e restrições	57
3.3	Previsão de recursos adicionais	59
3.4	Alternativas estudadas	59
3.5	Arquitetura de <i>hardware</i>	62
3.6	Arquitetura de <i>software</i>	63
3.6.1	Módulo de Aquisição	64
3.6.1.1	Procedimento de Calibração de Imagens	65
3.6.1.2	Procedimento de Montagem do Panorama	68
3.6.2	Módulo de Composição	71
3.6.3	Módulo de Exibição	71
3.7	Conclusão	72

4 IMPLEMENTAÇÃO E TESTE COM IMAGENS PANORÂMICAS

ESTÁTICAS

4.1	Introdução	73
4.2	Aquisição de imagens	73
4.3	Montagem do panorama	76
4.4	Projeção de imagens	79

4.5	Conclusão	81
5	IMPLEMENTAÇÃO E TESTES COM VÍDEOS PANORÂMICOS	82
5.1	Introdução	82
5.2	Equipamento de aquisição de vídeo	82
5.3	Aglomerado de computadores e rede	85
5.4	Módulo de Aquisição	88
5.4.1	Determinação das constantes de lente	88
5.4.2	Procedimento de Calibração de Imagens	89
5.4.2.1	Correção de distorções	90
5.4.2.2	Projeção cilíndrica	94
5.4.2.3	Alinhamento de imagens	96
5.4.3	Procedimento de Montagem do Panorama	100
5.4.3.1	Leitura de câmeras	100
5.4.3.2	Localização de coordenadas dos pixels no Mapa de Posicionamento	100
5.4.3.3	Rotação em 90° e costura de imagens	101
5.4.3.4	Compressão de imagens	103
5.4.3.5	Transmissão de imagens	104
5.5	Módulo de Composição	106
5.5.1	Recepção de imagens	106
5.5.2	Descompressão de imagens	107
5.5.3	Alinhamento e costura de imagens	107
5.5.4	Compressão e retransmissão de imagens	109
5.6	Conclusão	109
6	DISCUSSÃO DOS RESULTADOS	111
6.1	Introdução	111
6.2	Ajustes finais dos programas	111
6.3	Montagem dos equipamentos para os testes	113
6.4	Resultados obtidos	114

6.5	Conclusão	117
7	CONCLUSÕES E TRABALHOS FUTUROS	119
7.1	Introdução	119
7.2	Considerações sobre o trabalho	119
7.3	Contribuições	120
7.4	Trabalhos futuros	121
7.5	Consideração Final	122
	REFERÊNCIAS	123

LISTA DE FIGURAS

Figura 1: Reunião com presença virtual (WALKER; SHEPPARD, 1997)	4
Figura 2: Projeto GigaVR	8
Figura 3: Anel de câmeras (KURASHIMA, 2005)	9
Figura 4: Panorama de Chicago antes do incêndio (CURTIN, 2001)	12
Figura 5: Long Beach, Califórnia, Bathing Beauty Parade, 1927 (CURTIN, 2001)	12
Figura 6: Sequência de fotos com rotação da câmera (CURTIN, 2001)	13
Figura 7: Máquina panorâmica da Kodak (CURTIN, 2001)	13
Figura 8: Câmeras fotográficas panorâmicas (RIGG, 2001)	14
Figura 9: Máquina panorâmica giratória (CURTIN, 2001)	15
Figura 10: Máquina fotográfica com dois espelhos (CURTIN, 2001)	15
Figura 11: Câmeras de vídeo catadióptricas omnidirecionais (NAYAR, 1997)	16
Figura 12: Imagem panorâmica com projeção esférica (RIGG, 2001)	16
Figura 13: Imagem gerada por projeção esférica e corrigida (RIGG, 2001)	17
Figura 14: Diferentes homogeneidades dos pixels (NAYAR, 1997)	18
Figura 15: <i>Daguerreotype</i> estéreo (CURTIN, 2001)	19
Figura 16: Visualizadores e câmera estereoscópica (CURTIN, 2001)	20
Figura 17: Câmeras e dispositivos atuais (CURTIN, 2001)	20
Figura 18: Possíveis configurações de câmaras estereoscópicas	21
Figura 19: Imagem obtida pela PanDC® (CURTIN, 2001)	22
Figura 20: Câmera do tipo <i>Manifolds</i> capturando tiras estereoscópicas (PELEG; BEN-EZRA, 2001)	23
Figura 21: Duas câmeras omnidirecionais para captura de panoramas estereoscópicos (SHIMAMURA et al., 2000)	23

Figura 22: Uma imagem e seus pixels	26
Figura 23: Diagrama esquemático das principais características de uma lente	28
Figura 24: Vetores de deslocamento de distorção radial (BOUGUET, 2004)	30
Figura 25: Vetores de deslocamento de distorção tangencial (BOUGUET, 2004)	30
Figura 26: Cálculo da distância radial entre o pixel e o ponto principal	32
Figura 27: Projeção cilíndrica de um conjunto de imagens	34
Figura 28: Projeção da imagem plana no cilindro	34
Figura 29: Mudança nas coordenadas da imagem	35
Figura 30: Vista superior do cilindro e do plano da imagem	36
Figura 31: Vista superior do cilindro e do plano da imagem com X' e Z'	36
Figura 32: Projeção plana do panorama recebido	37
Figura 33: Projeção da imagem plana no cilindro	37
Figura 34: Vista superior do cilindro e do plano da imagem com X' e Z'	38
Figura 35: Diagrama do método de compressão JPEG	41
Figura 36: Exemplo de matrizes (a) de DCT, (b) de quantização, e o (c) resultado da divisão entre as duas (TANENBAUM, 1997)	44
Figura 37: Leitura em zigzag	45
Figura 38: Quadros no padrão H.261 (Rodriguez et al., 1999)	47
Figura 39: Tipos de quadros do padrão MPEG (Rodriguez et al., 1999)	48
Figura 40: Cave (CRUZ-NEIRA, 2006)	49
Figura 41: Maquete virtual da CAVERNA Digital® da EPUSP	49
Figura 42: CAVERNA Digital® da EPUSP	50
Figura 43: <i>Backbone</i> (barramento) da rede brasileira RNP2 (RNP, 2005)	54
Figura 44: Diagrama esquemático de um sistema de telepresença	58
Figura 45: Diferentes homogeneidades dos pixels (NAYAR, 1997)	60

Figura 46: Ângulo de tomada e zonas mortas	61
Figura 47: Ângulo de tomada e zonas de sobreposição	61
Figura 48: Diagrama esquemático dos equipamentos	62
Figura 49: Diagrama esquemático dos módulos de <i>software</i>	64
Figura 50: Seqüência de correções no Mapa de Posicionamento	66
Figura 51: Seqüência completa de montagem do Mapa de Posicionamento	66
Figura 52: Matrizes de imagens na memória dos microcomputadores	68
Figura 53: Localização de coordenadas no Mapa de Posicionamento	69
Figura 54: Rotação de imagens na construção do vídeo panorâmico	70
Figura 55: Máquinas montadas sobre dispositivo e tripé	74
Figura 56: Ilustração do mecanismo de deslocamento	75
Figura 57: Fotografias costuradas gerando um panorama de 360° da CAVERNA Digital®	75
Figura 58: Série de fotografias sem costura	76
Figura 59: Fotografias costuradas gerando um panorama de 360°	76
Figura 60: Série de fotos adquiridas da Sala São Paulo	77
Figura 61: Imagem montada em panorama	77
Figura 62: Série de fotos adquiridas da Sala São Paulo	78
Figura 63: Imagem montada da Sala São Paulo	79
Figura 64: Esquema ilustrando a Sala São Paulo em projeção cilíndrica	80
Figura 65: Esquema ilustrando a Sala São Paulo na CAVERNA Digital® da EPUSP	81
Figura 66: Diagrama esquemático dos anéis de câmeras, com doze e com oito câmeras	83
Figura 67: Anel de doze câmeras montadas em um tripé	83
Figura 68: Câmera <i>Dragonfly</i> (POINT GREY RESEARCH INC, 2004)	85
Figura 69: <i>Hub firewire</i>	86
Figura 70: Aglomerado de computadores com os equipamentos fora da estante (<i>rack</i>)	87

Figura 71: Imagens de um tabuleiro de xadrez, utilizadas na calibração da câmera	88
Figura 72: Painel do Módulo de Aquisição	89
Figura 73: Mapa de posicionamento inicial	91
Figura 74: Mapa de Posicionamento com a correção de distorções da lente	92
Figura 75: Imagens capturadas por duas câmeras	93
Figura 76: Imagens sem correção, rotacionadas em 90° para melhor visualização	93
Figura 77: Imagens corrigidas, rotacionadas em 90° para melhor visualização	94
Figura 78: Mapa de Posicionamento, corrigida a distorção da imagem e projetada cilindricamente	95
Figura 79: Imagens corrigidas e projetadas cilindricamente	95
Figura 80: Imagens costuradas sem terem sido previamente projetadas cilindricamente	96
Figura 81: Pontos equivalentes definidos (oito pontos)	97
Figura 82: Mapa de Posicionamento, após o alinhamento	98
Figura 83: Segunda imagem alinhada	99
Figura 84: Matriz de posicionamento, rotacionada em 90°	101
Figura 85: Montagem final da imagem panorâmica	102
Figura 86: Imagem panorâmica, com imagens corrigidas, projetadas, alinhadas e costuradas	103
Figura 87: Principais rotinas do WinSock 2 utilizadas no sistema	105
Figura 88: Montagem do quadro final	107
Figura 89: Costura final	108
Figura 90: Painel de controle do Módulo de Aquisição	112
Figura 91: Painel de controle do Módulo de Composição	112
Figura 92: Robô com o anel de câmeras no IEEE VR2006	114
Figura 93: Painel do Módulo de Composição mostrando a velocidade de recepção	115

Figura 94: Painel do Gerenciador de Tarefas do Windows, durante a execução do Módulo de
Aquisição _____ 116

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
ACM	Association for Computing Machinery
Bit	Menor unidade de informação computacional. É representado pelos algarismos zero ou um
<i>BitMap</i>	Mapa de bits. Formato de arquivo de imagem
Byte	Conjunto de oito bits
CAVE	<i>Cave Automatic Virtual Environment</i> (ambiente de realidade virtual do tipo multiprojeção, envolvendo os usuários)
CCD	<i>Charge Coupled Device</i> (dispositivo de carga acoplada). Dispositivo usado para captação de imagens em uma câmera digital
CODEC	Codificador e decodificador de vídeo ou de áudio
CPU	<i>Central Processing Unit</i> (unidade de processamento central)
DCT	<i>Discrete Cosine Transform</i> (transformada discreta do coseno)
DVD	<i>Digital Video Disk</i> (disco de vídeo digital)
EPUSP	Escola Politécnica da USP
fps	<i>Frames per second</i> (quadros por segundo)
<i>Firewire</i>	Barramento externo, também conhecido como i.Link ou IEEE1394
GigaVR	Plataformas para o Desenvolvimento de Aplicações de Realidade Virtual Imersiva e Distribuída sobre Redes de Altíssima Velocidade – projeto em desenvolvimento no LSI-EPUSP
H.261	Método de compressão de som e imagem utilizado para videoconferência
IEEE	Institute of Electrical and Electronic Engineers
ISO	International Organization for Standardization

JPEG	Joint Photographic Expert Group
LARC	Laboratório de Arquitetura e Redes de Computadores da EPUSP
LSI-EPUSP	Laboratório de Sistemas Integráveis da EPUSP
Mbps	Mega bits por segundo. Unidade utilizada na transmissão de dados
MIMD	<i>Multiple Instruction Multiple Data</i> (múltiplas instruções, múltiplos dados)
MIPS	Milhões de instruções por segundo
MISD	<i>Multiple Instruction Single Data</i> (múltiplas instruções, dado único)
MPEG	Moving Picture Experts Group
OpenCV	<i>Open Computer Vision Library</i> (biblioteca aberta de visão computacional)
OpenGL	<i>Open Graphic Library</i> (biblioteca gráfica aberta)
Pixel	<i>Picture Element</i> (elemento de imagem). Menor elemento de uma imagem
REMAVS	Redes Metropolitanas de Alta Velocidade
RA	Realidade Aumentada
RGB	<i>Red, Green, Blue</i> (Vermelho, Verde, Azul)
RNP	Rede Nacional de Pesquisa
RV	Realidade Virtual
SID	<i>Spatially Immersive Displays</i> (dispositivos de imersão espacial)
SISCOMPRO	Sistemas de Comunicação e Processamento Avançados e Aplicações Interativas, projeto em desenvolvimento no LSI-EPUSP
SIMD	<i>Single Instruction Multiple Data</i> (única instrução, múltiplos dados)
SISD	<i>Single Instruction Single Data</i> (única instrução, único dado)
String	Seqüência de bytes
TCP/IP	<i>Transmission Control Protocol/Internet Protocol</i> (protocolo de controle de transmissão / protocolo de Internet)
USB	<i>Universal Serial Bus</i> (Barramento Serial Universal)

USB-IF	<i>USB Implementers Forum</i> . Grupo de pesquisa, desenvolvimento e padronização do barramento USB
USP	Universidade de São Paulo
VOR	Veículos de Operação Remota
VTK	<i>The Visualization ToolKit</i> – Biblioteca gráfica com recursos para manipulação de imagens e dados volumétricos

NOTAÇÃO

Termos em inglês, para os quais não se encontraram traduções diretas até o momento da escrita da tese, são apresentados em itálico:

Exemplo: *pipeline*

Termos em inglês são apresentados em itálico, com a tradução em português colocada entre parênteses, na primeira ocorrência no texto.

Exemplo: *cluster* (aglomerado de computadores)

Quando se define um termo, este é apresentado em negrito:

Exemplo: **Telepresença** é uma forma de telecomunicação que possibilita a uma pessoa interagir a distância com outra pessoa ou objeto real, com a sensação de estar presente no local (WALKER; SHEPPARD, 1997).

1 INTRODUÇÃO

1.1 Contexto

Vivemos numa época denominada “era da informação”, em que o mundo está ligado por telefonia e Internet, e na qual é possível a comunicação entre pessoas ou objetos a longa distância. Entretanto, a sensação de proximidade e de presença proporcionada pelas soluções tecnológicas disponíveis ainda é limitada.

Às vezes é desejável mais do que um contato auditivo ou visual. Em certas tarefas, ainda é necessária a presença física de seres humanos no local, para que se possa ter completo entendimento do problema a ser resolvido e das dificuldades que o ambiente oferece, por meio da observação completa e tridimensional.

Por exemplo, na prospecção petrolífera em águas profundas (*offshore*), muitas vezes se faz necessária a presença de técnicos para vistoriar, instalar, treinar, simular e operar em grandes profundidades. Os submarinos que os levam para os locais de intervenção são equipamentos complexos e caros e a operação pode colocar-lhes em risco a vida. Outra possibilidade é a utilização de Veículos de Operação Remota (VOR), que são equipamentos controlados remotamente que, no entanto, por apresentarem limitações no campo de visão, tem restrições ou dificuldades em algumas operações.

Por outro lado, os custos dos equipamentos e das linhas de Internet de alta velocidade estão caindo e a velocidade de processamento e de transmissão de dados está em ascensão. Começa a tornar-se viável a construção de equipamentos e *software* que possibilitem a transmissão e o processamento de informações que melhorem a sensação de presença nas mais diversas áreas de aplicação.

Matthew Lombard (1997) define cinco formas de estar presente:

- Presença para o enriquecimento social. A presença física permite a sociabilização entre as pessoas. Crianças desenvolvem-se melhor na presença de outras crianças e de adultos, e aprendem outras formas de comunicação, como as simbólicas e as gestuais;
- Presença como realismo. A presença permite melhor compreensão dos fatos e das pessoas, o que sem ela não se poderia conseguir;
- Presença como transporte. Envolve a idéia de deslocamento virtual da pessoa. Pode ter as seguintes variações:
 - “Você está lá”. O observador se transporta para determinado local. A televisão permite este tipo de presença;
 - “Lá está aqui”. Comum em crianças pequenas que, ao assistirem à televisão, tentam tocar os objetos apresentados;
 - “Nós estamos juntos”. Típico de teleconferências e reuniões virtuais. Neste caso também não existe a presença física, mas as pessoas compartilham um mesmo espaço virtual;
- Presença como imersão. O conceito baseia-se na idéia de percepção e de imersão psicológica. Não se podem considerar apenas experiências com realidade virtual (RV), mas também cinema do tipo IMAX®, planetários e simulações diversas;
- Presença como ator virtual. A pessoa está envolvida em uma experiência de realidade virtual (RV), na qual pode interagir com o meio e dele obter retorno, atuando como se estivesse vivendo aquela situação e dirigindo suas ações e comportamentos.

Telepresença é uma forma de telecomunicação que possibilita a uma pessoa interagir a distância com outras pessoas ou objetos reais, com a sensação de estar presente no local contatado (WALKER; SHEPPARD, 1997); em outras palavras, é o termo utilizado para designar a presença virtual de uma ou mais pessoas em outro local, que pode estar afastado ou ser inacessível fisicamente.

Os principais recursos que podem ser utilizados atualmente para provocar a sensação de telepresença nas pessoas são:

- Vídeo: Consiste na transmissão de um sinal de vídeo ao vivo. Esse sinal deve ter velocidade superior a 15 quadros por segundo para não se perceber descontinuidade entre as imagens;
- Áudio: Para obter maior imersão é necessário capturar e transmitir os sons de modo estereofônico, de preferência com vários canais de áudio;
- Visão panorâmica: Uma câmera convencional captura um ângulo relativamente pequeno (por volta de 60°, dependendo da lente utilizada), e não causa boa sensação de proximidade, por não fornecer visão lateral nem posterior. O ideal é adquirir um vídeo com tomadas em 360°;
- Informações complementares: Podem-se transmitir dados para complementar as informações visuais e auditivas e melhorar a comunicação entre as pessoas, como textos, fotos, tabelas ou gráficos;
- Sinais de controle: Sinal de sincronismo, comandos para o controle remoto de máquinas e equipamentos e sinal de retorno com os parâmetros desses equipamentos.

Vislumbra-se grande quantidade de aplicações para a telepresença, como jogos interativos com inúmeros jogadores, os *multi-player games* (jogos de múltiplos jogadores), telemedicina, teleeducação, teleexploração de locais inóspitos e reuniões de negócios. Com a telepresença será possível, ainda, fazer reuniões virtuais com pessoas separadas por milhares de quilômetros, permitindo a todas a sensação de estarem sentadas à mesma mesa de reunião (Figura 1).

Figura 1: Reunião com presença virtual (WALKER; SHEPPARD, 1997)

1.2 Objetivos

O principal objetivo deste projeto é pesquisar, propor e implementar contribuições para o desenvolvimento de um sistema de telepresença capaz de realizar a aquisição, transmissão e projeção de vídeos panorâmicos de 360° em ambiente de projeção imersivo, com qualidade suficiente para causar sensação de presença nos usuários.

Após a comprovação da viabilidade técnica de construção e operação, pretende-se estabelecer uma metodologia de solução de problemas científicos e tecnológicos para a implementação deste sistema de telepresença com vídeos panorâmicos, a ser seguida em trabalhos futuros, e que possa fornecer subsídios para o projeto GigaVR - Plataformas para o Desenvolvimento de Aplicações de Realidade Virtual Imersiva e Distribuída sobre Redes de Altíssima Velocidade (RNP, 2004a) e (RNP, 2004b), em desenvolvimento no LSI-EPUSP.

1.3 Motivação

O interesse por adquirir e manipular imagens (fotografias e vídeos) surgiu como um passatempo e foi se tornando um tema de prospecção cada vez mais presente em minhas atividades de pesquisa e desenvolvimento.

No meu projeto de mestrado, o tema de pesquisa desenvolvido foi “Análise Comparativa dos Principais Formatos de Armazenamento de Imagem” (HU, 2000); no trabalho, são analisados os principais formatos de armazenamento de imagens, os algoritmos de compressão e seus respectivos desempenhos.

O interesse em aprofundar técnica e cientificamente essas questões levou-me a buscar integrar-me às atividades de pesquisa do Grupo de Meios Eletrônicos Interativos do LSI-EPUSP. À época, já existia e operava a CAVERNA Digital® do LSI-EPUSP, ambiente de realidade virtual imersivo de múltiplas projeções desenvolvido pelo LSI-EPUSP, sob a coordenação do Prof. Dr. Marcelo Knörich Zuffo. Estava também em andamento no LSI-EPUSP um projeto de pesquisa e desenvolvimento de ferramentas voltadas à Educação Musical, o Projeto DAPEM – “Desenvolvimento e Avaliação de um Portal Educativo-Musical: um Ambiente Colaborativo/Cooperativo”, sob a coordenação da Profa. Dra. Roseli de Deus Lopes, com financiamento do CNPq.

Em uma das reuniões de prospecção para confecção do projeto de pesquisa discutimos o tema “ambientes avançados de aprendizagem”. Uma das situações imaginadas foi a de um estudante de música que pudesse experimentar situações em que tocasse sozinho, mas com a sensação de estar tocando numa orquestra, sem os riscos da reprovação do maestro.

Daí surgiu a idéia de fazer os primeiros experimentos com fotografias panorâmicas, usando como espaço de testes a Sala São Paulo, local de apresentações da Orquestra Sinfônica do Estado de São Paulo.

Na mesma ocasião, a equipe do LSI-EPUSP iniciou a elaboração da proposta do projeto GigaVR (RNP, 2004a) (RNP, 2004b), que compreende a utilização da CAVERNA Digital® da EPUSP para simular presença em ambientes reais, porém distantes e/ou inacessíveis, por meio da captura de vídeo do mundo real, transmissão por redes de alta velocidade e exibição na CAVERNA Digital®.

Estava lançada a semente para o desenvolvimento desta tese, envolvendo temas e atividades que me trazem satisfação pessoal num projeto com grande potencial para contribuições científicas e tecnológicas.

1.4 Relevância

Identificam-se como potenciais áreas de aplicação, para esse tipo de sistema, aquelas em que se possa enviar um VOR (Veículo de Operação Remota) munido de um sistema com um conjunto de câmeras como o aqui proposto ao local de interesse, em lugar do(s) especialista(s), que atuaria(m) a distância, na observação e identificação de soluções para o problema.

