

# UM PROTÓTIPO DE REALIDADE AUMENTADA USANDO A FACE COMO MARCADOR

**Antonio Carlos dos Santos SOUZA (1), Luiz Claudio Machado dos SANTOS (2), Luiz Henrique Nascimento dos ANJOS (3), Pedro Oliveira RAIMUNDO (4), Aline de Oliveira MACHADO (5)**

- (1) Instituto Federal de Educação, Ciência e Tecnologia da Bahia, Rua Emídio dos Santos, s/n - Barbalho - Salvador - Bahia. CEP: 40301-015, e-mail: [antoniocarlos@ifba.edu.br](mailto:antoniocarlos@ifba.edu.br)  
(2) Instituto Federal de Educação, Ciência e Tecnologia da Bahia, Rua Emídio dos Santos, s/n - Barbalho - Salvador - Bahia. CEP: 40301-015, e-mail: [luizcms@ifba.edu.br](mailto:luizcms@ifba.edu.br)  
(3) Instituto Federal de Educação, Ciência e Tecnologia da Bahia, Rua Emídio dos Santos, s/n - Barbalho - Salvador - Bahia. CEP: 40301-015, e-mail: [lhenriqueanhos@gmail.com](mailto:lhenriqueanhos@gmail.com)  
(4) Instituto Federal de Educação, Ciência e Tecnologia da Bahia, Rua Emídio dos Santos, s/n - Barbalho - Salvador - Bahia. CEP: 40301-015, e-mail: [pedrooraimundo@gmail.com](mailto:pedrooraimundo@gmail.com)  
(5) Instituto Federal de Educação, Ciência e Tecnologia da Bahia, Rua Emídio dos Santos, s/n - Barbalho - Salvador - Bahia. CEP: 40301-015, e-mail: [alinemach@gmail.com](mailto:alinemach@gmail.com)

## RESUMO

A Realidade Aumentada proporciona ao usuário combinar elementos virtuais com o ambiente real, de forma interativa, com processamento em tempo real e em três dimensões. Isto é obtido, através de técnicas de visão computacional e de computação gráfica, resultando na sobreposição de objetos virtuais com o mundo real. As técnicas de Realidade Aumentada vêm despertando grande interesse na área da bioengenharia craniofacial, pois amplia as possibilidades de estudo e prática de variadas técnicas e procedimentos médicos que relacionam as imagens geradas na medicina com a situação instantânea do paciente. Este artigo descreve a implementação da primeira fase de um protótipo de realidade aumentada voltado para predição de cirurgias craniofaciais que ao invés de utilizar os tradicionais marcadores fiduciais de ferramentas livres como o ARTOOLKIT usa a detecção da face através das técnicas PCA e LDA como marcador.

**Palavras-chave:** realidade aumentada, detecção de faces, visão computacional

## 1 INTRODUÇÃO

No Brasil, é considerável o número de pacientes com traumas craniofaciais decorrentes de patologias oriundas de tumores ou acidentes. Felizmente, em muitos casos não há comprometimento neurológico visível. No entanto, o aspecto estético atinge a qualidade de vida do traumatizado craniofacial, pois este está sujeito a preconceitos e rejeição social, principalmente aqueles que possuem baixo poder aquisitivo. Em caso como afundamento do crânio ou face, ou tumores cancerígenos, há a necessidade de reabilitação urgente para evitar comprometimentos neurológicos [Campar 2008]. Portanto, há um apelo social bastante forte para que ferramentas computacionais possam auxiliar não só os diagnósticos, mas também o tratamento nesse tipo de trauma. Tais ferramentas devem atuar desde a identificação da região do trauma, passando pela modelagem de próteses, até a predição e o acompanhamento estético dos procedimentos pós-operatórios.

Como a reconstrução mesmo de uma simples fratura mandibular assume uma grande importância, um planejamento pré-cirúrgico deve ser realizado. Em um caso como esse, os cirurgiões plásticos na sala de cirurgia restauram a forma e a função dos elementos dos ossos fraturados no esqueleto craniofacial tipicamente expondo todos os fragmentos, no maxilar, com implantação de próteses ou enxertos ósseos retirados de outras partes do corpo. Em seguida, deslocam os componentes para uma configuração anatomicamente preferível e, finalmente, instalam-se parafusos rígidos nos ossos e laminas de metal para fixar os segmentos do osso em uma nova posição [Campar 2008]. Essas cirurgias, dependendo do caso, podem ser complexas e de tempo relativamente longo. Essas próteses ou os enxertos requerem um processo de adequação necessariamente particular para cada paciente, devido à anatomia do crânio ser única e a necessidade de ajustes finais dos modelos aos encaixes que geralmente exigem precisão.

As interfaces tradicionais utilizadas para comunicação com esses sistemas computacionais exigem do profissional o desenvolvimento do conhecimento simbólico (abstrato) e necessidade de treinamento. Apesar dos benefícios das novas tecnologias, a sofisticação das interfaces faz com que os operadores tenham que se adaptar aos equipamentos. Com o objetivo de reverter esse processo, diversas pesquisas buscaram maneiras de fazer com que as máquinas se ajustassem às pessoas. Surgiram assim interfaces de voz, interfaces tangíveis, interfaces hápticas [Machado 2007], entre outras, possibilitando ao usuário o acesso a aplicações como se estivessem atuando no mundo real, emitindo sons, fazendo gestos, sentindo resistência nos movimentos, etc [Kirner and Siscouto 2007].

A motivação para o desenvolvimento de aplicações e pesquisas na área de Realidade Aumentada vem do potencial de recursos e benefícios que tais técnicas podem trazer em várias áreas. No campo de registro e composição, as técnicas de Realidade Aumentada vêm despertando grande interesse na área médica, pois amplia as possibilidades de estudo e prática de variadas técnicas e procedimentos médicos que relacionam as imagens geradas na medicina com a situação instantânea do paciente. Isso acontece como uma evolução normal, pois diversos procedimentos médicos são guiados por imagens [Siemens 2008], que fornecem ao cirurgião uma visão anatômica interna do paciente para o planejamento da cirurgia.

Um ponto a ser destacado é a forma como o médico poderá controlar a observação em conjunto exame/paciente. A junção das imagens produzidas em geral por visualização volumétrica com as imagens da câmera é uma questão interessante, visto que essas últimas são geradas a partir de projeções em perspectiva, enquanto que as primeiras utilizam, em geral, a projeção ortográfica.

Ainda, no campo da realidade aumentada, um segundo desafio é como acompanhar os movimentos da face do paciente, considerando que o registro do paciente na realidade aumentada não é um problema trivial. As soluções propostas são limitadas a: manter o paciente totalmente imóvel, utilizar marcadores presos ao paciente [Sielhorst 2007] ou