Com esse tipo de sistema, problemas ocorridos em situações de alto risco ou em locais inacessíveis ao ser humano poderiam ser tratados com maior agilidade, sem necessidade de sua presença física. Será ainda passível de utilização nas mais diversas áreas de aplicações científicas e comerciais, dentre as quais se pode citar:

- Engenharia: controle de veículos, automóveis, trens, navios, aviões e robôs, remotamente;
- Indústria *offshore*: envio de VOR a grandes profundidades, em águas profundas;
- Ciências espaciais: vôos espaciais, exploração espacial e/ou controle de sondas;
- Medicina: procedimentos cirúrgicos a longa distância;
- Segurança: robôs que desarmam bombas e minas terrestres;
- Educação: visita a museus e espaços de patrimônio histórico a distância;

- Hiperguia: visitas automatizadas a museus e salas culturais;
- Arqueologia: visita a locais remotos ou inacessíveis a seres humanos;
- Entretenimento: *videogame*, novas formas de arte, filmes imersivos;
- Turismo: viagens a locais distantes.

No caso de aplicação como hiperguia, a Universidade Federal da Paraíba (UFPB), parceira no projeto GigaVR, está desenvolvendo o ambiente ICSPACE.

1.5 Atividades

Este trabalho envolveu as seguintes atividades:

- Pesquisa de alternativas de arquiteturas de *hardware* e *software* para a construção de um sistema de captura, tratamento, transmissão e exibição de vídeos panorâmicos em 360°, ao vivo;
- Escolha de uma das alternativas e detalhamento dos aspectos envolvidos nesta escolha;
- Projeto e desenvolvimento de um protótipo de hardware;
- Projeto e programação de um protótipo de software;
- Montagem do protótipo, compreendendo a integração entre o *hardware* e o *software*;
- Testes do protótipo;
- Avaliação dos resultados e identificação de trabalhos futuros.

1.6 Trabalhos correlatos

Este trabalho se insere nos objetivos do projeto GigaVR (Plataformas para o Desenvolvimento de Aplicações de Realidade Virtual Imersiva e Distribuída sobre Redes de Altíssima Velocidade), financiado pela RNP, sob a coordenação do LSI-EPUSP e com a colaboração das seguintes universidades: Universidade Federal do Rio Grande do Norte (UFRN), Universidade Federal da Paraíba (UFPB), Universidade Federal do Rio de Janeiro

(UFRJ), Universidade de Salvador (UNIFACS) e Universidade Federal de São Paulo (UNIFESP) (Figura 2).

Figura 2: Projeto GigaVR

O objetivo central do projeto GigaVR é desenvolver uma plataforma de telepresença imersiva, utilizando redes de alta velocidade para o tráfego de informações. As imagens serão capturadas nos *campi* das universidades envolvidas por meio de um robô, munido de um anel de câmeras, e transmitidas através da Internet 2 para o *campus* da USP, onde serão tratadas e projetadas nas telas da CAVERNA Digital® da EPUSP (ZUFFO et al., 2001).

No que se refere a outros trabalhos correlatos desenvolvidos no LSI-EPUSP, aquele que mais se relaciona a este projeto é o do Dr. Celso Setsuo Kurashima, que pesquisou em sua tese de doutorado, intitulada “Proposta de um Sistema de Teleconferência Imersiva Panorâmica”, a utilização da CAVERNA Digital® da EPUSP como ambiente de teleconferência imersiva (KURASHIMA, 2005). A montagem do primeiro conjunto, composto por doze câmeras, denominado **Anel de Câmeras**, foi responsabilidade sua (Figura 3). Este anel de câmeras foi o ponto de partida para a captura de vídeos do presente projeto.

Figura 3: Anel de câmeras (KURASHIMA, 2005)

Outros trabalhos relevantes na área são:

- O sistema de captura de ambientes utilizando câmeras catadióptricas, de Aliaga e Carlbom (2001), para aplicações em ambientes fechados e pequenos;
- O sistema de vídeo panorâmico imersivo desenvolvido com câmeras catadióptricas apresentado por Neumann (2000) e por Pintaric et al. (2003), no qual a reconstrução da cena é feita por mapeamento de textura sobre um cilindro;
- O sistema de projeção de vídeo panorâmico estereoscópico denominado TWISTER, de Tanaka et al. (2004), um sistema auto-estereoscópico em que o vídeo é projetado em um cilindro que envolve completamente o espectador.

1.7 Estrutura da tese

Esta tese está organizada em sete capítulos.

O capítulo 1 apresenta contexto, objetivos, motivação, relevância, atividades, requisitos do sistema e trabalhos correlatos.

O capítulo 2, os conceitos gerais e o estado da arte dos métodos e equipamentos e das tecnologias utilizadas no desenvolvimento do presente trabalho.

O capítulo 3, a proposta de sistema de telepresença.

O capítulo 4, a implementação e testes com imagens estáticas (fotografias), em projeção panorâmica de 360°, que constituiu o estágio inicial de todo o trabalho.

No capítulo 5 descrevem-se detalhes de implementação e testes com captura de vídeos em projeção panorâmica de 360°, transmissão via rede de alta velocidade e exibição ao vivo em ambiente imersivo.

O capítulo 6 discute os resultados obtidos no projeto.

O capítulo 7 registra as conclusões gerais, as contribuições obtidas e sugestões e direcionamentos para trabalhos futuros.

2 CONCEITOS E TECNOLOGIAS EM IMAGENS PANORÂMICAS

2.1 Introdução

Este capítulo apresenta uma visão geral considerando o histórico e o estado da arte dos conceitos e tecnologias relacionados com sistemas de imagens panorâmicas.

Apresenta-se inicialmente um histórico da aquisição e apresentação de imagens panorâmicas e de imagens estereoscópicas. O desenvolvimento das duas técnicas deu-se no século XIX, apesar de não terem utilização usual, e se confunde com o desenvolvimento da fotografia.

Na seqüência estão os conceitos básicos de imagens digitais.

Após a aquisição de imagens é necessário corrigir distorções de lente, projetar a imagem resultante de forma correta, e alinhar o conjunto de imagens, antes de agrupá-las em uma única imagem panorâmica. Todos estes conceitos são tratados na seqüência.

Registram-se também métodos e algoritmos utilizados para a compressão de imagens e vídeos, em particular o JPEG, o MJPEG e o MPEG.

O capítulo termina com a apresentação de tecnologias envolvidas em ambientes de Realidade Virtual Imersivos e de Telepresença, como comunicação e processamento de alto desempenho.

2.2 Fotos panorâmicas

Imagem panorâmica é a denominação utilizada para fotos ou pinturas que retratam uma paisagem em um ângulo muito maior do que a vista humana consegue alcançar, sem

movimentar o globo ocular. Esta característica já é explorada comercialmente em cinemas com imagem em 360° e em planetários.

Desde os primórdios da fotografia surgiram inúmeros aparelhos e dispositivos que procuravam registrar imagens em perspectiva. Existem registros de imagens panorâmicas desde 1839 (CURTIN, 2001), quando foram obtidas séries de imagens com uma câmera denominada *daguerreotype*, imagens estas que podiam ser emolduradas ou penduradas lado a lado. A partir daí, vários fotógrafos fizeram experiências com fotografias panorâmicas. (Figura 4 e Figura 5).

Figura 4: Panorama de Chicago antes do incêndio (CURTIN, 2001)

Figura 5: Long Beach, Califórnia, Bathing Beauty Parade, 1927 (CURTIN, 2001)

Tiram-se seqüências de fotos, cada uma de um determinado ângulo, variável em função da abertura da lente da máquina fotográfica (Figura 6). As fotos reveladas são copiadas em papel, recortadas e coladas lado a lado para formar as imagens panorâmicas.

Figura 6: Seqüência de fotos com rotação da câmera (CURTIN, 2001)

Em 1904 foi patenteada pela Kodak® uma máquina que girava sobre um eixo e fotografava imagens em 360°, a "Kodak Cirkut®". A câmera utilizava rolos de filme de grande formato (Figura 7).

Figura 7: Máquina panorâmica da Kodak (CURTIN, 2001)

Existe grande quantidade de modelos de câmeras panorâmicas comerciais que, no entanto, não conseguem fotografar 360°, fotografam um ângulo entre 90° e 180° (Figura 8).

Figura 8: Câmeras fotográficas panorâmicas (RIGG, 2001)

Para a aquisição de imagens panorâmicas de 360°, podem-se utilizar máquinas fotográficas convencionais, máquinas digitais ou máquinas panorâmicas especiais, usando técnicas de mosaico similares à técnica utilizada em 1839.

O uso de máquinas fotográficas convencionais permite reduzir o custo de aquisição dos equipamentos, porém exige um processo mais elaborado e cuidadoso, que aumenta o tempo para a aquisição de imagem e eleva o custo, pois os fotogramas necessitam ser processados em laboratório fotográfico. Após a aquisição das imagens é necessário digitalizá-las e utilizar *softwares* para a junção das mesmas e ajustes das cores, o que não seria necessário com o uso de máquinas panorâmicas (RIGG, 2001).

A utilização de máquinas panorâmicas especiais permite a criação da imagem em 360° com uma ou duas fotos, porém o custo desses equipamentos é elevado. Essas máquinas podem ser assim classificadas: máquinas com lente rotativa, máquinas fotográficas rotativas (Figura 9) ou máquinas cuja lente tem um ângulo de visão de até 180° (grande angular ou olho de peixe), as quais, no entanto, apresentam grande distorção de lente (CURTIN, 2001). Estas distorções podem ser corrigidas com a digitalização das imagens e o uso de programas de computadores.

Uma vez adquiridas as varias imagens que compõem um panorama, é possível criá-lo utilizando softwares específicos para isto. Alguns dos softwares comerciais que constroem panoramas são o PanEdit® da IBM®, o Photoshop CS® da Adobe® e o PhotoStitch® da Cânon®.

Figura 9: Máquina panorâmica giratória (CURTIN, 2001)

Outra possibilidade é o uso de máquinas que utilizam espelhos esféricos, parabólicos, ou com curvaturas diversas, que permitem obter-se em uma única foto um ângulo de 360° na horizontal. Tais fotos necessitam ser processadas digitalmente para que se corrijam as distorções introduzidas pelas curvaturas dos espelhos (Figura 10).

Figura 10: Máquina fotográfica com dois espelhos (CURTIN, 2001)

Nesse tipo de construção é utilizado um espelho cônico ou esférico com uma câmera apontada para o seu centro. A configuração é conhecida como *catadioptric omnidirectional video cameras* (câmeras de vídeo catadióptricas omnidirecionais) (Figura 11) (KANG, 1995).

Figura 11: Câmeras de vídeo catadióptricas omnidirecionais (NAYAR, 1997)

Os espelhos geram diferentes tipos de imagem, como as de projeção retilinear, esférica ou cúbica. (RIGG, 2001; WAACK, 2001) (Figura 12 e Figura 13).

Figura 12: Imagem panorâmica com projeção esférica (RIGG, 2001)

Figura 13: Imagem gerada por projeção esférica e corrigida (RIGG, 2001)

A maior vantagem desta alternativa é a simplicidade de sua construção (NAYAR, 1997). As principais desvantagens é a necessidade de se utilizar um *software* para a correção de imagens distorcidas e a baixa resolução destas imagens. Estes *softwares* são específicos para cada tipo de espelho. Por exemplo, o denominado “*Paracamera*”, desenvolvido e patentado por S. Nayar.

Câmeras com definição maior são caras e demandam o desenvolvimento de *software* mais complexo. Além disso, devido às características do espelho, não haverá homogeneidade no tamanho dos pixels (KANG; SZELISKI, 1995).

Como mostra a Figura 14, superior das imagens será formada pelos pixels tomados do círculo externo, enquanto a parte inferior das imagens será formada pelos pixels do círculo central. É possível perceber que o círculo externo contém maior quantidade de pixels que o círculo interno, logo a resolução das imagens superiores será maior que a das imagens inferiores.

Figura 14: Diferentes homogeneidades dos pixels (NAYAR, 1997)

Outra alternativa é utilizar uma filmadora girando a velocidade constante em relação ao seu centro óptico. Neste caso seriam utilizados os quadros gerados, retirando de cada quadro uma tira, aplicando as devidas correções ópticas e costurando as tiras em uma sequência, formando uma imagem panorâmica (PELEG, 2001; SHUM et al., 2005).

Esse tipo de construção é denominado *Manifolds*, e serve para gerar imagens estáticas, porém não é ideal para imagens dinâmicas (vídeos), porque as filmadoras conseguem filmar até 30 quadros por segundo. Supondo que para conseguir abranger os 360° sejam necessários 12 quadros, e para se ter conforto visual precise-se de uma taxa mínima de 15 quadros por segundo, a câmera deveria girar em 360° quinze vezes por segundo, e para 12 quadros a cada volta seria necessária uma câmera com velocidade de 180 quadros por segundo.

Não existe câmera comercial com tal velocidade. Utilizando uma câmera comercial com velocidade de 30 quadros por segundo seriam possíveis 2,5 rotações por segundo, obtendo-se uma taxa de 2,5 quadros por segundo, o que é insatisfatório para o nosso projeto.

2.3 Fotos estereoscópicas

Os princípios das imagens estereoscópicas foram descobertos em 1838, por Charles Wheatstone, um ano antes de ser inventada a fotografia 3D (Figura 15).

Figura 15: *Daguerreotype* estéreo (CURTIN, 2001)

Fotografia 3D é um conjunto de duas fotografias, também chamadas de par estereoscópico, tiradas ao mesmo tempo ou com um pequeno intervalo, focando um objeto ou pessoa, e com uma pequena diferença angular na posição de tomada de cada foto.

O problema, na época, era o alto custo para obter as fotografias, o que influenciava o custo de obtenção de imagens estereoscópicas. As fotos eram tiradas com equipamentos fotográficos normais, deslocados ligeiramente entre uma foto e outra para gerar a diferença angular.

Os visualizadores também eram equipamentos caros (Figura 16).

Figura 16: Visualizadores e câmera estereoscópica (CURTIN, 2001)

A partir de 1850 surgem métodos baratos para obter imagens em 3D e visualizá-las, o que popularizou a técnica. Atualmente, existem no mercado várias opções de câmeras e de dispositivos que permitem obter imagens estereoscópicas (Figura 17).

Figura 17: Câmeras e dispositivos atuais (CURTIN, 2001)

Para a criação de imagens estereoscópicas pode-se utilizar também uma máquina fotográfica convencional ou especial. Também nesse caso, o uso de máquinas convencionais permite redução de custo, porém resulta em um processo mais trabalhoso e sujeito a problemas.

As possíveis construções são apresentadas na Figura 18. O diagrama (A) mostra a utilização de uma única câmera, apenas alterando sua localização. Essa construção foi uma das utilizadas no projeto de imagens estáticas, e está detalhada no capítulo 4. O diagrama (B) mostra o uso de uma câmera com duas lentes, como apresentado nas duas imagens superiores da Figura 17. No diagrama (C) vê-se uma montagem do tipo duas câmeras distintas acopladas, como nas duas imagens inferiores da Figura 17. No diagrama (D) é apresentada mais uma máquina com duas lentes. A diferença em relação ao diagrama (B) é apenas a distância entre as lentes. No diagrama (E) apresenta-se a construção de uma máquina de uma única lente, com espelhos montados de forma a gerar duas imagens distintas no mesmo fotograma. Cada espelho filtra determinada cor, gerando uma imagem do tipo *Anaglyph*.

Figura 18: Possíveis configurações de câmaras estereoscópicas

Anaglyph é um método patenteado por Du Hauron (ACMI, 2006), em 1891, que consiste em filtrar cada uma das duas imagens estereoscópicas com filtros de cores diferentes, geralmente vermelho e azul, e sobrepô-las em um mesmo quadro. Para visualizar a imagem é

necessário utilizar óculos especiais, com lentes nas cores vermelha e azul. A Figura 19 é uma imagem *Anaglyph*.

2.4 Fotos panorâmicas e estereoscópicas

Imagens panorâmicas e estereoscópicas também não são novidades, e existem equipamentos, como o PanDC®, que permitem a obtenção de imagens em 3D com 120° (Figura 19). Dessa forma é possível obter um panorama em 360° com apenas três fotos (quatro, se forem consideradas áreas de sobreposição).

Figura 19: Imagem obtida pela PanDC® (CURTIN, 2001)

Outra proposta interessante é a de Peleg e Ben-Ezra (1999 e 2001), que sugerem a utilização de uma câmera de vídeo girando sobre seu eixo óptico, sistema denominado *Manifolds*, já explicado no item 2.2.

A diferença em relação ao modelo *Manifolds* é que, em lugar de se retirar uma tira (a central), retiram-se duas, uma do lado direito e outra do lado esquerdo do quadro, referentes ao olho direito e ao olho esquerdo, respectivamente. Montam-se então duas imagens, uma com as tiras do lado direito e outra com as tiras do lado esquerdo.

Esse tipo de construção pode ser observado na Figura 20.

Figura 20: Câmera do tipo *Manifolds* capturando tiras estereoscópicas (PELEG; BEN-EZRA, 2001)

Outra possibilidade é a construção de dois conjuntos de câmeras catadriópticas omnidirecionais opostas a 180° , de forma a gerar imagens ou vídeos estereoscópicos panorâmicos (GLUCKMAN; NAYAR; THORESZ, 1998; SHIMAMURA et al., 2000; TAN; HUA; AHUJA, 2002) (Figura 21).

Figura 21: Duas câmeras omnidirecionais para captura de panoramas estereoscópicos (SHIMAMURA et al., 2000)

O problema em tal construção é que ela se presta bem para visão robótica na determinação de distâncias de objetos e formas, porém não serve para gerar sensação de imersão, pois o campo de visão estereoscópico gerado é vertical, enquanto o ser humano tem visão estereoscópica horizontal (ZHU, 2001).

Outras configurações, usando diferentes tipos de lentes e espelhos, que possibilitam a construção de imagens panorâmicas e estereoscópicas, são apresentadas por Svoboda (2000) e por Fiala (2002), mas não serão tratadas neste trabalho.

Não foram encontradas referências, nas publicações pesquisadas, quanto a imagens estereoscópicas e panorâmicas de 360° para uso em sistemas de multiprojeções de realidade virtual (RV), como a CAVERNA Digital® da EPUSP, com captura de imagens ao vivo. Esta é uma das propostas de trabalho futuro.

2.5 Imagens analógicas e digitais

O princípio para a obtenção de uma imagem, tanto analógica, como digital, é o mesmo, sendo necessário alterar apenas o tipo sensor onde a luz é capturada.

A câmera fotográfica é uma caixa lacrada à entrada de luz, com um orifício no qual está acoplada uma lente. Este deve ser o único ponto por onde os raios de luz podem entrar na câmera, e esta entrada é regulada por um obturador (que permite ou veda totalmente a entrada de luz) e por um diafragma que regula a quantidade de luz que incide no sensor. Os raios de luz irão de encontro à superfície oposta à lente, onde está o sensor, que será um rolo de filme fotográfico, no caso de uma câmera analógica, ou um sensor CCD (*Charge Coupled Device*), no caso de uma câmera digital.

2.5.1 Imagens analógicas

Fotografias analógicas são as captadas por câmeras fotográficas que utilizam filmes fotográficos. O filme fotográfico é composto por uma película plástica, geralmente acetato de

celulose, revestida de uma fina camada de gelatina com cristais de prata, que é um material fotossensível. Os cristais de prata podem ter maior ou menor sensibilidade à luz, dependendo de sua composição química. Esta diferença de sensibilidade permite captar as imagens com mais ou menos luz e seus valores são padronizados pela ISO (*International Standards Organization*), sendo antigamente denominado de ASA ou DIN (KODAK, 2004).

Este filme, após as fotografias terem sido obtidas, são processados quimicamente, de forma a fixar as partículas fotossensíveis que formam as imagens (grãos) definidamente na película plástica, processo conhecido como revelação.

Existem dois tipos de filmes:

- O negativo, que forma uma imagem com as cores complementares, e deve ser transferido para papel fotográfico (processo conhecido como ampliação);
- O cromo, que forma a imagem nas cores originais, servindo para a criação de slides.

2.5.2 Imagens digitais

Uma imagem digital pode ser capturada através de câmeras digitais ou através de *scanner* (digitalizador) ligado ao computador. A câmera digital e o *scanner* contêm sensores que captam a luz emitida pelos objetos ou por eles refletida, e a decompõem nos seus componentes fundamentais de cor, denominada RGB: R - *red* (vermelho), G - *green* (verde) e B - *blue* (azul) (HU; RAUNHEITTE, 2000).

A imagem capturada é bidimensional e pode ser dividida em seus elementos fundamentais, os pixels. **Pixel** é o menor elemento de uma imagem digital. Sua representação é a de um ponto, porém, com um tamanho definido, diferentemente do conceito matemático de ponto com tamanho nulo. Este ponto eventualmente pode ter formato quadrado, e, no caso

de imagens coloridas, é formado pelos componentes fundamentais de cor: vermelho, verde e azul (ALBUQUERQUE, 2004) (Figura 22).

Figura 22: Uma imagem e seus pixels

Estes pixels são enviados para um conversor analógico / digital, e quantizados, ou seja, os valores dos componentes fundamentais RGB, que teoricamente poderiam variar de zero ao infinito, são convertidos para poucos valores inteiros, o que possibilita guardar a representação de cada componente em um conjunto de bits (por exemplo, para 256 níveis, em oito bits).

2.6 Correção e projeção de imagens

No processo de aquisição de imagens com câmeras podem ocorrer distorções, dependendo das características das câmeras. É necessário entender o que as ocasiona para se efetuarem as correções necessárias. Além disso, antes de exibir as imagens é necessário realizar operações de transformação e de projeção adequadas.

2.6.1 Características de lentes

As lentes apresentam algumas características físicas e de construção que precisam ser entendidas, para se costurarem corretamente as imagens geradas pelas câmeras.

Todas as lentes apresentam distorções, que podem ser causadas pelos processos de fabricação, por problemas de montagem da lente ou devidas a suas características de projeto. O entendimento dessas características torna possível montar modelos matemáticos de correção de distorções de imagens, geradas pelas lentes.

As principais características e informações relevantes para a correção de distorções são (Figura 23):

- Eixo óptico: É a reta que passa pelo centro da lente e pelo centro perspectivo;
- Plano da imagem: É o plano onde se forma a imagem projetada. No caso de câmera digital, é o plano onde se localiza a matriz de sensores, e, em máquinas analógicas, onde se localiza o filme;
- Centro perspectivo (CP): É o ponto de convergência de todos os raios ópticos;
- Distância focal: É a distância, ao longo do eixo óptico, do ponto de convergência dos raios luminosos até o plano da imagem;
- Ponto principal: Também conhecido como centro da imagem. Em uma câmera perfeitamente calibrada, pode ser definido como a interseção do eixo óptico do conjunto de lentes com o plano da imagem;
- Coeficientes de distorção (K_1 , K_2 , K_3 , K_4 e K_5): Coeficientes de distorção radial e tangencial.

Figura 23: Diagrama esquemático das principais características de uma lente

A esse conjunto de informações e características dá-se o nome de **parâmetros intrínsecos de uma lente**.

Os parâmetros extrínsecos são determinados pela orientação espacial da imagem em relação ao sistema de coordenadas global que está sendo utilizado e podem ser representados pela matriz de translação e pela matriz de rotação.

É necessário determinar os valores desses parâmetros para calibrar as imagens geradas por uma câmera (HEIKKILÄ; SILVÉN, 1997).

Normalmente, tais parâmetros são obtidos no sistema métrico. Assim, uma das primeiras operações é a conversão para a unidade pixel, utilizada no cálculo da correção da imagem em sistemas de processamento de imagens digitais.

2.6.2 Distorções de lentes de câmeras

A distorção da imagem pode ser definida como um desvio do raio de luz em relação à trajetória ideal, depois de este atravessar uma lente, fazendo com que um ponto qualquer do objeto observado esteja em posição diferente da esperada na imagem gerada.

Dado um ponto cuja localização deveria estar na posição (x, y) do plano da imagem, ele estará na posição (x_0, y_0) , devido à distorção da lente. As novas coordenadas serão dadas por:

$$x_0 = x + \delta x(x; y) \quad (1)$$

$$y_0 = y + \delta y(x; y) \quad (2)$$

onde (δx) e (δy) representam o erro posicional devido às distorções geométricas introduzidas pela lente, nos dois eixos (SILVA, 2003).

Pode-se classificar as câmeras em métricas e não-métricas.

As câmeras métricas são utilizadas para propósitos específicos, como a fotogrametria, e utilizam lentes de precisão (FRAGASSI, 2004), que apresentam baixa distorção, mas são caras e de uso em equipamentos específicos.

Além disso, mesmo tais lentes apresentam alguma forma de distorção, pois “a geometria perfeita só pode ser obtida para um único ponto no objeto e para um único ponto na imagem, lentes não podem ser projetadas para ter um imageamento perfeito sob todos os campos de visada e variação de foco” (FRAGASSI, 2004).

Os equipamentos de uso mais comum utilizam lentes denominadas não-métricas, projetadas para apresentarem baixa perda de luminosidade, à custa de distorções na imagem.

Existem duas formas de distorção:

- Distorção Radial: É a distorção causada pelo deslocamento radial de um ponto de imagem, podendo ser de dentro para fora ou de fora para dentro. É causada por uma curvatura radial defeituosa da lente ou por utilização de lente com distância focal pequena (BOUGUET, 2004) (Figura 24);

Figura 24: Vetores de deslocamento de distorção radial (BOUGUET, 2004)

- Distorção Tangencial: É a distorção causada pelo deslocamento na direção tangencial, gerada por lentes cujos eixos ópticos estão descentralizados (Figura 25);

Figura 25: Vetores de deslocamento de distorção tangencial (BOUGUET, 2004)

- Distorção Geométrica: É a combinação das distorções radiais e tangenciais.

2.6.3 Correção de distorções de lentes de câmeras

O primeiro modelo de correção de lente que considerava tanto a distorção radial quanto a distorção tangencial foi proposto por Brown, em trabalho intitulado *Decentering Distortion of Lenses* (Distorção Descentralizada de Lentes) (1996).

Considerando as constantes da lente:

- Distância focal: em pixels, dada por fc_1 (horizontal) e fc_2 (vertical);
- Ponto principal: coordenada do ponto principal dada por cc_1 (horizontal) e cc_2 (vertical);
- Coeficiente de *Skew*: define o ângulo entre os eixos x e y dos pixels nos casos em que estes apresentam uma inclinação (não estão alinhados verticalmente), e é denominado $alpha_c$. Assume-se que o ângulo dos eixos x e y dos pixels será de 90° , sendo que a inclinação em relação a este ângulo será zero. Na prática, este coeficiente não é utilizado no projeto.

De posse destas constantes, obtidas geralmente no manual do fabricante da lente, ou ainda através de testes, pode-se calcular a matriz da lente, denominada KK.

Os valores foram obtidos utilizando o *software* MatLab® e o aplicativo *Camera Calibration Toolbox for Matlab*, desenvolvido por Bouguet (2004); inseriu-se uma série de imagens de um tabuleiro de xadrez e no programa, e este calcula todos os coeficientes, como descrito no item 5.4.1.

$$KK = \begin{bmatrix} fc_1 & alpha_c \bullet fc_1 & cc_1 \\ 0 & fc_2 & cc_2 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

Além destes parâmetros, há os coeficientes de distorção, que definem a distorção radial e tangencial e são dados por K_1 , K_2 , K_3 , K_4 e K_5 .

A distância radial entre a posição dos pixels e o ponto principal pode ser calculada utilizando o teorema de Pitágoras:

Figura 26: Cálculo da distância radial entre o pixel e o ponto principal

$$r = \sqrt{(x - cc_1)^2 + (y - cc_2)^2} \quad (4)$$

Onde (x) e (y) representam a coordenada do pixel com origem no centro da imagem e (cc₁) e (cc₂), coordenadas do ponto principal em relação ao centro da imagem.

Pode-se descrever a distorção radial (δ_r) de uma lente pela equação:

$$\delta_r = K_1 \cdot r^3 + K_2 \cdot r^5 + K_5 \cdot r^7 + \dots + K_n \cdot r^{2n-3} \quad (5)$$

onde (K₁, K₂, K₅,..., K_n) são coeficientes do polinômio que representam a distorção radial.

É possível desconsiderar as constantes K acima da 7ª ordem, pois sua contribuição na série será de um valor pequeno, que não influenciará o resultado final; logo, não se considerará de K₆ em diante (BROWN, 1966; SILVA, 2003; BOUGUET, 2004; FRAGASSI, 2004).

É possível desmembrar a distorção radial em relação a cada um dos eixos cartesianos, acrescidos de sua posição original, passando a representá-la como:

$$\delta x_r = (K_1 \cdot r^3 + K_2 \cdot r^5 + K_5 \cdot r^7) \cdot \frac{x'}{r} \quad (6)$$

$$\delta y_r = (K_1 \cdot r^3 + K_2 \cdot r^5 + K_5 \cdot r^7) \cdot \frac{y'}{r} \quad (7)$$

onde:

$$x' = x - cc_1 \quad (8)$$

e

$$y' = y - cc_2. \quad (9)$$

Então:

$$\delta x_r = (K_1 \bullet r^2 + K_2 \bullet r^4 + K_5 \bullet r^6) \bullet x' \quad (10)$$

$$\delta y_r = (K_1 \bullet r^2 + K_2 \bullet r^4 + K_5 \bullet r^6) \bullet y' \quad (11)$$

As distorções tangenciais podem ser representadas pela distorção de desalinhamento em relação ao centro óptico e pelos efeitos de prisma da lente, e são dadas pelas seguintes expressões:

$$\delta x_t = 2 \bullet K_3 \bullet x \bullet y + K_4 \bullet (r^2 + 2 \bullet x^2) \quad (12)$$

$$\delta y_t = K_3 \bullet (r^2 + 2 \bullet y^2) + 2 \bullet K_4 \bullet x \bullet y \quad (13)$$

A combinação de todas as expressões de distorções pode ser representada pelas expressões:

$$x_d = (K_1 \bullet r^2 + K_2 \bullet r^4 + K_5 \bullet r^6) \bullet x' + 2 \bullet K_3 \bullet x \bullet y + K_4 \bullet (r^2 + 2 \bullet x^2) \quad (14)$$

$$y_d = (1 + K_1 \bullet r^2 + K_2 \bullet r^4 + K_5 \bullet r^6) \bullet y' + K_3 \bullet (r^2 + 2 \bullet y^2) + 2 \bullet K_4 \bullet x \bullet y \quad (15)$$

Note que se consideram nestas equações tanto as componentes radiais como as tangenciais.

O problema está em achar as constantes K_1 , K_2 , K_3 , K_4 e K_5 . Para encontrá-las novamente pode-se utilizar o MatLab® e o aplicativo *Camera Calibration Toolbox for Matlab* (BOUGUET, 2004).

Uma vez calculados estes coeficientes, pode-se descobrir, para cada pixel, sua nova posição, utilizando a equação (HEIKKILÄ; SILVÉN, 1997; AGAPITO et al., 2001; BOUGUET, 2004):

$$\begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} = KK \bullet \begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix} \quad (16)$$

Onde x_p e y_p são as coordenadas do pixel após a correção das distorções radiais e tangenciais.

2.6.4 Projeção cilíndrica

Após a correção de distorções da lente tem-se uma imagem plana, porém, para costurar as imagens na construção de um ambiente de 360°, deve-se corrigir novamente as imagens para produzir uma projeção cilíndrica ou esférica (KANG, 2001) (Figura 27).

Figura 27: Projeção cilíndrica de um conjunto de imagens

No caso de se usar uma projeção cilíndrica, será preciso projetar cada ponto da imagem em um cilindro de raio “Z”, de mesma altura da imagem (AGAPITO, 2001) (Figura 28).

Figura 28: Projeção da imagem plana no cilindro

Para calcular a posição dos pontos projetados no cilindro usa-se trigonometria.

Considerando que o plano da imagem está em um espaço tridimensional, pode-se construir um cilindro encostado ao plano da imagem.

Considerando a origem dos eixos no centro deste cilindro, e assim coincidente com o centro óptico da lente, o raio do cilindro (Z) será a distância focal da lente. Para a imagem de uma câmera o eixo “ Z ” deverá coincidir com o eixo óptico.

É necessária uma conversão das coordenadas dos pontos. Normalmente considera-se a origem dos eixos da imagem em seu canto superior esquerdo. Para efetuar a projeção considera-se o centro da imagem, coordenada $(0, 0)$, no ponto principal do plano da imagem.

Figura 29: Mudança nas coordenadas da imagem

Cada ponto da imagem terá agora uma coordenada no sistema global dada como (X, Y, Z) . Esta posição projetada no cilindro terá a coordenada (v, θ) , sendo que “ v ” é a altura do ponto de intersecção entre o cilindro e a reta que liga a origem ao ponto (X, Y, Z) . O ângulo “ θ ” é aquele formado entre o eixo “ Z ” e esta reta (SZELISKI; SHUM, 1997; SHUM, 2001).

Observando o cilindro pela vista superior, vê-se a imagem apresentada na Figura 30.

Figura 30: Vista superior do cilindro e do plano da imagem

Para calcular o ângulo θ , tem-se:

$$\theta = \text{tg}^{-1}\left(\frac{X}{Z}\right) \quad (17)$$

Para calcular “v” tem-se que calcular antes X' .

Figura 31: Vista superior do cilindro e do plano da imagem com X' e Z'

$$X' = Z \cdot \text{seno}(\theta) \quad (18)$$

Agora a altura “v” será calculada pela relação de triângulos:

$$\frac{Y}{v} = \frac{X}{X'} \Rightarrow v = \frac{Y \cdot X'}{X} \quad (19)$$

Substituindo X' e isolando “v”, tem-se:

$$v = \frac{Y \cdot Z \cdot \text{seno}(\theta)}{X} \quad (20)$$

2.6.5 Conversão de imagem projetada cilindricamente em imagem plana

Este problema é exatamente o contrário do exposto no item 2.6.4. A exibição final da imagem recebida será um panorama de 360°, projetado cilindricamente, e num conjunto formado por telas planas de projeção. No caso de quatro telas perpendiculares é necessário dividir a imagem em quatro imagens e projetar cada uma destas em seu respectivo plano (Figura 32).

Figura 32: Projeção plana do panorama recebido

Agora, dado um ponto (v e θ) é necessário achar o ponto correspondente (x , y) (Figura 33).

Figura 33: Projeção da imagem plana no cilindro

Figura 34: Vista superior do cilindro e do plano da imagem com X' e Z'

O raio do cilindro (Z) utilizado foi a distância focal da lente. Como visto na Figura 34, que é uma vista superior, pode-se determinar X' e Z' , por:

$$\text{sen}(\theta) = \left(\frac{X'}{Z} \right) \Rightarrow X' = Z \cdot \text{sen}(\theta) \quad (21)$$

$$\text{cos}(\theta) = \left(\frac{Z'}{Z} \right) \Rightarrow Z' = Z \cdot \text{cos}(\theta) \quad (22)$$

Agora, “ X ” será calculado pela relação de triângulos:

$$\frac{Z}{Z'} = \frac{X}{X'} \Rightarrow X = \frac{Z \cdot X'}{Z'} \quad (23)$$

A partir daí é possível calcular Y :

$$Y = \frac{v \cdot X}{Z \cdot \text{seno}(\theta)} \quad (24)$$

2.6.6 Alinhamento de imagens e transformação de perspectiva

Mesmo tendo corrigido as distorções da lente e projetado cilindricamente as imagens, não é possível ainda costurá-las, porque apresentam pequenas distorções de posição, ângulo e escala. Essas distorções são causadas pelo fato de não se conseguir posicionar as câmeras de

forma a ficarem perfeitamente alinhadas (distorções de posição e de ângulo), ou por diferenças na regulação do foco das lentes (escala).

É necessário alinhar as imagens e ajustar a escala para que seja possível costurá-las (SHUM, 2001). Nesse alinhamento pode-se usar o método dos oito parâmetros para calcular a matriz de transformação da imagem (KANG, 2001; FORSYTH; PONCE, 2002; SUN, C., 2004).

Dadas duas imagens de uma série, obtidas com câmeras posicionadas no mesmo centro óptico, e que tenham os mesmos parâmetros intrínsecos de câmera, pode-se calcular a relação entre suas posições da seguinte maneira:

$$P' = E \bullet P \quad (25)$$

onde P e P' são pontos correspondentes em cada uma das imagens. A matriz é conhecida como **matriz essencial**, sendo representada pelo símbolo “E”, caso a imagem esteja **calibrada**, ou seja, os parâmetros intrínsecos são conhecidos (KANG, 1995).

Caso a imagem não esteja calibrada, essa matriz é conhecida como **matriz fundamental**, sendo representada pelo símbolo “F”.

Pode-se escrever a equação da seguinte forma:

$$P' = F \bullet P. \quad (26)$$

Escrevendo as equações de maneira matricial, tem-se:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & m_8 \end{bmatrix} \bullet \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (27)$$

onde (x', y', 1) e (x, y, 1) são as coordenadas dos pontos correspondentes em cada uma das imagens.

Pode-se reescrever a equação da seguinte maneira:

$$x' = \frac{m_0 x + m_1 y + m_2}{m_6 x + m_7 y + m_8} \quad (28)$$

e

$$y' = \frac{m_3x + m_4y + m_5}{m_6x + m_7y + m_8} \quad (29)$$

O problema está em determinar os parâmetros m_n . Como são necessários vários pontos para calibrar as imagens, é preciso resolver um sistema de equações. No modelo proposto utilizam-se oito pontos, quatro para cada imagem.

Uma das maneiras de resolver esse sistema é através de algoritmos iterativos (MASE, 1996; SHUM; SZELISKI, 2001).

2.7 Compressão de imagens

A aquisição de vídeos, neste caso a aquisição com múltiplas câmeras, para obter vídeos panorâmicos, gera um volume de dados que precisa ser compactado para permitir transmissão, mesmo considerando a possibilidade de se utilizarem redes de alta velocidade. Aqui são apresentados dois métodos utilizados para a compressão de imagens e vídeos, o JPEG e o MJPEG.

2.7.1 Compressão JPEG

JPEG, de *Joint Photographic Expert Group* (Grupo de Especialistas em Imagens Fotográficas), é a denominação do método de compressão de imagens do grupo de especialistas que desenvolveu o método e posteriormente o normalizou na ISO *International Organization for Standardization* (Organização Internacional de Normalização) como um padrão internacional.

O método utiliza dois modos básicos: o método de compressão com perdas, que utiliza um algoritmo baseado na DCT - *Discrete Cosine Transform* (Transformada Discreta de Coseno), e o método de compressão sem perdas, que se baseia em métodos preditivos. Existem no total 44 variações do método de compressão JPEG, e algumas delas são utilizadas

em aplicações específicas, tais como aplicações médicas e imagens astronômicas, entre outras (HU; RAUNHEITTE, 2000).

O diagrama completo da compressão JPEG é apresentado na Figura 35.

Figura 35: Diagrama do método de compressão JPEG

Após a aquisição das imagens, no formato RGB, esses são convertidos para componentes de luminância (“Y”) e cromância (“U” e “V”). A luminância é uma escala de representação numérica do nível de cinza, e a cromância, duas escalas numéricas, que juntas representam as cores.

As matrizes de conversão entre os componentes “RGB” e os componentes “YUV” são:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0,2990 & 0,5870 & 0,1140 \\ -0,1687 & -0,3313 & 0,5000 \\ 0,5000 & -0,4187 & -0,0813 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (30)$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1,00000 & 0,00000 & 1,40200 \\ 1,00000 & -0,34414 & -0,71414 \\ 1,00000 & 1,77200 & 0,00000 \end{bmatrix} \cdot \begin{bmatrix} Y \\ U \\ V \end{bmatrix} \quad (31)$$

As matrizes de cromância são reduzidas na taxa de 2:1 horizontalmente (cada duas linhas são convertidas em uma), e 1:1 ou 2:1 verticalmente (cada duas colunas são convertidas em uma). Estas conversões são chamadas “*sampling* (amostragem) 2h1v” e “*sampling* 2h2v”. A matriz de luminância não é reduzida, pois o olho humano é mais sensível à luminância (tonalidade de cinza) do que à cromância (tonalidades de cores), o que permite maior taxa de perda de cromância sem que esta seja percebida pelo espectador (LI, 2000).

Nessa etapa as informações do quadro são compactadas em duas vezes (2h1v) ou três vezes (2h2v), pois a matriz de luminância permanece no seu tamanho original, e as duas de cromância são reduzidas para a metade (2h1v) ou para um quarto (2h2v) de seu tamanho original (HU; RAUNHEITTE, 2000).

O algoritmo de compressão JPEG divide a matriz de luminância e as duas matrizes de cromância, que são as três matrizes que descrevem uma imagem colorida, em inúmeras matrizes, cada uma com tamanho de 8 x 8 elementos (Figura 35). Com isto, haverá várias matrizes compostas de 64 elementos, conhecidas como *sample values* (valores de amostragem).

Sobre tais matrizes aplica-se o algoritmo DCT - *Discrete Cosine Transform* (Transformada Discreta de Coseno), gerando outras matrizes, denominadas Coeficientes de DCT, onde a maioria dos elementos tem valor próximo a zero. Esse processo transforma a informação do domínio do tempo para o domínio da frequência (AHMED et al., 1974).

A **Transformada Discreta de Coseno** converte uma matriz de valores altamente correlacionados em uma matriz de valores menos correlacionados. Em outras palavras, este algoritmo converte uma matriz numérica em outra, sendo que boa parte dos elementos da nova matriz, especificamente os do canto inferior direito, tem valores próximos a zero, que serão desprezados.

O valor médio dessa matriz representa a cor fundamental dos 64 pixels, é chamado de componente **DC - Direct Current (Corrente Contínua)** e está localizado no canto superior esquerdo da matriz de coeficientes de DCT. Os outros coeficientes são denominados coeficientes **AC - Alternating Current (Corrente Alternada)** e representam os valores de pequenas variações de tonalidade e coloração do bloco.

A Transformada Discreta de Coseno para matrizes de duas dimensões e tamanho de 8 x 8 é dada pela seguinte equação (AHMED et al., 1974):

$$F(u, v) = \left(\frac{1}{4}\right) C(u) C(v) \sum_{u=0}^7 \sum_{v=0}^7 f(i, j) \cos\left((2i+1)u \frac{\pi}{16}\right) \bullet \cos\left((2j+1)v \frac{\pi}{16}\right) \quad (32)$$

Onde:

$F(u, v)$ = coeficiente no domínio da transformada;

u = eixo horizontal no domínio da transformada;

v = eixo vertical no domínio da transformada;

$C(x)$ = $1 / \sqrt{2}$ para $x = 0$;

$C(x)$ = 1 para $x \neq 0$;

x = u ou v ;

$f(i, j)$ = valor do pixel no domínio do espaço;

i = eixo horizontal no domínio do espaço;

j = eixo vertical no domínio do espaço.

Somente a aplicação da DCT resulta em nenhuma compactação, pois a matriz resultante terá o mesmo tamanho que a original. Entretanto, a maioria desses coeficientes tem valores próximos de zero, Assim, a compactação virá a partir da aplicação do processo de quantização. Neste processo de compressão os elementos menos importantes da matriz de coeficientes DCT serão descartados.

A matriz de coeficientes DCT é dividida por outra matriz (chamada de **tabela de quantização**), que transforma todos os elementos em uma nova matriz. O processo irá acentuar as características da matriz de coeficientes de DCT, forçando todos os elementos a se aproximarem do valor zero.

Todos os valores da nova matriz serão arredondados, o que implica a introdução de erro na matriz final, o qual poderá ser notado quando se aplicar o processo inverso na descompactação e geração da imagem. Na montagem do arquivo compactado serão armazenados somente os valores não nulos (Figura 36).

Matriz de Coeficientes DCT							
150	80	40	14	4	2	1	0
2	75	36	10	6	1	0	0
52	38	26	8	7	4	0	0
12	8	6	4	2	1	0	0
4	3	2	0	0	0	0	0
2	2	1	1	0	0	0	0
1	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Matriz de Quantização							
1	1	2	4	8	16	32	64
1	1	2	4	8	16	32	64
2	2	2	4	8	16	32	64
4	4	4	4	8	16	32	64
8	8	8	8	8	16	32	64
16	16	16	16	16	16	32	64
32	32	32	32	32	32	32	64
64	64	64	64	64	64	64	64

Matriz Resultante							
150	80	20	4	1	0	0	0
92	75	18	3	1	0	0	0
26	19	13	2	1	0	0	0
3	2	2	1	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Figura 36: Exemplo de matrizes (a) de DCT, (b) de quantização, e o (c) resultado da divisão entre as duas (TANENBAUM, 1997)

É assim possível obter alta taxa de compressão através da transformação dos Coeficientes da DCT em valores com precisão não superior à necessária para gerar a imagem na qualidade desejada. Nessa etapa ocorre perda de informação, porém, como a vista humana é mais sensível a baixas frequências do que a altas, e as perdas são referentes às altas frequências, logo, dependendo da imagem, podem ser pouco notadas.

Não existe definição na norma JPEG para a tabela de quantização. Cada aplicação pode definir sua própria tabela. A variação na taxa de compressão e conseqüentemente na qualidade da imagem compactada se dá alterando a tabela. Quanto maiores os coeficientes da tabela de quantização, maior a taxa de compressão e, conseqüentemente, pior a qualidade da imagem gerada (GIBSON, 1998).

A matriz de valores quantizados é lida em ziguezague da esquerda para a direita e de cima para baixo, fazendo com que os primeiros elementos lidos sejam valores não zerados. Ignora-se a leitura a partir do momento em que se encontram elementos com valores zerados (RODRIGUEZ, 1996) (Figura 37).

Figura 37: Leitura em ziguezague

No exemplo da Figura 36, a leitura gerará um vetor com 64 elementos, sendo que 44 deles terão valor zero. Após uma leitura em ziguezague da matriz, basta substituir os últimos 44 zeros por um identificador que informe esta quantidade.

Após a leitura em ziguezague, os coeficientes serão rearranjados para possibilitar nova compressão, que pode utilizar os algoritmos de Huffman ou a codificação aritmética. Os algoritmos são do tipo *lossless* (sem perdas), ou seja, não introduzem novos erros nas informações; assim, a nova compressão não afetará a qualidade da imagem. Este algoritmo irá gerar uma tabela de conversão. (LI; ZHONG, 2000).

Finalmente, adicionam-se os *headers* (cabeçalhos) e *trailers* (rodapés) de controle no arquivo, com todos os parâmetros e tabelas a serem utilizados na descompressão. Dentre as tabelas que serão incluídas no arquivo estão as de quantização e a de Huffman.

2.7.2 Compressão MJPEG

No caso de vídeos, a compressão é ainda mais importante. Considerando-se que um pixel utiliza três bytes, o tamanho do quadro normalmente é de 640 pixels por 480 pixels, e a velocidade de um vídeo normalmente é de 30 quadros por segundo, o fluxo de dados é da ordem de 27 milhões de bytes por segundo. Com essa quantidade de informações completa-se um CD em trinta segundos, aproximadamente.

As primeiras idéias de compressão de vídeo visavam comprimir cada quadro com o algoritmo JPEG e armazenar a seqüência de quadros, método conhecido como MJPEG, com o qual se consegue eliminar a redundância espacial, porém não a temporal, que se elimina utilizando o método de compressão MPEG.

As principais vantagens do método de compressão MJPEG são:

- Compressão e descompressão praticamente simétricas, ou seja, o tempo de compressão de um quadro é praticamente o mesmo que o de descompressão;
- Consumo de CPU menos intenso que o do MPEG;
- Possibilidade de uso de *hardware* mais barato;
- Menor complexidade da implementação em *software*;
- Aceita qualquer tamanho de quadro.

Por outro lado, não resolve o problema de redundância espacial, gerando um arquivo ou utilizando uma largura de banda grande, se comparada ao MPEG (NG; CHAN; SHUM, 2005).

2.7.3 Compressão MPEG

O algoritmo de compressão de vídeo MPEG procura eliminar a redundância espacial que não é eliminada pelo algoritmo MJPEG. Utiliza o algoritmo JPEG para eliminar a

redundância temporal de métodos especificados na norma H.261 (HU; RAUNHEITTE, 2004).

A norma H.261 padroniza dois diferentes tipos de quadros, os quadros tipo “I” (ou Intraquadros) comprimidos com o algoritmo JPEG, e os quadros do tipo P (ou Interquadros), que apontam apenas as diferenças em relação ao quadro imediatamente anterior (Figura 38) (Rodriguez et al, 1999).

Figura 38: Quadros no padrão H.261 (Rodriguez et al., 1999)

Porém, a compressão H.261 não permite passar a imagem de trás para frente de forma contínua, pois assim apenas os quadros do tipo “I” conteriam informação útil. Os quadros do tipo “P”, por serem preditivos em relação ao quadro imediatamente anterior, se lidos de trás para frente não acrescentariam nenhuma informação adicional que permitisse compor uma imagem. Um retrocesso de vídeo gravado em H.261 mostrará o vídeo de forma não contínua, sem a sensação de movimento.

A norma MPEG solucionou o problema acrescentando o quadro do tipo “B”, também chamado de bidirecional, que em termos de formato e codificação é semelhante ao quadro do tipo “P”, porém preditivo tanto em relação ao quadro anterior como ao quadro posterior (Figura 39).

Figura 39: Tipos de quadros do padrão MPEG (Rodriguez et al., 1999)

2.8 Ambientes avançados

É importante apresentar também os conceitos gerais e as tecnologias envolvidas em ambientes de Realidade Virtual (RV) e de Telepresença, como comunicação e processamento de alto desempenho.

2.8.1 Ambientes de realidade virtual (RV)

A expressão **realidade virtual (RV)** se refere a um conjunto de métodos e tecnologias relacionadas com a imersão do usuário em ambientes com aparência realística sendo o comportamento e a interação do usuário com o ambiente simulados pelo computador (Foley, 1990).

Dentre as principais aplicações para a realidade virtual (RV) pode-se citar o uso em entretenimento — jogos, cinema interativo 360° —, as aplicações em engenharia (simulação sísmológica de prospecção de petróleo, maquetes de veículos na indústria automobilística, aeronáutica), em medicina (simuladores cirúrgicos, operação a distância) e em arquitetura (modelos de edifícios, planejamento urbano).

Até o final da década de 90 incrementaram-se muito os dispositivos denominados SID - *Spatially Immersive Displays* (Dispositivos de Exibição de Imersão Espacial). Um dos dispositivos mais avançados de realidade virtual (RV) da atualidade é a denominada *cave* (caverna) (CRUZ-NEIRA et al., 1993) (Figura 40).

Figura 40: Cave (CRUZ-NEIRA, 2006)

O LSI-EPUSP projetou, construiu e mantém o primeiro sistema do tipo cave da América Latina, a CAVERNA Digital®. Trata-se de uma estrutura em formato de cubo, com as paredes feitas de material plástico semitransparente, com três metros de lado. O sistema tem cinco projetores de alta intensidade luminosa e alta resolução, que projetam imagens estereoscópicas nas quatro laterais e no chão (Figura 41 e Figura 42). As imagens projetadas nas telas são geradas por um supercomputador gráfico, do tipo aglomerado, desenvolvido no LSI-EPUSP (ZUFFO, 2001).

Figura 41: Maquete virtual da CAVERNA Digital® da EPUSP

Figura 42: CAVERNA Digital® da EPUSP

A CAVERNA Digital®, por ser um ambiente de pesquisa e desenvolvimento, está em constante mudança em sua arquitetura, sendo que em 2006 foi desenvolvido um novo aglomerado, e foi incorporado um novo sistema de sonorização de múltiplos canais.

Podem-se utilizar esses ambientes para projetar imagens previamente preparadas na forma de vídeos (como cinemas 3D), mas normalmente são usados para projetar imagens sintetizadas pelo computador, que interage com o usuário. Em algumas aplicações é interessante combinar imagens reais, capturadas na forma de vídeos, e combiná-las com imagens sintetizadas pelo computador, permitindo aos usuários maior imersão nos problemas apresentados e uma nova interpretação da realidade, isto é, proporcionando-lhes uma **Realidade Aumentada (RA)** (*augmented reality*) (HOU, 2003).

Existem outros modelos de sistema de multiprojeção imersivos, como a *Blue-C* © (STAADT, 2000; GROSS, 2003), que tem apenas três lados, mas por meio de um engenhoso sistema de paredes ftopolarizadas, que podem se tornar ora opacas ora transparentes, permite não apenas projetar imagens nas paredes da caverna, mas também captar imagens de pessoas

ou objetos que estão dentro dela, misturando-as e possibilitando o uso de aplicações de realidade aumentada (RA).

Outro modelo é a caverna em V, onde as imagens são exibidas em duas paredes e o usuário fica em um ponto central (JACOBSON, 2002). Este tipo de caverna é mais simples e barato, mas a sensação de imersão é menor.

A construção de um sistema de multiprojeção imersivo demanda uma série de conhecimentos interdisciplinares, como a construção e o controle de aglomerado, desenvolvimento de *softwares* de rede, controle de múltiplos equipamentos, renderização das imagens ao vivo, *softwares* de vídeo e áudio, projeção sincronizada, arquitetura e design, engenharia elétrica e civil.

Sistemas de multiprojeção imersivo são muito utilizados por cientistas, na indústria, para aplicações que impliquem risco ao usuário, impossibilidade de realização real ou custos extremamente elevados. Por exemplo, na indústria automobilística é muito mais barato construir um modelo de carro virtual, para pesquisa de aceitação com os consumidores, do que construir um modelo em tamanho real.

2.8.2 Comunicação entre microcomputador e periféricos

É importante investigar as possibilidades de comunicação entre os equipamentos de aquisição de imagens, as câmeras digitais e os computadores.

Até pouco tempo atrás a conexão de periféricos ao computador só era possível por meio de portas seriais e paralelas, ou por meio de placas especiais conectadas ao barramento do computador. As portas seriais e paralelas apresentavam sérias limitações quanto à velocidade e à quantidade de periféricos passíveis de serem conectados ao mesmo tempo. Atualmente, as portas seriais e paralelas estão sendo substituídas por outros padrões de conexão:

- USB: Foi a primeira interface com periféricos para PC com a característica *Plug and Play* (conecte e utilize - detecção automática do periférico conectado) e permite que até 127 periféricos sejam conectados ao mesmo tempo, em uma única porta. Sua velocidade é de 12 Mbps, cerca de 100 vezes mais que uma porta serial. Esta interface foi padronizada pela *USB Implementers Forum* (USB-IF, 2006);
- USB 2.0: Desenvolvido pela Intel®, tem velocidade de 480 Mbps e permite a ligação de até 127 periféricos ao mesmo tempo. Também padronizada pela *USB Implementers Forum* (USB-IF, 2006);
- IEEE-1394: Criada pela *Apple®* e padronizada pela IEEE com o número 1.394, também conhecida como *firewire*, é uma conexão serial de alta velocidade, que utiliza topologia flexível, permitindo construir pontes, nós, interconexões, conexões virtuais e portais (MENDONÇA, 2004). Sua velocidade é de 400 Mbps, e possibilita ligar até 63 periféricos ao mesmo tempo. Novas revisões da norma permitirão que o *firewire* trabalhe com velocidades de 800 Mbps a 3200 Mbps. Uma de suas vantagens é permitir sincronizar todos os equipamentos que estejam conectados à interface. Essa característica e a sua velocidade fazem com que seja muito utilizado em equipamentos de áudio e de vídeo, além dos computadores.

2.8.3 Comunicação de alta velocidade - Gigabit Ethernet

Um dos primeiros padrões de rede, o *Ethernet*, possibilitava a transmissão de dados a uma velocidade de até 10Mbps e foi denominado 10BaseT.

Seu sucessor foi o 100BaseT, também chamado *Fast Ethernet*, com velocidade de até 100Mbps. A evolução foi simples, pois esse padrão mantinha compatibilidade com o padrão anterior, permitindo que as empresas trocassem alguns poucos equipamentos e cabos inicialmente e, na medida em que substituíssem os microcomputadores, migrassem para o novo padrão.

Esses padrões estão definidos na norma IEEE 802.3.

Com a proliferação de usuários ligados às redes locais e o aumento de aplicativos como vídeos, áudio, imagens e grandes bancos de dados, que usam maior banda da rede local, houve necessidade de se desenvolver nova tecnologia de rede, pois a anterior não conseguia suprir as necessidades dos usuários.

Em 1996 iniciou-se o desenvolvimento de um novo padrão de rede, o IEEE 802.3z, conhecido como Gigabit Ethernet, cujos requisitos foram:

- Apresentar velocidade de até 1.000Mbps;
- Manter compatibilidade com o padrão anterior;
- Usar o formato do quadro Ethernet 802.3;
- Usar o método de acesso CSMA/CD.

As seguintes características foram definidas para o meio físico de transmissão:

- Uso de fios de cobre para distâncias curtas (até 25m);
- Enlace de fibra óptica em multimodo para distância até 550m;
- Enlace de fibra óptica em monomodo para distâncias até 3.000m.

Observa-se migração incremental para esta nova tecnologia de rede (FARIAS, 1998; LOPES, 2005).

2.8.4 Comunicação de alta velocidade - Internet 2

A Internet 2 foi iniciativa do governo norte-americano para interligar suas universidades com uma rede de alta velocidade. O projeto se iniciou com a interligação de 180 universidades americanas e depois se estendeu para empresas particulares e para universidades de outros países, através de convênios entre o governo norte-americano e os governos dos países interessados.

No Brasil, o Ministério de Ciências e Tecnologia (MCT) e a Rede Nacional de Pesquisas (RNP) firmaram um acordo de cooperação em tecnologia e educação com o

governo norte-americano, em outubro de 1997, e criaram as REMAVs (Redes Metropolitanas de Alta Velocidade), interligando universidades e órgãos governamentais em uma rede de alta velocidade.

A primeira REMAV foi implantada em abril de 1999, na cidade do Rio de Janeiro, interligando a Universidade Federal do Rio de Janeiro, a Fundação Osvaldo Cruz, a PUC do Rio de Janeiro e o Centro Brasileiro de Pesquisas Físicas, a uma velocidade 75 vezes maior que a das redes convencionais da época.

Em São Paulo, a REMAV foi implantada em dezembro de 1999, coordenada pelo LARC (Laboratório de Arquitetura e Redes de Computadores da EPUSP) (SEI, 2005).

A ligação das REMAVs à Internet 2 ocorreu em agosto de 2001; atualmente, a rede, administrada pela RNP (Rede Nacional de Pesquisa), é denominada RNP2 e tem uma velocidade que varia entre 4Mbps e 622Mbps (RNP, 2005) (Figura 43).

Figura 43: *Backbone* (barramento) da rede brasileira RNP2 (RNP, 2005)

2.8.5 Processamento de alto desempenho – Aglomerado de Computadores

Segundo a classificação de Flynn, há quatro tipos de computadores (STALLINGS, 2002):

- **SISD:** *Single Instruction Single Data* (Instrução Única e Dado Único): Arquitetura de von Neumann. É o computador doméstico, cuja CPU recebe um dado e executa uma operação;
- **SIMD:** *Single Instruction Multiple Data* (Única Instrução e Múltiplos Dados): Supercomputadores clássicos, como os computadores vetoriais. Apresentam N processadores, que processam uma única instrução sobre N dados. Excelente para trabalhar com matrizes ou vetores;
- **MISD:** *Multiple Instruction Single Data* (Múltiplas Instruções e Dado Único): O mesmo fluxo de dados percorre uma matriz linear de processadores (*pipeline*). Esta estrutura nunca foi implementada;
- **MIMD** *Multiple Instruction Multiple Data* (Múltiplas Instruções e Múltiplos Dados): Arquitetura formada por processadores que executam instruções independentemente. São microcomputadores completos ligados por rede de alta velocidade, que executam o processamento como se fossem um único computador e são conhecidos como **aglomerado de computadores**.

Das quatro configurações, a única viável para este projeto foi a quarta, visto que a primeira (SISD) não atenderia a demanda, a segunda (SIMD) é cara e a terceira (MISD) não existe.

Um aglomerado de computadores pode ser definido como um conjunto de computadores interligados entre si, que ao trabalharem juntos criam a impressão de que se trata de um único computador (STALLINGS, 2002).

As principais vantagens de um aglomerado de computadores são:

- Escalabilidade absoluta: possibilidade de criação de aglomerado de computadores com milhares de computadores;
- Escalabilidade incremental: possibilidade de se adicionarem computadores novos ao aglomerado de computadores;
- Alta disponibilidade: a perda de um computador interfere pouco no desempenho do aglomerado de computadores;
- Baixo custo por MIPS (milhões de instruções por segundo): seu custo é uma fração do custo de um computador vetorial.

2.9 Conclusão

Neste capítulo foram apresentados o histórico e os principais métodos e técnicas relacionados ao projeto.

Todas as técnicas são conhecidas, e algumas foram descobertas há quase 200 anos, como as de fotos panorâmicas e estereoscópicas. Outras são descobertas recentes, como os ambientes de Realidade Virtual Imersivos e de Telepresença, e os sistemas de comunicação e de processamento de alto desempenho.

Procurou-se apresentar em detalhes os algoritmos de correção de distorções de lentes, projeções de imagens e alinhamento, que são a base do projeto e ponto central para a proposta desta tese de doutorado.

3 PROPOSTA DE UM SISTEMA DE TELEPRESENÇA EM AMBIENTES IMERSIVOS DE VÍDEOS PANORÂMICOS

3.1 Introdução

Este capítulo contém a proposta de um sistema de telepresença por meio da aquisição, transmissão e projeção em ambientes imersivos de vídeos panorâmicos em 360°.

Dada a complexidade técnica do projeto, a arquitetura aqui apresentada é o resultado de um trabalho em equipe desenvolvido pelos pesquisadores envolvidos no projeto GigaVR do LSI-EPUSP.

Este trabalho de doutoramento se concentra na concepção, no detalhamento das etapas de processamento do sistema, na implementação e em testes, bem como na interação com a equipe do GigaVR para acompanhamento da implementação e adequação da infra-estrutura para aquisição e exibição dos vídeos panorâmicos. Um dos seus principais objetivos é verificar a viabilidade técnica de se criarem ambientes de telepresença baseados em vídeos panorâmicos 360°, para transmissão por rede de alta velocidade, e de projetar as imagens na CAVERNA Digital® da EPUSP.

3.2 Requisitos e restrições

Retomando-se a definição de telepresença como uma forma de telecomunicação que possibilita a uma pessoa ou mais pessoas interagir à distância com outras pessoas ou objetos reais, com a sensação de estar presente no local (Walker; Sheppard, 1997), fica evidente que

todo sistema de telepresença deve ocorrer ao vivo, ou com atraso máximo adequado a sua aplicação específica em cada caso. Atrasos maiores do que o máximo tolerável para uma dada aplicação podem comprometer a interação entre pessoas ou com objetos, perdendo-se a característica de telepresença (Figura 44).

Figura 44: Diagrama esquemático de um sistema de telepresença

Os principais requisitos que devem ser atendidos pelo sistema de telepresença proposto neste trabalho são:

- Adquirir os quadros de forma a permitir a transmissão e montagem de vídeos panorâmicos em 360°, ao vivo ou respeitando um atraso máximo definido para cada aplicação específica;
- Imagens coloridas representadas por 24 bits;
- Velocidade mínima de exibição de 15 quadros por segundo. Abaixo desta velocidade o sistema se torna desconfortável para o usuário, pois o vídeo não fornece sensação de continuidade das imagens;
- As imagens adquiridas devem sofrer correções, transformações, alinhamentos, costuras e projeções adequados para exibição no ambiente imersivo.

3.3 Previsão de recursos adicionais

Como requisitos para trabalhos futuros têm-se:

- Captação, transmissão e apresentação de sinais sonoros, estereofônicos e com múltiplos canais de som;
- Captação, transmissão e exibição de imagens estereoscópicas;
- Transmissão de sinais de controle para dispositivos hápticos (dispositivos que fornecem ao usuário sensações táteis, como textura de objetos, peso, densidade e temperatura).

3.4 Alternativas estudadas

Foram estudadas diferentes alternativas de montagem de câmeras para a construção do sistema de captura de imagens.

A primeira foi a de se construir um sistema de captura com câmera de vídeo catadióptrica omnidirecional (*catadioptric omnidirectional video cameras*). Os problemas encontrados nesta arquitetura são a falta de homogeneidade no tamanho dos pixels e a baixa resolução proporcionada por uma única câmera. Como mostra a Figura 45, as imagens que serão exibidas no lado superior serão formadas pelos pixels tomados do círculo externo, e as imagens do lado inferior, pelos pixels do círculo central.

Figura 45: Diferentes homogeneidades dos pixels (NAYAR, 1997)

É possível perceber na figura que o círculo externo contém uma quantidade maior de pixels que o círculo interno, logo a resolução das imagens no lado superior será maior que no lado inferior. Esta diferença de resolução pode ser diminuída, porém não eliminada, criando-se pixels intermediários, com o uso de algoritmos de interpolação.

A segunda alternativa, esta adotada, foi a criação de um anel de câmeras com ângulo de visão de 360°.

A principal vantagem dessa configuração, em relação à montagem com câmera de vídeo catadióptrica omnidirecional, é a resolução maior, que será a soma das resoluções das câmeras utilizadas no dispositivo, descontada a área de sobreposição de imagens. Outra vantagem é a homogeneidade dos pixels na imagem.

O principal compromisso de utilizar o anel de câmeras é a necessidade de garantir o sincronismo de câmeras e de realizar um processamento mais complexo nas imagens a fim de fazer as correções de distorções, projeções, transformações, alinhamento e costuras.

Outra limitação é que nessa configuração surgem zonas mortas, em que nenhuma das câmeras consegue obter imagens, e zonas onde ocorre sobreposição de imagens. O ângulo das zonas mortas varia em função da abertura focal da lente e da distância dos objetos em relação

à câmera. O problema pode ser minimizado diminuindo-se o raio do círculo de montagem das câmeras (Figura 46).

Uma alternativa seria interpolar as imagens da zona morta entre duas câmeras, gerando uma imagem sintética. Não se trata aqui desse problema, o qual poderia ser objeto de um trabalho futuro.

Figura 46: Ângulo de tomada e zonas mortas

Outra característica é a sobreposição de imagens das câmeras, a qual aumenta em função da distância entre os objetos focalizados e a câmera, como se pode ver na Figura 47.

Figura 47: Ângulo de tomada e zonas de sobreposição

Apesar das limitações, foi esta a configuração adotada, porque, para exibição em ambiente imersivo com múltiplas telas de projeção, como a CAVERNA Digital®, é necessária resolução maior que a fornecida por uma única câmera.

3.5 Arquitetura de *hardware*

Para este projeto, a proposta é a montagem de um aglomerado de computadores, formado por três ou quatro microcomputadores, dependendo do número de câmeras utilizado no anel de câmeras, cada um deles biprocessado e todos interligados entre si por uma rede de alta velocidade.

A arquitetura geral de *hardware* é apresentada na Figura 48.

Figura 48: Diagrama esquemático dos equipamentos

O local 1, em que se realiza a aquisição das imagens, deve dispor da seguinte infraestrutura:

- Um anel de câmeras, com oito ou doze câmeras, ligadas de quatro em quatro;

- *Hubs firewire* (IEEE-1394) ligados às câmeras e aos microcomputadores (um para cada quatro câmeras);
- Rede Gigabit, interligando os microcomputadores;
- Microcomputadores capturando, processando e transmitindo as imagens (três computadores no caso de oito câmeras ou quatro, para doze câmeras).

O local 2, em que se exibem as imagens em ambiente imersivo, deve dispor da seguinte infra-estrutura:

- Um aglomerado de computadores recebendo as imagens;
- Um sistema de exibição imersivo de múltiplas projeções (no caso, a CAVERNA Digital® da EPUSP).

Os microcomputadores ligados às câmeras serão denominados **Microcomputadores Servidores**, e efetuarão a captura e o primeiro processamento das imagens. Deverão ser ligados a um equipamento, denominado **Microcomputador Cliente / Servidor**, que consolidará as imagens geradas pelos Microcomputadores Servidores em uma única imagem panorâmica, e a transmitirá para o aglomerado de computadores da CAVERNA Digital®.

3.6 Arquitetura de *software*

O *software*, principal responsabilidade deste projeto, pode ser dividido em um **Módulo de Aquisição**, que será executado nos Microcomputadores Servidores, um **Módulo de Composição**, executado no Microcomputador Cliente / Servidor, e um módulo que opera no aglomerado de computadores da CAVERNA Digital® da EPUSP, que será chamado de **Módulo de Exibição**.

A Figura 49 apresenta uma visão geral dos módulos de *software*.

Figura 49: Diagrama esquemático dos módulos de *software*

3.6.1 Módulo de Aquisição

O Módulo de Aquisição tem este nome por ser aquele que gera as imagens e as transmite para o Módulo de Composição. Nele as imagens são adquiridas de até quatro câmeras, corrigidas as distorções de lentes, projetadas cilíndricamente, alinhadas, rotacionadas em 90°, costuradas, gerando uma imagem panorâmica de 180° (para oito câmeras) ou de 120° (para doze câmeras), comprimidas e transmitidas para o Módulo de Composição.

Com a atual tecnologia de computadores, a quantidade de operações matemáticas envolvidas no processo de geração da imagem panorâmica demanda um esforço computacional elevado, viabilizando a elaboração de algoritmos que permita a simplificação do processo. Considerando 640 x 480 pixels (formato VGA) à velocidade de 15 quadros por

segundo, o fluxo de dados é da ordem de 13,8 milhões de bytes por segundo. Como se trabalha com quatro câmeras por microcomputador, o fluxo total é de 55,3 milhões de bytes por segundo.

Para cada conjunto de três bytes (um pixel) é necessário calcular sua nova posição cinco vezes (para a correção da lente, a projeção cilíndrica, o alinhamento, a rotação em 90° e a costura das imagens), de acordo com as equações apresentadas no Capítulo 2.

Este trabalho propõe e implementa um algoritmo, o qual, a partir das imagens capturadas pelas câmeras, corrige distorções da lente, projeção cilíndrica, alinhamento de imagens, rotação em 90° e costura de imagens em um único passo, possibilitando que os equipamentos atuais gerem um vídeo panorâmico em 360°, ao vivo.

Para viabilizar este algoritmo, o Módulo de Aquisição foi alterado, dividindo-se os oito processos em dois procedimentos, o **Procedimento de Calibração de Imagens**, executado uma única vez no início do processamento e o **Procedimento de Montagem do Panorama**, que roda continuamente, gerando o fluxo de imagens transmitidas.

3.6.1.1 Procedimento de Calibração de Imagens

Cada cálculo de posicionamento de um pixel exige inúmeros cálculos, que demandariam centenas de milhões de operações com ponto flutuante a cada segundo. Não seria possível efetuá-los todos com os equipamentos disponíveis atualmente.

Para viabilizar o projeto desenvolveu-se a estratégia de criar para cada câmera uma matriz, que foi chamada de **Mapa de Posicionamento**, com o mesmo tamanho da imagem gerada pela câmera, que será chamada de **Matriz de Pixels da Câmera**. Cada elemento do Mapa de Posicionamento é um ponteiro, que aponta uma posição da Matriz de Pixels da Câmera.

Inicialmente, é gerado o Mapa de Posicionamento com os ponteiros voltados para as posições originais da Matriz de Pixel de Câmera, sem nenhuma alteração de localização.

Depois efetuam-se os cálculos de correção da distorção da lente, aplicados então sobre o Mapa de Posicionamento. A seguir, sobre o Mapa já corrigido, efetua-se a projeção cilíndrica e, por fim, o alinhamento das imagens (Figura 50 e Figura 51).

Figura 50: Sequência de correções no Mapa de Posicionamento

Figura 51: Sequência completa de montagem do Mapa de Posicionamento

Tal procedimento deverá ser executado antes do Procedimento de Montagem do Panorama e todas as vezes em que se constatar que a costura das imagens não está correta.

Inicialmente, criam-se matrizes de ponteiros, uma para cada câmera, cada uma com o mesmo tamanho da Matriz de Pixels da Câmera.

Em seguida, a matriz é preenchida com os endereços correspondentes da Matriz de Pixels da Câmera. Por exemplo, a posição (4,1) do Mapa de Posicionamento aponta para o endereço (4,1) da Matriz de Pixels da Câmera (Figura 51 – a).

Executa-se depois a rotina de correção de distorção de lente, deslocando os ponteiros no Mapa de Posicionamento. No exemplo, o ponteiro do pixel (4,1) foi deslocado para a posição (4,2) no Mapa de Posicionamento (Figura 51 – b). As posições localizadas nas bordas, que não apontarem nenhum endereço, serão preenchidas com cor preta, como na posição (3,1).

Após a correção efetua-se a projeção cilíndrica, e deslocam-se novamente as posições dos ponteiros dos pixels. No exemplo, o ponteiro do pixel (4,1) é deslocado da posição (4,2) para a posição (4,1) no Mapa de Posicionamento (Figura 51 – c). Alguns ponteiros poderão repetir posições para completar as lacunas que possam surgir, como ocorre na posição (4,2) do Mapa de Posicionamento, que passou a conter o mesmo endereço que a posição (4,1).

Por fim, faz-se o alinhamento. O pixel (4,1) é deslocado para a posição (5,2) no Mapa de Posicionamento (Figura 51 – d).

A rotação de 90° e a costura são efetuadas no Procedimento de Montagem do Panorama, no momento da montagem da imagem panorâmica.

Uma vez montado o Mapa de Posicionamento de cada câmera, termina o Procedimento de Calibração de Imagens, o qual só será executado novamente por intervenção do usuário.

O Procedimento de Calibração de Imagens pode ser considerado como o “coração” do sistema, pois graças a ele se consegue diminuir a complexidade computacional em níveis passíveis de processamento pela atual tecnologia de computadores, pois o posicionamento de cada pixel se dará em único passo, como se verá no próximo item.

3.6.1.2 Procedimento de Montagem do Panorama

Este procedimento tem os seguintes processos a cada ciclo:

- Leitura dos quadros capturados pelas câmeras. Cada câmera que está no barramento *firewire* (IEEE-1394) será lida até 15 vezes por segundo, e a cada leitura a imagem será armazenada em uma matriz (*buffer*), criada antecipadamente na memória do computador. Existe uma matriz para cada câmera (Figura 52).

Figura 52: Matrizes de imagens na memória dos microcomputadores

- Localização das coordenadas dos pixels no Mapa de Posicionamento de cada câmera. O Mapa de Posicionamento é uma matriz de ponteiros, onde cada elemento aponta a posição correta de cada pixel da matriz de imagem. Após o carregamento das matrizes de imagem na memória, o Mapa de Posicionamento de cada câmera é varrido da primeira à última célula, onde se verifica para qual pixel ele aponta. Se a posição for

uma posição não-nula, o pixel daquela posição na matriz de imagem será pintado na posição (m, n) da imagem panorâmica. Se não for uma posição não-nula, este pixel será pintado de preto (Figura 53).

Figura 53: Localização de coordenadas no Mapa de Posicionamento

Usando-se como exemplo a Figura 51, o endereço $(5,2)$ do Mapa de Posicionamento tem um ponteiro que indica a posição $(4,1)$ da Matriz de Pixels, o qual será utilizado para determinar o pixel da posição $(4,1)$ da Imagem Panorâmica, pintado em azul, conforme o exemplo.

- Rotação em 90° e costura de imagens. Usou-se a abordagem de ler os pixels de cada imagem na horizontal e colocá-los diretamente no endereço de memória do panorama, na vertical. Desta forma, ao mesmo tempo em que se executa a costura do panorama, rotacionam-se as imagens em 90° .

A grande vantagem é que não se gasta tempo algum além do usado na montagem do panorama, pois não existe operação adicional. Para uma rotação de 90° , 180° e 270° não é necessário efetuar a interpolação dos pixels, porque estes apenas mudam de posição, da horizontal para a vertical, no quadro. Para outros ângulos, a posição dos pixels no quadro poderá ser a de um número fracionário, o que demanda uma interpolação para ajuste de posição (Figura 54).

Figura 54: Rotação de imagens na construção do vídeo panorâmico

- Após obter-se a posição correta do pixel, com a utilização do Mapa de Posicionamento, monta-se a imagem panorâmica, colocando-se cada pixel na posição apontada pelo Mapa de Posicionamento, já rotacionada em 90°.
- Compressão de imagens. Como já se explicou, a transmissão do quadro panorâmico sem alguma forma de compressão demandaria grande largura de banda de transmissão, algo em torno de 885 Mbits por segundo, caso se use a velocidade de 15 quadros por segundo.

$$640 \bullet 480 \text{ pixels} \bullet 8 \text{ câmeras} \bullet 3 \text{ bytes / pixel} \bullet 8 \text{ bits} \bullet 15 \text{ quadros / s} = 884.736.000 \text{ bits / s}$$

(33)

Como não é possível transmitir adequadamente tal fluxo de dados, torna-se necessário comprimir o arquivo antes da transmissão. A alternativa utilizada foi compactar os quadros em JPEG e transmiti-los individualmente, método conhecido como MJPEG.

- Transmissão de imagens. Para a transmissão de imagens, tanto do Microcomputador Servidor para o Microcomputador Cliente / Servidor como deste para o aglomerado de computadores da CAVERNA Digital® da EPUSP, via Internet 2, utilizou-se o protocolo de comunicação TCP/IP.

3.6.2 Módulo de Composição

O Módulo de Composição tem dupla função: a de cliente, na recepção de imagens geradas pelos vários Microcomputadores Servidores, e a de servidor, que transmite a imagem panorâmica final para o Módulo de Exibição da CAVERNA Digital® da EPUSP.

Os computadores que rodam o Módulo de Aquisição são ligados por uma rede local GigaBit a um computador que roda o Módulo de Composição. Este, por sua vez, está ligado à rede de alta velocidade Internet 2.

Seus principais processos são:

- Recepção de imagens. Receber as imagens geradas pelos Microcomputadores Servidores;
- Alinhamento e costura de imagens. No Módulo de Composição é necessário alinhar os dois ou três conjuntos de imagem e costurá-las para obter a imagem panorâmica de 360°. Cada Microcomputador Servidor envia imagens em 180° (quando se usam dois microcomputadores) ou 120° (quando se usam três microcomputadores). Existem duas ou três imagens a serem costuradas;
- Compressão e retransmissão de imagens. Novamente usa-se compressão JPEG para comprimir os quadros.

3.6.3 Módulo de Exibição

O próximo passo é o recebimento de imagens pelo Módulo de Exibição, localizado na CAVERNA Digital® da EPUSP.

Neste módulo as imagens são:

- Descomprimidas: Os quadros estão comprimidos no formato JPEG, e é necessário transferi-los para a memória do microcomputador no formato *BitMap*, para tratamento e exibição;

- Divididas: Para projetar o panorama na CAVERNA Digital®, que tem quatro lados, é necessário dividir a imagem panorâmica em quatro imagens de tamanho aproximadamente igual;
- Projetadas em superfície plana: Como a imagem está projetada cilindricamente, é necessário projetá-la em superfície plana para que não ocorra distorção durante a exibição;
- Exibidas nas telas: O aglomerado de computadores da CAVERNA Digital® enviará as quatro imagens, uma para cada um dos projetores da caverna, para exibição.

3.7 Conclusão

Este capítulo dedica-se a apresentar, de forma geral, a proposta de solução para geração de vídeo panorâmico 360°, ao vivo, para uso em aplicações de telepresença, os requisitos, as restrições, a arquitetura de *hardware* e *software* e seu detalhamento.

As principais técnicas incluídas no desenvolvimento deste projeto foram explanadas no capítulo 2, e a forma como se integraram e desenvolveram técnicas, equipamentos e programas encontram-se detalhados no capítulo 5.

4 IMPLEMENTAÇÃO E TESTE COM IMAGENS PANORÂMICAS ESTÁTICAS

4.1 Introdução

Este capítulo descreve as primeiras implementações e testes realizados a fim de investigar os materiais e métodos para estabelecer uma metodologia de aquisição, armazenamento e projeção de imagens panorâmicas em 360° (HU et al., 2002). Os testes iniciais foram realizados com imagens estáticas.

4.2 Aquisição de imagens

A idéia inicial era obter duas imagens panorâmicas de 360° e posteriormente processá-las, utilizando o método *Anaglyph* (capítulo 2) para gerar uma única imagem panorâmica e estereoscópica.

Nas primeiras tentativas de aquisição de imagens, estas foram obtidas por meio da rotação de duas máquinas fotográficas analógicas convencionais de 35 mm, montadas em um mecanismo apoiado sobre um tripé. O objetivo era gerar fotos panorâmicas e estereoscópicas simultaneamente.

As câmeras fotográficas utilizadas foram uma Canon® FTb com lente 55mm e uma Canon® EOS-5 com lente 28-105, regulada para distância focal de 55mm (Figura 55).

Figura 55: Máquinas montadas sobre dispositivo e tripé

A disposição não funcionou a contento porque, apesar de ambas as máquinas estarem reguladas para as mesmas distâncias focais, utilizavam lentes diferentes, resultando em fotos com diferentes tamanhos de quadro.

Com quadros de tamanhos diferentes seria possível construir os dois panoramas, pois cada panorama seria o resultado da seqüência de fotos de uma única máquina, mas isso afetaria a montagem da imagem estereoscópica, uma vez que não haveria correspondência entre a imagem dos dois panoramas.

Também se questionou o tempo perdido na revelação do filme, que inviabilizava a eventual correção de fotograma com problema. Para contornar isso, implementou-se outra solução, que se mostrou até mais econômica, pois usa uma única câmera.

Realizou-se a montagem com uma única máquina, acoplada a um mecanismo especialmente desenvolvido, que permitia deslocamento lateral de aproximadamente 65 mm. Esta distância, recomendada para gerar um efeito estereoscópico confortável para o ser humano, corresponde aproximadamente à distância entre os centros ópticos dos dois olhos, nos seres humanos (LIPTON, 1982) (Figura 56).

Figura 56: Ilustração do mecanismo de deslocamento

Para facilitar a aquisição e a visualização, passou-se a utilizar uma câmera digital, da marca Minolta®. Recomenda-se que o tripé utilizado possua marcação em graus no eixo horizontal, para permitir melhor precisão nas tomadas fotográficas.

O número de fotos necessárias para gerar a imagem com 360° varia em função da distância focal da lente utilizada. Uma distância focal de 35 mm permite um ângulo de visão horizontal (HFOV) de 54° (BURIAN, 1999; CANON, 1998). Além disso, é necessária uma área de sobreposição nas fotos para possibilitar a montagem da imagem panorâmica.

Para os testes, adquiriram-se imagens na CAVERNA Digital® da EPUSP (Figura 57), na casa do autor (Figura 58 e Figura 59), na Sala São Paulo (sala de espetáculos em que se apresenta a Orquestra Sinfônica do Estado de São Paulo) (Figura 60), e no campus da USP (HU et al., 2002).

Figura 57: Fotografias costuradas gerando um panorama de 360° da CAVERNA Digital®

Figura 58: Série de fotografias sem costura

Figura 59: Fotografias costuradas gerando um panorama de 360°

Obtiveram-se assim duas séries de fotos para cada panorama desejado, uma correspondente ao olho esquerdo e a outra, ao olho direito. Todas as séries sofreram o processo de costura descrito a seguir, gerando imagens panorâmicas em 360°, comprimidas em JPEG.

Após processar os primeiros panoramas, constatou-se que não seria possível gerar o par estereoscópico, porque os panoramas continham distorções geradas pelo *software* de costura de imagens, que podem ser vistas na Figura 59. Novamente, não havia correspondência entre os dois panoramas, de forma a permitir a construção da imagem estereoscópica.

Em função disso, decidiu-se pelo abandono momentâneo da montagem estereoscópica, para concentrar esforços na montagem de imagens panorâmicas.

4.3 Montagem do panorama

Após a aquisição das séries de fotos, é necessário montá-las para transformá-las em imagens panorâmicas, processo conhecido como *stitch* (costura).

A costura deve ser efetuada sobrepondo os trechos comuns de fotos (recomenda-se uma sobreposição de 15% a 30%) (CURTIN, 2001) e a correção de cores e da intensidade luminosa.

A Figura 60 apresenta uma sequência de três fotos e a Figura 61, o resultado após a costura.

Figura 60: Série de fotos adquiridas da Sala São Paulo

Figura 61: Imagem montada em panorama

Para realizar a costura foi utilizado o *software* PanEdit® (PANEDIT, 2002). Cada foto foi adquirida com resolução de 2048 x 1536 pixels em JPEG, com 24 bits de profundidade de cores. Cada série compunha-se por dezoito fotos, cada uma cobrindo um ângulo de 20°, e mais uma área de sobreposição de aproximadamente 20°.

A utilização do *software* apresentou problemas devido ao tamanho e à quantidade de imagens. Não foi possível efetuar os testes utilizando um microcomputador PC AMD® 700, com 128Mbytes de memória. Acrescentando um pente de memória, com 512 Mbytes,

completando 640 Mbytes de memória total, foi possível realizar a costura das dezoito fotos na resolução original.

Além disso, desabilitou-se a memória virtual do Windows 98®. O sistema operacional, dessa forma, não paginou mais em disco, somente em memória, aumentando significativamente o desempenho do sistema em relação à situação anterior; no entanto, a costura de uma série de fotos ainda consumia aproximadamente 3 horas.

Com o programa PanEdit®, dada uma série de fotos, foi possível efetuar automaticamente a montagem e a correção de cores de cada série, apenas com alguns ajustes manuais.

A Figura 62 mostra uma série antes da montagem e a Figura 63, a imagem panorâmica correspondente da Sala São Paulo (HU et al., 2002).

Figura 62: Série de fotos adquiridas da Sala São Paulo

Figura 63: Imagem montada da Sala São Paulo

Após a carga do *software* PanEdit® na memória, este solicita a indicação dos arquivos que comporão a imagem. Devem-se montar os arquivos na sequência exata de composição da imagem panorâmica, na janela auxiliar que surge no canto inferior da tela. O segundo passo, após a carga dos arquivos, é indicar na janela se a imagem é uma panorâmica de 360° ou o ângulo que esta abrangerá.

Após estes passos, inicia-se o processo de costura de imagens. O panorama será gerado em projeção cilíndrica. Após o processamento é possível salvar a imagem panorâmica, que será a composição de todas as fotos incluídas no início do processo. A recomendação é salvar no formato JPEG, pois o tamanho da imagem gerada será menor do que com os outros formatos disponíveis no PanEdit®.

4.4 Projeção de imagens

Para a apresentação na CAVERNA Digital® utilizou-se a sequência de procedimentos que se segue.

A imagem gerada apresentava um tamanho de 3069 x 640 pixels em 24 bits de profundidade de cores.

Primeiro seccionou-se a imagem panorâmica original em quatro imagens de tamanhos aproximadamente iguais, para serem apresentadas uma em cada tela da CAVERNA Digital® da EPUSP. Para isso utilizou-se o *software* PhotoShop® versão 5.0.

Após a divisão em quatro imagens, iniciaram-se os testes. Nos primeiros já foi possível notar que o tamanho das imagens não era compatível com o apresentado nas telas. Foram então redimensionadas para 1048 x 840, que é a maior resolução suportada pelos projetores, utilizando o *software* Image View®. Novamente, utilizou-se no redimensionamento o PhotoShop® versão 5.0.

Não houve a preocupação, naquele momento, de fazer nenhuma correção de projeção cilíndrica para projeção plana, mesmo sabendo-se que ocorreriam distorções nas imagens projetadas nas telas.

Após vários ajustes manuais da imagem foi possível apresentar toda a Sala São Paulo na CAVERNA Digital® em imagem panorâmica estática (Figura 64 e Figura 65).

Figura 64: Esquema ilustrando a Sala São Paulo em projeção cilíndrica

Figura 65: Esquema ilustrando a Sala São Paulo na CAVERNA Digital® da EPUSP

Trabalhou-se da mesma forma para o processamento de outras imagens panorâmicas adquiridas na casa do autor, no *campus* da USP e dentro do prédio da CAVERNA Digital® da EPUSP.

4.5 Conclusão

Este capítulo descreveu implementações e testes preliminares, cujo sucesso garantiu que se continuasse na linha de pesquisa pretendida. Os resultados obtidos foram publicados no 5th Brazilian Symposium on Virtual Reality (HU et al., 2002), e forneceram subsídios para a especificação, identificação de requisitos e restrições, bem como para o detalhamento do sistema proposto.

5 IMPLEMENTAÇÃO E TESTES COM VÍDEOS PANORÂMICOS

5.1 Introdução

Este capítulo descreve os detalhes de implementação e os testes realizados visando à transmissão de imagens panorâmicas 360° ao vivo, a distância, fazendo da CAVERNA Digital® da EPUSP um ambiente de telepresença, com a projeção de vídeos panorâmicos em 360°.

Os programas foram escritos em C++ e utilizam a biblioteca FlyCapture, oferecida pela Point Grey Research®, fabricante das câmeras DragonFly (POINT GREY RESEARCH INC, 2004), e a biblioteca OpenCV, da Intel® (OPENCV, 2004).

O foco principal deste trabalho está voltado para a construção dos Módulos de Aquisição e de Composição, que tratam especificamente dos problemas de aquisição, correção, costura e transmissão de vídeos (imagens dinâmicas).

5.2 Equipamento de aquisição de vídeo

Desenvolveu-se para o projeto um equipamento de aquisição de imagens denominado **anel de câmeras**, que utiliza uma série de câmeras digitais dispostas em círculo, cada uma delas cobrindo um determinado ângulo de visão. Montaram-se dois conjuntos, um com doze câmeras e outro com oito. (Figura 66 e Figura 67).

Figura 66: Diagrama esquemático dos anéis de câmeras, com doze e com oito câmeras

Figura 67: Anel de doze câmeras montadas em um tripé

As câmeras foram colocadas deitadas, em ângulo de 90° , conforme pode ser visto na Figura 67, pois com este posicionamento tem-se 480 pixels na horizontal e 640 pixels na vertical para cada câmera, proporcionando aquisição de maior quantidade de pixels na vertical (640 pixels).

De acordo com a especificação da lente de 4 mm de distancia focal (POINT GREY RESEARCH INC, 2004), tem-se um ângulo de cobertura de $68,8^\circ$ na horizontal e de $51,9^\circ$ na vertical (com a câmera na posição normal).

Inicialmente, montou-se um conjunto com doze câmeras. Para as câmeras posicionadas em 90° , o ângulo de abertura é de $51,9^\circ$ na horizontal. Como para um anel de

doze câmeras são necessários apenas 30° de abertura de cada câmera, é possível deduzir que a área de sobreposição é de 42,2%.

$$\left(1 - \frac{30^\circ}{51,9^\circ}\right) \bullet 100\% = 42,20\% \quad (34)$$

É possível calcular aproximadamente a resolução horizontal estimada do anel de câmeras:

$$480 \text{ pixels} \bullet \frac{30^\circ}{51,9^\circ} \bullet 12 \text{ câmeras} = 3329,48 \text{ pixels} \quad (35)$$

Após alguns testes, concluiu-se que era possível diminuir o anel de câmeras para apenas oito, e ainda assim restar área de sobreposição suficiente para efetuar a costura — cada câmera cobrindo 45°. A sobreposição será de:

$$\left(1 - \frac{45^\circ}{51,9^\circ}\right) \bullet 100 = 13,29\% \quad (36)$$

A resolução horizontal continua igual:

$$480 \text{ pixels} \bullet \frac{45^\circ}{51,9^\circ} \bullet 8 \text{ câmeras} = 3329,48 \text{ pixels} \quad (37)$$

A resolução total do anel de câmeras é de aproximadamente 3.330 pixels na horizontal por 640 pixels na vertical. Com esta resolução, em uma tela de 3 m de lado, tem-se o seguinte tamanho de pixel:

$$\text{Horizontal: } \frac{4 \text{ lados}}{3330 \text{ pixels}} \bullet 300 \text{ cm} = 0,36 \text{ cm} / \text{ pixel} \quad (38)$$

$$\text{Vertical: } \frac{1}{640 \text{ pixels}} \bullet 300 \text{ cm} = 0,47 \text{ cm} / \text{ pixel} \quad (39)$$

Quanto à decisão sobre qual marca e modelo de câmeras utilizar, optou-se pelas já existentes no LSI da EPUSP, um conjunto de câmeras Dragonfly®, da Point Grey Research®. Após pesquisa com outros fabricantes de câmeras, não se encontrou nenhuma que apresentasse vantagem significativa sobre as já existentes, de forma a justificar nova compra.

As principais características das câmeras Dragonfly® são:

- Possibilidade de sincronização de várias câmeras, tanto sobre um mesmo barramento como em barramentos diferentes;
- Utilização de barramento de alta velocidade, o barramento *firewire* (IEEE-1394), que permite a ligação de quatro câmeras simultaneamente;
- Resolução de cada câmera de 640 x 480 pixels;
- Velocidade de 15 quadros por segundo;
- Dimensões físicas compatíveis para a montagem de um dispositivo portátil com várias câmeras;
- Bibliotecas de *software*, com os códigos fontes em C++, o que facilita o desenvolvimento de aplicativos e programas utilizando as câmeras (Figura 68) (POINT GREY RESEARCH INC, 2004);
- Disponibilidade imediata e custo já absorvido.

Figura 68: Câmera *Dragonfly* (POINT GREY RESEARCH INC, 2004)

5.3 Aglomerado de computadores e rede

É possível utilizar oito ou doze câmeras sincronizadas, e a quantidade de câmeras delimita o número de computadores utilizados no aglomerado.

O barramento *firewire* (IEEE-1394) permite a conexão de até quatro câmeras simultaneamente no mesmo barramento, acopladas por um *hub firewire* (Figura 69).

Figura 69: *Hub firewire*

Logo, para oito câmeras são necessários dois microcomputadores, com a função de Microcomputador Servidor, e um com a função de Microcomputador Cliente / Servidor. Para doze câmeras, três microcomputadores com a função de Microcomputador Servidor e um, com a função de Microcomputador Cliente / Servidor.

Montou-se inicialmente um aglomerado de computadores composto por quatro microcomputadores da marca Dell Dimension®, cada um com duas CPUs Intel® Pentium 4® de 1.7 MHz, todos com barramento *firewire*, sendo que em cada um dos barramentos conectaram-se quatro câmeras, por meio de um *hub firewire* (Figura 70).

Figura 70: Aglomerado de computadores com os equipamentos fora da estante (*rack*)

Posteriormente, aproveitaram-se apenas dois destes microcomputadores para atuarem com a função de Microcomputador Cliente / Servidor e se substituíram os outros dois por um microcomputador marca Itaotec, com duas CPUs Intel® Pentium 4® de 3.06 MHz, servindo como Microcomputador Cliente / Servidor. A mudança objetivou obter-se maior velocidade de processamento na execução do Módulo de Composição.

Três tecnologias de rede foram utilizadas para fazer a captura e transmissão de imagens:

- Para a comunicação entre as câmeras e o microcomputador foi utilizado um *hub firewire* para cada conjunto de quatro câmeras. Das câmeras até o *hub* e deste até o microcomputador usou-se o barramento *firewire* (IEEE-1394), por ser aquele utilizado pelas câmeras;
- Para a ligação dos Microcomputadores Servidores com o Microcomputador Cliente / Servidor sugeriu-se a utilização de um barramento GigaBit Ethernet, com velocidade de 1Gbps (IEEE 802.3z). Uma rede com tal velocidade atenderia a demanda de banda de transmissão gerada pelos conjuntos de câmeras. Na montagem da rede serão

utilizados um *backbone* (fiação) em fibra óptica, placas de rede com interface óptica e um *hub* Intel de alta velocidade. No entanto, os testes foram feitos com equipamentos Ethernet de 100 Mbps, por indisponibilidade de equipamentos GigaBit Ethernet;

- Para ligar o ambiente remoto com a CAVERNA Digital® da EPUSP foi proposta a utilização da Internet 2, devido à largura de banda oferecida.

5.4 Módulo de Aquisição

5.4.1 Determinação das constantes de lente

Utilizaram-se o *software* MatLab® e o aplicativo *Camera Calibration Toolbox for Matlab*, desenvolvido por Bouguet (2004). O aplicativo apresenta uma série de imagens tiradas com a câmera e a lente que se deseja calibrar. As imagens devem ser necessariamente de um tabuleiro de xadrez padrão, posicionado em diversos ângulos (Figura 71).

Figura 71: Imagens de um tabuleiro de xadrez, utilizadas na calibração da câmera

Para cada imagem do tabuleiro, o sistema solicita a posição de seus vértices, e, de posse de todos os pontos de todas as imagens, efetua o cálculo das constantes da lente.

Utilizaram-se esses parâmetros no Procedimento de Calibração de Imagens, para correção das imagens capturadas pelas câmeras.

5.4.2 Procedimento de Calibração de Imagens

O procedimento deve ser feito antes que se execute o Procedimento de Montagem do Panorama, pois ele gera o Mapa de Posicionamento, que será utilizado durante o processamento da imagem.

Caso não se efetue o Procedimento de Calibração de Imagens antes do Procedimento de Montagem do Panorama, as imagens panorâmicas não serão costuradas, mas sim agrupadas lado a lado, sem nenhuma correção e nenhum corte.

Depois que se inicia a utilização do Módulo de Aquisição é apresentado o painel de controle do *software*. Deve-se acionar o botão <Open Stitch> para iniciar o Procedimento de Calibração de Imagens. Após a conclusão do procedimento, deve-se acionar o botão <Save Stitch> a fim de gravar em disco os parâmetros calculados, para uso posterior.

Figura 72: Painel do Módulo de Aquisição

As etapas executadas neste procedimento estão descritas a seguir.

5.4.2.1 Correção de distorções

Após a determinação das constantes de lente, deve-se utilizá-las sobre a imagem capturada por câmera.

É importante notar que cada lente tem os seus coeficientes de distorção. Apesar de todas as lentes usadas neste projeto serem iguais, pode ocorrer que tenham coeficientes diferentes, devido ao processo de fabricação. É necessário calcular para cada uma das câmeras os coeficientes de distorção, e aplicá-los individualmente.

A imagem capturada pela câmera será chamada de **Matriz de Pixels da Câmera**. Esta matriz ainda não sofreu nenhuma correção. O conteúdo de cada elemento seu é a cor do pixel, expresso no padrão RGB e com 24 bits de profundidade de cores.

Na correção do posicionamento dos pixels aplicam-se as equações:

$$KK = \begin{bmatrix} fc_1 & \alpha_c \cdot fc_1 & cc_1 \\ 0 & fc_2 & cc_2 \\ 0 & 0 & 1 \end{bmatrix} \quad (40)$$

$$x_d = (K_1 \cdot r^2 + K_2 \cdot r^4 + K_5 \cdot r^6) \cdot x' + 2 \cdot K_3 \cdot x \cdot y + K_4 \cdot (r^2 + 2 \cdot x^2) \quad (41)$$

$$y_d = (K_1 \cdot r^2 + K_2 \cdot r^4 + K_5 \cdot r^6) \cdot y' + K_3 \cdot (r^2 + 2 \cdot y^2) + 2 \cdot K_4 \cdot x \cdot y \quad (42)$$

$$\begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} = KK \cdot \begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix} \quad (43)$$

já discutidas no item 2.6.3 (Correção de distorções de lentes de câmeras).

A aplicação destas equações a cada quadro resulta numa quantidade de operações matemáticas que inviabiliza o projeto. Trabalha-se em cada Microcomputador Servidor com 15 quadros por segundo, com o tamanho de 640 x 480 pixels, com quatro câmeras, ou seja, 18.432.000 pixels a serem tratados por segundo. Não existe atualmente equipamento comercial que consiga processar todas as transformações geométricas necessárias para o correto posicionamento dos pixels com este fluxo de informações.

No entanto, os coeficientes de distorção não se alteram com o passar do tempo. Eles fazem parte das características físicas das lentes. Isso significa que cada pixel se desloca para um outro local na matriz de pixels, em função de sua posição, e o deslocamento será sempre o mesmo.

Para solucionar o problema de processamento, criou-se para cada câmera uma matriz de mesmo tamanho que a Matriz de Pixels da Câmera (640 x 480 elementos), na qual cada elemento é um ponteiro que mostra onde os pixels da Matriz de Pixels devem se posicionar (Figura 73). Esta matriz será chamada de **Mapa de Posicionamento**.

MATRIZ DE PIXELS DA CÂMERA								
	1	2	3	4	5	6	7	8
1	Red	Brown	Green	Blue	Purple	Orange	Pink	Yellow
2	Brown	Green	Blue	Purple	Orange	Pink	Yellow	Red
3	Green	Blue	Purple	Orange	Pink	Yellow	Red	Brown
4	Blue	Purple	Orange	Pink	Yellow	Red	Brown	Green
5	Purple	Orange	Pink	Yellow	Red	Brown	Green	Blue
6	Orange	Pink	Yellow	Red	Brown	Green	Blue	Purple
7	Pink	Yellow	Red	Brown	Green	Blue	Purple	Orange
8	Yellow	Red	Brown	Green	Blue	Purple	Orange	Pink

MAPA DE POSICIONAMENTO COM AS POSIÇÕES ORIGINAIS								
	1	2	3	4	5	6	7	8
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)	(1,7)	(1,8)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)	(2,7)	(2,8)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)	(3,7)	(3,8)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)	(4,7)	(4,8)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)	(5,7)	(5,8)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)	(6,7)	(6,8)
7	(7,1)	(7,2)	(7,3)	(7,4)	(7,5)	(7,6)	(7,7)	(7,8)
8	(8,1)	(8,2)	(8,3)	(8,4)	(8,5)	(8,6)	(8,7)	(8,8)

Figura 73: Mapa de posicionamento inicial

A Figura 73 apresenta a matriz de pixels, que foi capturada por uma das câmeras, e seu tamanho é de 8 x 8 pixels.

O Mapa de Posicionamento não é criado durante o processamento, mas antes, em procedimento separado, o **Procedimento de Calibração de Imagens**. As matrizes poderão ser alteradas apenas neste procedimento.

No momento da criação do Mapa de Posicionamento, todos os seus elementos apontam para posições coincidentes na Matriz de Pixels, sem nenhum deslocamento, como visto na Figura 73.

Após o cálculo das novas posições dos pixels, os ponteiros do Mapa de Posicionamento são alterados para mostrar a nova posição. Para ilustrar isso, alteraram-se as

posições dos pixels na Matriz de Pixels da Figura 74, mostrando como esta ficará após a aplicação do Mapa de Posicionamento, com a correção da distorção da lente.

Figura 74: Mapa de Posicionamento com a correção de distorções da lente

No exemplo, inicialmente o ponteiro da posição (4, 2) do Mapa de Posicionamento apontava para a posição (4, 2) da Matriz de Pixels da Câmera (Figura 73). Após efetuar os cálculos de correção de distorções da lente, conclui-se que nesta posição deveria estar o pixel da posição (4, 1) da Matriz de Pixels da Câmera. Colocou-se então este novo ponteiro na posição (4, 2) do Mapa de Posicionamento (elipse vermelha na Figura 74). Apesar de ser conveniente para melhorar o resultado visual, optou-se por não efetuar nenhuma interpolação para não comprometer a velocidade de processamento.

Quando o deslocamento resultar em uma posição fora da imagem, como as posições (1, 3), (1, 4), (1, 5) e (1, 6), nela será colocado um ponteiro nulo, e na montagem do panorama utilizar-se-á um pixel preto, representando o *background* (a cor de fundo).

A correção das imagens ocorrerá durante a execução do Procedimento de Montagem do Panorama, quando se fará a aplicação do Mapa de Posicionamento sobre a Matriz de Pixels das imagens capturadas pelas câmeras (Figura 75 e Figura 76).

Figura 75: Imagens capturadas por duas câmeras

Figura 76: Imagens sem correção, rotacionadas em 90° para melhor visualização

O resultado da correção de distorções de lentes, com a aplicação do Mapa de Posicionamento, pode ser observado na Figura 77.

Figura 77: Imagens corrigidas, rotacionadas em 90° para melhor visualização

5.4.2.2 Projeção cilíndrica

Como já explicado no item 2.6.4 (Projeção cilíndrica), é necessário projetar cilíndricamente as imagens, já corrigidas de suas distorções. Tal projeção faz-se necessária para que se possam costurar corretamente as imagens.

A projeção cilíndrica alterará novamente as posições dos pixels e será outra vez necessário tratar a restrição de volume computacional. O cálculo individual da posição de cada pixel demandará algumas dezenas de operações matemáticas por pixel, as quais, multiplicadas pela quantidade de pixels envolvidos (18,4 milhões por segundo), inviabilizará mais uma vez o processo.

Ao analisar essa transformação geométrica, percebe-se que se repete, ou seja, uma vez encontradas as novas posições dos pixels, estas serão as mesmas em todo o processamento.

Utiliza-se a mesma estratégia para a correção das distorções de lente. Alteram-se as posições dos ponteiros do Mapa de Posicionamento, já modificados pela aplicação das

equações de correção de distorção da lente, e tem-se o Mapa de Posicionamento corrigido e projetado cilindricamente (Figura 78).

Figura 78: Mapa de Posicionamento, corrigida a distorção da imagem e projetada cilindricamente

Da mesma forma que no exemplo anterior (Figura 74), o Mapa de Posicionamento já está corrigido. Após a aplicação das equações de projeção cilíndrica, tem-se um novo deslocamento na posição dos ponteiros do Mapa de Posicionamento. A posição (4,1), que estava com um ponteiro nulo (Figura 74), volta a dirigir-se para a posição (4,1) (elipse vermelha na Figura 78). O resultado da projeção cilíndrica sobre a imagem já corrigida aparece na Figura 79. As equações utilizadas estão apresentadas no item 2.6.4.

Figura 79: Imagens corrigidas e projetadas cilindricamente

Caso não se projete cilíndricamente a imagem, o ajuste da costura não será perfeito, como se pode ver na Figura 80. Nesta imagem panorâmica não se fez a projeção cilíndrica das imagens, apenas a correção das distorções de lente, e é possível perceber, principalmente no teto, que as linhas não se encontram, apresentando quebras de continuidade.

Figura 80: Imagens costuradas sem terem sido previamente projetadas cilíndricamente

5.4.2.3 Alinhamento de imagens

A próxima etapa do processamento é o alinhamento de imagens, duas a duas, para posterior costura. O alinhamento novamente altera as posições dos pixels.

Como se viu no item (2.6.6), efetua-se uma multiplicação matricial entre a matriz fundamental (F) e a matriz da posição de cada pixel (P).

$$P' = F \bullet P \quad (44)$$

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & m_8 \end{bmatrix} \bullet \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (45)$$

No cálculo da matriz fundamental utilizam-se quatro pontos para cada imagem, escolhidos manualmente durante o Procedimento de Calibração da Imagem.

A escolha se faz em tela, como mostra a Figura 81.

Figura 81: Pontos equivalentes definidos (oito pontos)

De posse das coordenadas destes oito pontos (quatro por imagem), calcula-se a matriz fundamental (F), através da resolução de um sistema linear de equações.

Adotando m_g (também conhecido como fator de escala do eixo z) como igual a um, portanto sem alteração na escala da figura em relação ao seu eixo z , tem-se um sistema linear com oito equações e oito incógnitas, passível de resolução por vários métodos matemáticos. Utiliza-se a biblioteca OpenCV (OPENCV, 2004) para calcular as incógnitas e determinar a Matriz Fundamental.

O alinhamento sempre será feito com pares de imagens e convencionou-se que no projeto o alinhamento da imagem dar-se-á na imagem do lado direito.

Mais uma vez há um gargalo no processamento: fazer nove multiplicações e seis somas para cada pixel (15 operações x 18,4 milhões de pixels por segundo, ou seja, 276 milhões de operações por segundo). No entanto, a menos que se altere a posição de uma câmera em relação à outra, o alinhamento será sempre igual. Como as câmeras estão fixadas no anel de câmeras, não sofrerão alterações de posicionamento relativo.

Após ter alterado as posições dos ponteiros de pixels no Mapa de Posicionamento, para corrigir a distorção da lente e projetar cilindricamente a imagem, efetua-se o alinhamento. É possível realizar as três transformações no mesmo Mapa de Posicionamento (Figura 82).

Figura 82: Mapa de Posicionamento, após o alinhamento

Utilizando o mesmo exemplo da Figura 74, após o alinhamento as posições dos ponteiros de pixel do Mapa de Posicionamento são rearranjadas, e a posição (5,2) passa a apontar para a posição (4,1) da Matriz de Pixels (elipse vermelha da Figura 82).

No Procedimento de Montagem do Panorama, as três transformações descritas anteriormente — correção, projeção cilíndrica e alinhamento —, serão efetuadas durante a operação de montagem da imagem panorâmica, sem consumir CPU na execução de cálculos de transformação geométrica, dado que estes já foram feitos previamente e se refletem na posição dos ponteiros do Mapa de Posicionamento.

Outro parâmetro extraído neste procedimento é a posição de corte das imagens, da direita e da esquerda, para poder encaixá-las duas a duas. O alinhamento será feito com pares de imagens e a transformação da imagem dar-se-á na imagem do lado direito.

A Figura 83 mostra a imagem à direita alinhada. O Mapa de Posicionamento alterado é sempre o relativo à imagem à direita.

Figura 83: Segunda imagem alinhada

Ainda outro parâmetro advém desse procedimento: a posição de corte da imagem da direita e da imagem da esquerda, a fim de que se possam encaixar as duas imagens. Para melhor compreensão, cabe um exemplo: partindo da premissa de que as duas imagens que serão alinhadas têm 480 pixels na horizontal por 640 na vertical (já rotacionadas), e que foram escolhidos na imagem da esquerda os pontos $P_{e1} = (400, 180)$, $P_{e2} = (460, 190)$, $P_{e3} = (420, 500)$ e $P_{e4} = (470, 520)$ e na imagem da direita os pontos equivalentes à primeira imagem nas posições $P_{d1} = (120, 180)$, $P_{d2} = (200, 190)$, $P_{d3} = (140, 500)$ e $P_{d4} = (190, 520)$, para calcular a posição de corte de cada imagem usar-se-á a seguinte equação:

$$X_{Corte} = X_{Max} - \frac{(X_{Max} - X_{Min})}{2} \quad (46)$$

Para a imagem esquerda, tem-se:

$$470 - \frac{(470 - 400)}{2} = 435 \quad (47)$$

Para a imagem direita:

$$200 - \frac{(200 - 120)}{2} = 160 \quad (48)$$

Logo, efetua-se um corte na imagem da esquerda, na posição 435 da horizontal, e utilizam-se os pixels que estão à esquerda desta posição, ou seja, a imagem esquerda terá o tamanho de $480 - 35 = 435$ pixels por 640 pixels.

Na imagem direita, corta-se na posição 160 e utiliza-se a porção direita da imagem, a qual terá o tamanho de $480 - 160 = 320$ pixels por 640 pixels.

Neste cálculo utiliza-se a posição média dos pontos escolhidos para ser a posição de corte da imagem, e com isto diminuem as distorções, que ocorrem com mais intensidade nos extremos da imagem.

5.4.3 Procedimento de Montagem do Panorama

A principal preocupação da rotina foi reduzir ao máximo o tempo de processamento. Dentro da rotina tem-se um *loop* (laço), no qual é executada a leitura das câmeras, quadro a quadro, e o processamento das imagens, e dentro do *loop* existem as seguintes rotinas:

5.4.3.1 Leitura de câmeras

As imagens são carregadas e armazenadas temporariamente em matrizes, criadas no início do processamento. Há uma matriz para cada câmera. Para essas rotinas utilizaram-se, na confecção do programa, as bibliotecas do fabricante das câmeras.

5.4.3.2 Localização de coordenadas dos pixels no Mapa de Posicionamento

A operação se efetua pixel a pixel, para todos os pixels de todas as câmeras. Existe um Mapa de Posicionamento para cada matriz de imagem, e, por sua vez, uma matriz por câmera, como já visto na Figura 53.

5.4.3.3 Rotação em 90° e costura de imagens

As Figuras 76, 77, 79, 81 e 83 mostram as imagens em posição normal, mas esta é uma apresentação didática. As imagens estão na realidade deitadas, ou seja, necessitam ser rotacionadas em 90° para que se possa costurá-las.

Uma das alternativas para rotacionar a imagem é rotacionar o Mapa de Posicionamento, da mesma forma como foram feitas as transformações de correção, projeção e alinhamento, conforme apresentado na Figura 84.

Figura 84: Matriz de posicionamento, rotacionada em 90°

A alternativa citada acima foi descartada, porque exigiria alterar as dimensões do Mapa de Posicionamento, de 640x480 para 480x640, o que aumentaria a complexidade de montagem da matriz.

Outra alternativa é efetuar a rotação no momento da montagem da imagem panorâmica. Esta foi a opção adotada, por não aumentar a quantidade de operações na montagem da imagem panorâmica nem a complexidade da montagem da Matriz de Posicionamento (Figura 85).

Figura 85: Montagem final da imagem panorâmica

Calculando o número operações efetuadas no *loop* central, com a utilização do Mapa de Posicionamento, tem-se:

$$\frac{3330}{2} \bullet 640 \bullet 15 = 15.984.000 \text{ pixels / segundo} \quad (49)$$

No *loop* central da rotina de costura têm-se as seguintes operações para cada pixel:

- Oito somas de números inteiros;
- Dois produtos de números flutuantes;
- Uma comparação;
- Seis atribuições.

Podem-se considerar todas estas operações com dezessete operações em ponto flutuante, o que resulta em:

$$15.984.000 \bullet 17 = 271.728.000 \text{ operações / segundo} \quad (50)$$

quantidade passível de ser tratada computacionalmente com os equipamentos atuais.

Um exemplo de imagem panorâmica, resultado da costura de três imagens, é apresentado na Figura 86.

Figura 86: Imagem panorâmica, com imagens corrigidas, projetadas, alinhadas e costuradas

5.4.3.4 Compressão de imagens

Uma das opções para o desenvolvimento do projeto foi a utilização do algoritmo de compressão MPEG-4, pelas inúmeras vantagens que possui. Chegou-se a implantar uma rotina: após a montagem de cada imagem panorâmica, compactava-se esta numa sequência de imagens, de tipo “avi” - *audio & video interface* (interface de áudio e vídeo), utilizando um CODEC (Codificador e Decodificador) MPEG-4 instalado no equipamento.

Os **CODEC** são rotinas de codificação e decodificação de áudio e vídeo, as quais podem ser instaladas independentemente nos equipamentos, que lêem e gravam arquivos utilizando diferentes métodos de compressão. Podem ser produzidas por desenvolvedores independentes ou pelo desenvolvedor do sistema operacional.

Em termos de compressão e geração de arquivo, a rotina funcionou a contento, porém este produzia uma lentidão no sistema que impossibilitaria a transmissão de vídeo com taxa

de 15 quadros/s. Os filmes gerados, quando reproduzidos, mostravam as imagens em alta velocidade, o que ocorre quando se grava um arquivo com o cabeçalho indicando uma velocidade, mas as imagens são gravadas em velocidade menor.

O motivo provável do atraso é a complexidade do algoritmo de compressão MPEG-4, que demanda maior esforço computacional. O computador demorava muito tempo para compactar os quadros em MPEG-4, atrasando todo o processo.

A segunda alternativa foi compactar os quadros em JPEG e transmiti-los individualmente, método denominado MJPEG. Entre suas vantagens estão a facilidade de uso e a velocidade de compressão. Por outro lado, tem como desvantagem pior taxa de compressão, quando comparado ao método MPEG-4.

Utilizou-se a biblioteca “ezJ”, desenvolvida no LSI-EPUSP a partir de códigos-fontes originais do *Independent JPEG Group*. Optou-se por sua utilização pelo fato de ter rotinas que geram o arquivo JPEG em memória, pronto para ser transmitido. A biblioteca original do *Independent JPEG Group* não tem rotinas semelhantes, gerando o arquivo JPEG sempre em disco.

5.4.3.5 Transmissão de imagens

Para a transmissão das imagens utiliza-se o **protocolo de comunicação TCP/IP**, desenvolvido pelo Departamento de Defesa dos Estados Unidos da América, por volta dos anos 70, primeiramente para fins militares, depois acadêmicos, e por fim se disseminando pela Internet (ALMEIDA, 2004).

Na utilização desse protocolo, usou-se a biblioteca WinSock-2 da Microsoft®, que contém todas as rotinas necessárias para implementar a comunicação entre dois equipamentos, utilizando o protocolo TCP/IP.

As principais rotinas utilizadas estão na Figura 87, e serão descritas a seguir.

Figura 87: Principais rotinas do WinSock 2 utilizadas no sistema

No Módulo de Aquisição utilizam-se as seguintes rotinas para estabelecer a conexão e transmitir os dados:

- *Socket*: cria um objeto de comunicação;
- *Bind*: associa um nome ao *socket* criado;
- *Listen*: indica que o *socket* pode receber uma conexão;
- *Accept*: bloqueia o processo até que o processo CLIENTE estabeleça conexão;
- *Read*: recebe uma mensagem do *socket* remoto, autorizando o início da transmissão;
- *Write*: envia as imagens sequencialmente.

O vídeo é enviado quadro a quadro, cada um deles comprimido no formato JPEG, precedido de um cabeçalho que contém o tamanho do *frame* em bytes, a largura e a altura em pixels.

5.5 Módulo de Composição

5.5.1 Recepção de imagens

O Módulo de Composição tem duas atribuições. Uma é atuar como cliente, na recepção de imagens dos vários Módulos de Aquisição e na costura de imagens geradas por eles. A outra é atuar como Microcomputador Servidor, para poder transmitir as imagens panorâmicas completas para o Módulo de Exibição na CAVERNA Digital® da EPUSP.

A atribuição CLIENTE tem as seguintes rotinas (Figura 87):

- *Socket: Cria um objeto de comunicação.*
- *Connect: Estabelece uma comunicação entre o socket CLIENTE e o socket SERVIDOR.*
- *Write: Envia o pedido de envio de imagens.*
- *Read: Recebe as imagens.*

Recebe-se primeiro o cabeçalho, que contém o tamanho do *frame* em bytes, a largura e a altura em pixels.

O tamanho do *frame*, em bytes, permite verificar se o quadro chegou completo. Caso isso não aconteça, ele é descartado e utiliza-se na criação da imagem panorâmica o quadro imediatamente anterior.

A largura e a altura servem para verificar qual o tamanho da matriz de recepção da imagem. Caso haja alguma alteração durante o processamento, o sistema automaticamente redimensiona todas as matrizes.

O quadro recebido é armazenado em uma posição de memória para posterior descompressão da imagem.

5.5.2 Descompressão de imagens

Após a recepção e a verificação do tamanho do quadro, caso o tamanho se mantenha igual ao do quadro anterior efetua-se a descompressão e montagem da matriz de imagem. Caso contrário, redimensionam-se as matrizes de imagem para que estas possam acondicionar adequadamente os quadros recebidos e montar de maneira correta o quadro panorâmico.

Na descompressão utiliza-se a biblioteca “ezJ” com as funções de descompressão, a qual permite que se leia o arquivo JPEG direto de uma posição de memória, e não de arquivos gravados em disco.

5.5.3 Alinhamento e costura de imagens

Como se recebem duas imagens panorâmicas de 180° para o anel de oito câmeras (ou três de 120° para o anel de doze câmeras), cada uma das imagens necessita ser alinhada e costurada, para gerar o quadro panorâmico de 360° (Figura 88).

Figura 88: Montagem do quadro final

Como em toda montagem de panoramas, é necessário alinhar as duas ou três imagens. Note-se que nesta etapa não será necessário corrigir distorções de lente nem projetar cilíndricamente os quadros, porque as imagens já estão corrigidas e projetadas, etapa elaborada no Módulo de Aquisição, como mostra a Figura 88.

É necessário apenas fazer o alinhamento e a montagem final, ou seja, costurar as imagens recebidas em uma única imagem panorâmica.

Usa-se novamente o Mapa de Posicionamento para efetuar o alinhamento de imagens. A diferença deste mapa para o descrito no item 5.4.2 (Procedimento de Calibração de Imagens) é o tamanho, que neste módulo é pelo menos três vezes maior, e o posicionamento, pois como as imagens não vêm dos Módulos de Aquisição deitadas em 90° , mas sim em posição normal (Figura 89), não necessitam ser rotacionadas em 90° .

Figura 89: Costura final

Mais uma vez se escolhem quatro pontos equivalentes em cada uma das imagens recebidas e, de posse deles, calcula-se a matriz fundamental e utiliza-se a equação:

$$P' = F \bullet P \quad (51)$$

que equivale a:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & m_8 \end{bmatrix} \bullet \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (52)$$

de forma a se ter um Mapa de Posicionamento com as posições dos ponteiros alinhadas.

5.5.4 Compressão e retransmissão de imagens

Como já se explicou no item 5.5.1, o Módulo de Composição tem também a atribuição de transmitir a imagem panorâmica completa para a CAVERNA Digital® da EPUSP.

Novamente, usa-se compressão JPEG para comprimir os quadros e a biblioteca “ezJ”. Esta compressão é similar à descrita no item 5.4.3.4.

As rotinas de transmissão utilizadas neste módulo são as mesmas descritas no item 5.4.3.5, e encontram-se na Figura 87.

5.6 Conclusão

Neste capítulo se mostrou detalhadamente a metodologia utilizada para integrar todas as tecnologias apresentadas nos itens anteriores, os detalhes de implementação e as dificuldades e inovações que surgiram no decorrer do desenvolvimento do projeto.

O sistema completo é formado por um conjunto de equipamentos composto de um aglomerado, um anel de câmeras, *hubs* IEEE-1394 (*firewire*), *hub* GigaEthernet, cabos IEEE-1394 (*firewire*), cabos de fibra óptica, e, por fim, um aglomerado de computadores receptor e a CAVERNA Digital® da EPUSP.

Quanto ao *software*, foram desenvolvidos dois módulos, um denominado Módulo de Aquisição, instalado nos equipamentos ligados às câmeras, um denominado Módulo de

Composição, que consolida todas as imagens, e um que se encontra em desenvolvimento, denominado Módulo de Exibição, que é executado no aglomerado de computadores da CAVERNA Digital® da EPUSP.

Todos os equipamentos estão interligados por redes, permitindo a transmissão de imagens ao vivo.

6 DISCUSSÃO DOS RESULTADOS

6.1 Introdução

Este capítulo engloba a fase de implantação do projeto, as dificuldades encontradas e as alternativas adotadas para saná-las. Apresentam-se os resultados obtidos e os testes efetuados.

A medida de desempenho utilizada para a avaliação do sistema é a quantidade de quadros processados por segundo (*frame rate*), em função do tamanho do quadro processado e transmitido.

É importante ressaltar que a velocidade ideal está entre 15 e 30 quadros por segundo. Abaixo de 15 quadros inicia-se a percepção de descontinuidade do vídeo por parte do espectador. Acima de 30 quadros há desperdício na utilização de banda de comunicação, pois o ser humano não perceberá nenhuma diferença em um filme apresentado com 30, 40 ou mais quadros por segundo.

6.2 Ajustes finais dos programas

Em janeiro de 2006, concluiu-se a versão 1 dos Módulos de Aquisição e de Composição. Estes *softwares* foram escritos em C++, utilizando o compilador C++ *Visual Studio 6.0®*, da Microsoft®. A partir daí, os engenheiros Olavo da Rosa Belloc e Victor H. P. Gomes entraram no projeto para acelerar otimização do código e a montagem final dos equipamentos.

O desempenho da versão 1 estava ruim, obtendo velocidade de transmissão abaixo de 3,0 *frames* por segundo. O programa apresentava problemas de estabilidade, com travamento do Módulo de Aquisição. O Módulo de Composição também tinha problemas de estabilidade. Após alguns testes, concluiu-se que a causa dos travamentos era a transmissão dos dados dos

Módulos de Aquisição para o Módulo de Composição: qualquer flutuação na rede causava o travamento.

A primeira providência foi a conversão do sistema, de *Visual Studio 6.0®* para *Visual Studio Net 2005®*, e passou-se a utilizar bibliotecas mais recentes. Com a nova versão da biblioteca Winsock 2 e mais algumas alterações nas rotinas de transmissão e recepção, o sistema ficou estável.

Foram incluídas as rotinas de conversão de *BitMap* para JPEG no Módulo de Aquisição e de JPEG para *BitMap* no Módulo de Composição. Para essas rotinas foi utilizada a biblioteca “ezJ”.

A seguir, o painel de controle foi traduzido para o inglês, para possibilitar a apresentação do sistema no IEEE Virtual Reality Conference 2006, na seção Demonstração de Laboratórios (Figura 90, Figura 91) em março de 2006 em Alexandria, Virgínia, EUA (CABRAL, 2006).

Figura 90: Painel de controle do Módulo de Aquisição

Figura 91: Painel de controle do Módulo de Composição

Até o presente momento encontra-se em elaboração o Módulo de Exibição, que receberá as imagens transmitidas pelo Módulo de Composição e as exibirá nas paredes da CAVERNA Digital® da EPUSP. Os testes efetuados até o momento concentraram-se na comunicação entre os Módulos de Aquisição e de Composição. Desenvolveu-se uma função, executada no Microcomputador Cliente / Servidor, que simula o ambiente da CAVERNA Digital®, mapeando a imagem panorâmica em um cubo e exibindo o resultado na tela do próprio Microcomputador Cliente / Servidor, o que permite que os espectadores visualizem a imagem panorâmica simulando a apresentação na CAVERNA Digital®.

6.3 Montagem dos equipamentos para os testes

Utilizaram-se os seguintes equipamentos na montagem do sistema para testes com oito câmeras na CAVERNA Digital®:

- Dois microcomputadores da marca Dell Dimension®, cada um com duas CPUs Intel® Pentium 4® de 1.7 MHz, utilizados como Microcomputadores Servidores e executando o Módulo de Aquisição;
- Um microcomputador da marca Itautec, com duas CPUs Intel® Pentium 4® de 3.06 MHz, servindo como Microcomputador Cliente / Servidor e rodando o Módulo de Composição;
- Oito câmeras *firewire* Dragonfly®;
- Dois *hubs* *firewire*;
- Um *hub* Ethernet 100/10;
- Rede local Ethernet 100MBits.

Na elaboração dos testes utilizou-se uma rede local Ethernet de 100 Mbits por não se ter disponível uma rede GigaBit Ethernet.

Para a apresentação no IEEE Virtual Reality Conference 2006 a equipe montou um robô sobre o qual foi instalado o anel de câmeras (Figura 92).

Figura 92: Robô com o anel de câmeras no IEEE VR2006

Os equipamentos utilizados para a apresentação na conferência foram:

- Dois *notebooks* Dell Precision M60 - Centrino 1.7 GHz, 1GB de RAM, cada um operando como Microcomputador Servidor e processando o Módulo de Aquisição;
- Um *notebook* Sharp – Pentium 4 - 2.8GHz, 512MB de RAM trabalhando como Microcomputador Cliente / Servidor e processando o Módulo de Composição;
- Oito câmeras *firewire* Dragonfly®;
- Dois *hubs* *firewire*;
- Um *hub* Ethernet 100/10;
- Rede local Ethernet 100MBits.

Para simular a exibição no ambiente da CAVERNA Digital®, a imagem panorâmica 360° foi mapeada em um cubo e exibida na tela de um computador (a mesma máquina utilizada para executar o Módulo de Composição).

6.4 Resultados obtidos

Nos equipamentos montados na CAVERNA Digital® da EPUSP foram obtidas as velocidades de recepção do Módulo de Composição, as primeiras medidas apontavam uma

velocidade entre cinco e seis quadros por segundo. Essa velocidade de recepção pode ser vista nas elipses vermelhas (Figura 93) do painel de controle do Módulo de Composição, que está sendo executado no Microcomputador Cliente / Servidor. Cada uma das elipses apresenta a velocidade de transferência de quadros gerados por cada um dos Microcomputadores Servidores conectados ao sistema.

Figura 93: Painel do Módulo de Composição mostrando a velocidade de recepção

Ao se verificar o desempenho dos Microcomputadores Servidores com o programa Gerenciador de Tarefas do Windows, constatou-se que é utilizada apenas uma parte da capacidade total de processamento, como se pode ver na Figura 94.

Figura 94: Painel do Gerenciador de Tarefas do Windows, durante a execução do Módulo de Aquisição

Os computadores montados no robô, para a apresentação no IEEE Virtual Reality Conference 2006, têm menor desempenho quando comparados aos computadores utilizados nos testes na CAVERNA Digital®. No entanto, apresentavam uma velocidade de recepção semelhante à observada nos testes. Ambas as montagens apresentavam velocidade entre 5 e 6 quadros por segundo.

Observando o consumo de CPU, no Gerenciado de Tarefas, percebe-se que os computadores operam com folga. Inferiu-se que não se pode alcançar maior velocidade de exibição porque a rede local não dá vazão suficiente aos quadros enviados pelos Microcomputadores Servidores, retardando o processamento do conjunto. Utilizou-se para os

testes uma rede de 100 Mbits e é provável que com uma rede Gigabit seja possível atingir uma velocidade de transmissão próxima ou igual à desejada.

Também foram detectados problemas de sincronismo entre os dois Módulos de Aquisição, como se vê nas elipses vermelhas da Figura 94. Enquanto um Microcomputador Servidor transmite 5,56 quadros por segundo, o outro transmite 6,21 quadros por segundo. É provável que a causa do problema seja o tráfego na rede entre os Microcomputadores Servidores e o Microcomputador Cliente / Servidor.

O problema pode ser resolvido com o Módulo de Composição enviando um sinal de que está pronto para receber as imagens para os Módulos de Aquisição; estes só enviariam novas imagens após receberem o comando (sinal de sincronismo). Assim se garantiria que os dois conjuntos de imagens foram gerados no mesmo instante. Esta alteração está na lista de prioridades do projeto e deverá ser feita em curto espaço de tempo.

Gravou-se uma série de vídeos, com imagens panorâmicas do átrio da CAVERNA Digital®, os quais provam a viabilidade de se implantar a metodologia de tratamento e costura de imagens apresentada neste trabalho, que possibilita a geração de vídeos panorâmicos de 360° ao vivo.

6.5 Conclusão

Após a implantação dos Módulos de Aquisição e de Composição, e dos testes efetuados, é possível concluir que o projeto é plenamente factível, graças à metodologia proposta, de utilização de um algoritmo de transformação geométrica de imagens baseado no fluxo de deslocamento de pixels, a partir de um Mapa de Posicionamento.

Se não se obteve o desempenho esperado, em função de não se utilizarem equipamentos adequados e das configurações de rede propostas, foi no entanto possível demonstrar na prática a viabilidade do método.

Verificou-se que, utilizando-se cerca de 25% da capacidade computacional de microcomputadores antigos (com velocidade de 1.7 GHz), é possível transmitir imagens a uma velocidade de cinco quadros por segundo. O que está freando a execução do sistema é a rede local, de 100 Mbits.

Com a montagem de um novo aglomerado, com processadores de última geração, e com o uso de uma rede local com barramento de tecnologia GigaBit, deve-se atingir a velocidade planejada de, no mínimo, 15 quadros por segundo.

7 CONCLUSÕES E TRABALHOS FUTUROS

7.1 Introdução

Este trabalho apresentou conceitos, alguns antigos, outros recentes, sobre imagens panorâmicas e suas utilizações, e sugeriu um sistema de telepresença por meio da aquisição, transmissão e projeção em ambientes imersivos de vídeos panorâmicos em 360°.

O presente capítulo traz as considerações finais sobre a pesquisa e o desenvolvimento do protótipo, descreve as principais contribuições do trabalho e apresenta propostas de trabalhos futuros e de evolução do sistema desenvolvido.

7.2 Considerações sobre o trabalho

O projeto, ao se iniciar, era menos ambicioso. Seu objetivo era obter imagens da vida real na CAVERNA Digital®, até mesmo uma imagem gravada previamente, sem muita preocupação quanto a resolução e velocidade.

Com o surgimento do projeto GigaVR, visualizou-se a oportunidade de dar um salto qualitativo e quantitativo. Iniciou-se um projeto mais ousado, que deveria ter necessariamente um sistema de captura de imagens panorâmicas com 360° de cobertura, de alta resolução e ao vivo, a longa distância e com exibição em um ambiente de alta imersibilidade.

Devido à definição de imagem desejada foi necessário construir um equipamento, denominado Anel de Câmeras (para atingir até 3.400 pixels x 640 pixels), e desenvolver um conjunto de programas para efetuar todo o processamento e transmissão de imagens. O primeiro anel de câmeras, desenvolvido pelo Dr. Celso Kurashima, foi posteriormente aperfeiçoado, para simplificar a integração e o processamento.

Como a demanda computacional é elevada para esse tipo de sistema, foi necessária a montagem de um aglomerado de computadores, para se obter um processamento de alto desempenho. Para tanto, utilizaram-se inicialmente quatro e, posteriormente, três computadores biprocessados do LSI-EPUSP. A montagem e a manutenção do aglomerado de computadores foram realizadas pela equipe da CAVERNA Digital®.

7.3 Contribuições

Procurou-se apresentar todos os conceitos conhecidos necessários ao desenvolvimento de um projeto de telepresença utilizando imagens panorâmicas, além de algumas idéias novas de transformações geométricas de imagens, objetivando alto desempenho e buscando fornecer uma contribuição para futuros trabalhos de pesquisadores da área.

A principal contribuição científica do trabalho foi o desenvolvimento de métodos otimizados para a correção de distorções, projeções e costura de imagens para sistemas de telepresença baseados em vídeos panorâmicos em ambientes imersivos. A utilização de tais sistemas atende a inúmeras aplicações, em vários campos científicos, da arqueologia à pesquisa espacial, da prospecção de petróleo ao controle remoto de veículos.

Sua contribuição tecnológica foi o desenvolvimento do protótipo de *software* e a montagem do sistema de *hardware* para captura e costura de imagens panorâmicas em 360°, ao vivo. Devido à grande quantidade de operações matemáticas e de movimentação de registros de memória, necessárias para a correção e a costura de imagens, isso somente foi possível com o desenvolvimento de um método no qual se utilizou uma matriz de ponteiros de pixels, denominada Mapa de Posicionamento.

Durante o desenvolvimento do projeto, foi apresentado em Fortaleza, Ceará, em 2002, o trabalho “*Experiments with Panoramic Imaging in Multiprojection Immersive Environments*” no *5th Brazilian Symposium on Virtual Reality*, onde se demonstrou a implementação com imagens estáticas.

No 4th *International Information and Telecommunication Technologies Symposium*, promovido pelo IEEE Região 9 e pela SBC (Sociedade Brasileira de Computação), que ocorreu em Florianópolis, Santa Catarina, em dezembro de 2005, foi apresentado o trabalho “Geração de Imagens Panorâmicas para Sistemas de Telepresença”, versando sobre a pesquisa e implementação do sistema aqui proposto.

Foi apresentado um protótipo do sistema no IEEE Virtual Reality Conference 2006, na seção Demonstração de Laboratórios em março de 2006 em Alexandria, Virgínia, EUA .

Atualmente, encontra-se em elaboração um artigo descrevendo detalhadamente o projeto e seus resultados.

7.4 Trabalhos futuros

O primeiro trabalho, a ser finalizado em futuro próximo, é o Módulo de Exibição, a ser executado no ambiente da CAVERNA Digital®, que servirá para a exibição de imagens geradas em locais distantes pelos Módulos de Aquisição e de Composição, transmitidas pela Internet 2.

Não foi o escopo deste projeto, mas outros elementos deverão se incorporar ao sistema, de forma a aumentar a sensação de imersão:

- Transmissão de informações sonoras, de forma a permitir aos usuários ouvir ruídos e falas remotos. As informações poderiam ser estereofônicas e utilizar vários canais de áudio;
- Geração e transmissão de imagens estereoscópicas, que ao mostrar os objetos em três dimensões permitam maior imersão (SUN; PELEG, 2004);
- Realidade aumentada (RA). Inclusão de elementos virtuais às imagens do mundo real, de forma a simular situações de risco em ambiente habitual ao usuário. Um exemplo seria, simulando a condução de um carro, utilizar imagens reais das ruas e avenidas, e repentinamente colocar um pedestre virtual em situação de risco de atropelamento.

Dessa forma, poder-se-iam avaliar as reações do motorista e treiná-lo para situações similares no mundo real;

- Instalação de sensores e atuadores com dispositivos hápticos (dispositivos que fornecem ao usuário sensações táteis, como textura de objetos, peso, densidade e temperatura) na CAVERNA Digital® e braços robóticos no local remoto, de forma a permitir que o usuário atue em um ambiente distante.

Outro trabalho que se sugere é o desenvolvimento de rotinas que efetuem o alinhamento de forma automática, desonerando o usuário de executar o Procedimento de Calibração de Imagens.

Também seria interessante o desenvolvimento de rotinas para corrigir as variações de cores entre duas imagens, conhecidas como *blending*, de forma a não deixar o usuário perceber quando termina uma imagem e começa outra (LEVIN et. al., 2005; SU et. al., 2004).

7.5 Consideração Final

Os resultados obtidos neste trabalho confirmam a viabilidade da proposta e apontam como promissor o investimento em aprimoramentos no protótipo atual, incorporando a visualização de imagens de locais distantes, ao vivo, com 360°, visão estereoscópica, informações sonoras estereofônicas e inclusão de elementos virtuais, tudo apresentado em ambiente de alta imersão, como a CAVERNA Digital®, possibilitando experiências de telepresença nas mais diferentes áreas de aplicação.

REFERÊNCIAS

- ACMI – Australian Centre for the Moving Image. Disponível em: http://acmi.net.au/AIC/DU_HAURON_BIO.html. Acesso em 07/05/06.
- AGAPITTO, L.; HAYMAN, E.; REID, I.D. Self-Calibration of Zooming Cameras from a Single Viewpoint. In: BENOSMAN, R.; KANG, S. **Panoramic Vision**. New York: Springer-Verlag, 2001. p.269-289.
- AHMED, N.; NATARAJAN, T.E.; RAO, K.R. On Image Processing and a Discrete Cosine Transform. **IEEE Transactions on Computers**, v. C-23 (1), p. 90-93, 1974.
- ALBUQUERQUE, M.P. **Processamento de Imagens: Métodos e Análises**. Disponível em: http://mesonpi.cat.cbpf.br/e2002/cursos/NotasAula/Apostila_Imagens.pdf. Acesso em 22/12/04.
- ALIAGA, D.; CARLBOM, I. Plenoptic Stitching: A Scalable Method for Reconstructing 3D Interactive Walkthroughs. In: Proceedings of ACM SIGGRAPH 2001, Los Angeles, CA., 2001. **Anais**. ACM, 2001. p.443–450.
- ALMEIDA, R. Q. **Arquitetura TCP/IP**. Disponível em <http://www.dicas-l.unicamp.br/dicas-l/Treinamentos/tcpip/tcpip-10.html>. Acesso em 07/2005.
- BOUGUET, J.Y. **Camera Calibration Toolbox for Matlab**. Disponível em: http://www.vision.caltech.edu/bouguetj/calib_doc/index.html. Acesso em 30/11/2004.
- BROWN, D.C. Decentering Distortion of Lenses. **Photometric Engineering**, v. 32, n. 3, p. 444-462, 1966.
- BURIAN, P.K.; CAPUTO, R. **National Geographic Photography Field Guide**. National Geographic, 1999.
- CABRAL, M. et al. **Panoramic Image Capture and Display**. IEEE Virtual Reality Conference 2006 Laboratory Demonstrations. Proceedings of IEEE VR2006, Alexandria, Virginia, USA, 2006. Disponível em: http://www.vr2006.org/lab_demos.htm.
- CANON. **Lens Works - EF Lens Work II**. Canon Inc, 1998.
- CRUZ-NEIRA, C.; SANDIN, D.J.; AND DEFANTI, T.A. Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the Cave. In: Proceedings of ACM SIGGRAPH, 1993. **Anais**. ACM, 1993. p. 135-142.
- CRUZ-NEIRA, C.; SANDIN, D.J.; AND DEFANTI, T.A. Virtual Reality: The Design and Implementation of the CAVE®. Disponível em: <http://www.evl.uic.edu/core.php?mod=4&type=3&indi=135>. Acesso em 08/05/2006.
- CURTIN, D. P. **Imagem Digital**. Disponível em: <http://www.imagem-digital.com> . Acesso em 15/10/2001.

FARIAS, J.B.; OLIVEIRA, K.K.C. **Redes ATM x Gigabit Ethernet**. Disponível em <http://www.di.ufpe.br/~flash/ais98/atm/gigabit_atm.htm>. Acesso em 21/11/2005.

FIALA, M.; BASU, A. Panoramic Stereo Reconstruction Using Non-Svp Optics. In: International Conference on Pattern Recognition (ICPR 2002). Quebec, Canada, 2002. **Anais**. 2002. v. 4, p. 27-30.

FRAGASSI, A.P.M.; SILVA, A.S. **Calibração de câmaras fotográficas não métricas**. Disponível em: <<http://geodesia.ufsc.br/Geodesia-online/ARQUIVO/Conea7/513.PDF>>. Acesso em 30/11/2004.

FORSYTH, D.; PONCE, J. **Computer Vision: A Modern Approach**. New York: Prentice Hall. 2002.

GIBSON, J.D. et al. **Digital Compression for Multimedia: Principles & Standards**. San Diego: Morgan Kaufmann Publishers. 1998.

GLUCKMAN, J.; NAYAR, S.K.; THORESZ, K.J. Real-Time Omnidirectional and Panoramic Stereo. In: Proceedings of DARPA Image Understanding Workshop. Monterey, California, 1998. **Anais**. 1998, p. 299-303.

GROSS, M. et al. Blue-c - A Spatially Immersive Display and 3D Video Portal for Telepresence. In: Proceedings of ACM SIGGRAPH, 2003. **Anais**. ACM, 2003. p. 819-827.

HEIKKILÄ, J.; SILVÉN, O. A Four-step Camera Calibration Procedure with Implicit Image Correction. In: Conference on Computer Vision and Pattern Recognition (CVPR'97), San Juan, Puerto Rico, 1997. **Anais**. IEEE Computer Society, 1997. p. 1106-1112.

HOU, M. User Experience with Alignment of Real and Virtual Objects in a Stereoscopic Augmented Reality Interface. In: 7th International Conference on Information Visualization. 2003. **Anais**. IEEE Computer Society, 2003.

HU, O.R.T. **Análise Comparativa dos Principais Formatos de Armazenamento de Imagem**. Dissertação de Mestrado. Universidade Presbiteriana Mackenzie. 2000.

HU, O.R.T.; CABRAL, M.; ZUFFO, M.; LOPES, R. Experiments with Panoramic Imaging in Multiprojection Immersive Environments. In: Proceedings of the 5th Brazilian Symposium on Virtual Reality, Fortaleza, CE, 2002. **Anais**. SBC, 2002. p. 262-270.

HU, O.R.T.; KURASHIMA, C.; LOPES, R. D.; ZUFFO, M.K. Geração de Imagens Panorâmicas para Sistemas de Telepresença. In: 4th International Information and Telecommunication Technologies Symposium, Florianópolis, 2005. **Anais**. IEEE Região 9 – SBC, 2005. 1 CD-ROM.

HU, O.R.T.; RAUNHEITTE, L.T.M. Padrão JPEG de Compactação de Imagens. **Revista Mackenzie de Engenharia e Computação**. São Paulo, v. 1, p. 139-152, 2000.

HU, O.R.T.; RAUNHEITTE, L.T.M. **Processamento e Compressão Digital de Imagens**. São Paulo: Editora Mackenzie, 2004.

ISO/IEC JTC1/SC29/WG11. **Coding of Moving Pictures and Audio**. Disponível em: <<http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm>> Acesso em 25/11/2004.

JACOBSON, J. Configuring Multiscreen Displays with Existing Computer Equipment. In: Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting, 2002. **Anais**. 2002. p.761-765

KANG, S.B.; SZELISKI, R. **3-D Scene Data Recovery using Omnidirectional Multibaseline Stereo**. Digital Equipment Corporation, Cambridge Research Lab, 1995.

KANG, S.B.; WEISS, R. Characterization of error in compositing panoramic images. In: BENOSMAN, R., KANG, S. **Panoramic Vision**. New York: Springer-Verlag, 2001. p.205-226.

KODAK. **Curso básico de fotografia para amadores**. CEK – Centro Educacional Kodak. 2004.

KURASHIMA, C.S. **Proposta de um Sistema de Teleconferência Imersiva Panorâmica**. Tese de doutorado. Escola Politécnica da USP. 2005.

LEVIN, A.; ZOMET, A.; PELEG, S.; WEISS, Y. **Seamless Image Stitching in the Gradient Domain**. Hebrew University. Disponível em <<http://leibniz.cs.huji.ac.il/tr/acc/2003/huji-cse-ltr2003-82blending.pdf>>. Acesso em 09/05/2005.

LI, Z.N.; ZHONG, W. **Multimedia Systems**. Simon Fraser University, Canada. Disponível em: <http://www.cs.sfu.ca/CourseCentral/365/li/index_prev.html>. Acesso em 26/03/00.

LIPTON, L. **Foundations of the Stereoscopic Cinema**. New York: Van Nostrand Reinhold Company Inc. 1982.

LOMBARD, M.; DITTON, T. **At the Heart of it All: The Concept of Presence**. JCMC, v. 2 Setembro de 1997.

LOPES, J.A.S. **Gigabit Ethernet**. Disponível em <http://www.gta.ufrj.br/grad/98_2/joao>. Acesso em 21/11/2005.

MASE, K.; NISHIRA, H. Computing the Field-of-View of a Stitched Panorama to Create FoV Sensitive Virtual Environments. In: Proceedings of ICPR '96. **Anais**. IEEE, p. 151-155, 1996.

MELO, H. Compressão digital de vídeo. **Revista de Engenharia de Televisão**. 1999.

MENDONÇA, A.; ZELENOSKY, R. **Barramento Firewire (IEEE 1394)**. Disponível em: <<http://www.clubedohardware.com.br/firewire.html>>. Acesso em 19/12/2004.

NAYAR, S. K. Catadioptric Omnidirectional Camera. Conference on Computer Vision and Pattern Recognition (CVPR'97), San Juan, Puerto Rico, 1997. **Anais**. IEEE Computer Society, 1997.

NAYAR, S.K. Omnidirectional Vision. In: Proceedings of International Symposium on Robotics Research, Japan, 1997 **Anais**. IEEE, 1997.

NAYAR, S.K.; KARMARKAR, A.D. 360 x 360 Mosaic: Regular and Stereoscopic. In: BENOSMAN, R., KANG, S. **Panoramic Vision**. New York: Springer-Verlag, 2001. p.291-307.

NEUMANN, U.; PINTARIC, T.; RIZZO, A. Immersive Panoramic Video. In: Proceedings of the 8th ACM International Conference on Multimedia. Los Angeles, CA, 2000. **Anais**. ACM, 2000. p.493-494.

NG, K-T.; CHAN, S-C.; SHUM, H-Y. **Data Compression and Transmission Aspects of Panoramic Videos**. IEEE Transactions On Circuits And Systems For Video Technology, v. 15, n. 1, p. 82-95, January 2005

NRV - NÚCLEO DE REALIDADE VIRTUAL - LSI - USP, Disponível em: <<http://www.lsi.usp.br/~rv>>. Acesso em 11/10/2001.

OPENCV. versão beta 4, Biblioteca Gráfica. Intel Corp. Disponível em: <<http://sourceforge.net/projects/opencvlibrary/>>. Acesso em agosto de 2004.

PANEDIT. Versão 1.0.9, IBM. Disponível em <<http://www.ibm.com/hotmedia>>. Acesso em Julho de 2002.

PELEG, S.; HERMAN, J. Panoramic Mosaics by Manifold Projection. In: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97). **Anais**. IEEE, p. 338-343, 1997.

PELEG, S., et al. Mosaicing with Strips on Adaptive Manifolds. In. BENOSMAN, R., KANG, S. **Panoramic Vision**. New York: Springer-Verlag, 2001. p.291-307.

PELEG, S.; BEN-EZRA, M. Stereo Panorama with a Single Camera In: Proceedings of the 1999 Conference on Computer Vision and Pattern Recognition (CVPR '99). **Anais**. IEEE, p. 1395-1401, 1999.

PELEG, S.; BEN-EZRA, M.; PRITCH, Y. **Omnistereo: Panoramic Stereo Imaging**. IEEE Transactions on Pattern Analysis and Machine Intelligence, v.23, n.3, p 279-290, 2001.

PEREIRA, F. **Tele-Visões: Novas Tecnologias, Novos Serviços**. Instituto Superior Técnico - Instituto de Telecomunicações. Lisboa, Portugal. 1999.

PINTARIC, T.; RIZZO, A.; NEUMANN, U. Video-based Virtual Environments. In: Proceedings of ACM SIGGRAPH 2003, Conference on Sketches and Applications, San Diego, CA., 2003. **Anais**. ACM, 2003.

POINT GREY RESEARCH INC. **DRAGONFLY: Technical Reference Manual**. Point Grey Research Inc. (PGR). Setembro de 2004.

RIGG, J. **Panoguide**. Disponível em: <<http://www.panoguide.com>>. Acesso em 11/10/2001.

RNP. Projeto Giga, 2004a. - Disponível em: <<http://www.projetogiga.com.br/>> Acesso em julho de 2004 .

RNP. Projeto Giga, 2004b - Chamada RNP 02/2003: **Pesquisa e Desenvolvimento em Protocolos e Serviços de Rede e em Serviços e Aplicações Científicas**, 2004b. Disponível em: <http://www.rnp.br/editais/giga_200302.html>. Acesso em julho 2004 .

RNP. **Mapa do Backbone RNP2**. Disponível em <<http://www.rnp.br/backbone/index.php>>. Acesso em 22/11/2005.

RODRIGUEZ, A.; FIBUSH, D.K.; BILOW, S.C. **MPEG-2 Fundamentals for Broadcast and Post-production Engineers**. Tektronix. 1996. Disponível em: <http://www.tek.com/VND/Support/White_papers/Profile/mpeg2.html>. Acesso em 17/10/99.

SAYOOD, K. **Introduction to Data Compression**. Morgan Kaufmann Publishers. San Diego. 2000.

SEI – Sistemas Estratégicos de Informações – Casa Civil do Governo do Estado de São Paulo. Disponível em: <<http://www.sei.sp.gov.br/portalsei/portal.nsf/V03.02/Internet2?OpenDocument>>. Acesso em 22/11/05.

SHIMAMURA, J.; YOKOYA, N.; TAKEMURA, H.; YAMAZAWA K. Construction of an Immersive Mixed Environment Using an Omnidirectional Stereo Image Sensor. In: IEEE Workshop on Omnidirectional Vision. **Anais**. p. 62–69, 2000.

SHUM, H-Y.; SZELISKI, R. Construction of Panoramic Image Mosaic. In. BENOSMAN, R., KANG, S. **Panoramic Vision**. New York: Springer-Verlag, 2001. p.227-268.

SHUM, H-Y.; NG, K-T.; CHAN, S-C. **A Virtual Reality System Using the Concentric Mosaic: Construction, Rendering and Data Compression**. IEEE Transactions on Multimedia, v.7, n.1, p. 85-95, 2005.

SILVA, L.C. **Método Robusto para a Calibração de Câmeras em Estereofotogrametria**. Tese de doutorado. Programas de pós-graduação de Engenharia da Universidade Federal do Rio de Janeiro. 2003.

STAADT, O.; GROSS, M.; KUNZ, A.; MEIER, M. The Blue-C.: Integrating Real Humans into a Networked Immersive Environment. In: Proceedings of ACM Collaborative Virtual Environments 2000. **Anais**. ACM, 2000 p. 202-202.

STALLINGS, W. **Arquitetura e Organização de Computadores**. Pearson-Prentice Hall. São Paulo. 2002.

SU, M.S.; HWANG, W.L.; CHENG, K.Y. **Analysis on Multiresolution Mosaic Images**. IEEE Transactions on Image Processing, v. 13, n.7, p. 952-959, 2004.

SUN, C.; PELEG, S. **Fast Panoramic Stereo Matching Using Cylindrical Maximum Surfaces**. IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, v. 34, n. 1, p. 760-765, February 2004.

SUN, X.; DUBOIS, E. A novel algorithm to stitch multiple views in image mosaics. In: IEEE Int. Conf. Acoustics Speech Signal Processing - 2004. **Anais**. IEEE, p.481-484, 2004.

SVOBODA, T. ; PAJDLA, P. Panoramic Cameras for 3D Computation. In: Proceedings of the Czech Pattern Recognition Workshop, Prague, Czech Republic, 2000. **Anais**. Czech Society for Pattern Recognition. 2000. p. 63-70.

SZELISKI, R.; SHUM, H.-Y. Creating Full View Panoramic Image Mosaics and Environment Maps. In: Proceedings of 24th International Conference on Computer Graphics and Interactive Techniques. **Anais**. ACM, P.251-258, 1997.

TAN, K-H.; HUA, H.; AHUJA, N. Multiview Panoramic Cameras Using a Mirror Pyramid. In: Proceedings of the IEEE Workshop on Omnidirectional Vision 2002 held in conjunction with ECCV 2002, Copenhagen, Denmark, 2002. **Anais**. IEEE, 2002.

TANAKA, K.; HAYASHI, J.; INAMI, M.; TACHI, S. TWISTER: An Immersive Autostereoscopic Display. In: Proceedings of IEEE VR2004, Chicago, 2004. **Anais**. IEEE, 2004.

TANENBAUM, A. S. **Redes de Computadores**. Rio de Janeiro, Editora Campos, 1997.

USB-IF. Requirements. Disponível em <<http://www.usb.org>>. Acesso em 15/04/2006.

WAACK, F. G. Stereo Photography, An Introduction to Stereo Photo Technology and Practical Suggestions for Stereo Photography. Disponível em: <<http://www.stereoscopy.com/faq/>>. Acesso em 15/10/2001.

WALKER, G. R.; SHEPPARD, P. J. Telepresence - the Future of Telephony. **BT Technology Journal**, v. 15, n. 4, Outubro de 1997.

ZHU, Z. Omnidirectional Stereo Vision. Workshop on Omnidirectional Vision. In. 10th IEEE ICAR, Budapest, Hungary, 2001. **Anais**. IEEE, 2001.

ZUFFO, J. A.; SOARES, L. P.; ZUFFO, M. K.; LOPES, R. D. CAVERNA Digital - Sistema de Multiprojeção Estereoscópico Baseado em Aglomerados de PC's para Aplicações Imersivas em Realidade Virtual. In: Proceedings of the 4th Brazilian Symposium on Virtual Reality - SVR'01. Florianopolis, SC, 2001. **Anais**: SBC, 2001. p.139-147.

ZUFFO, M. K. A. **Convergência da Realidade Virtual e Internet Avançada em Novos Paradigmas de TV Digital Interativa**. São Paulo, SP, 2001. Tese (livre docência) - Escola Politécnica da Universidade de São Paulo, Departamento de Engenharia de Sistemas Eletrônicos.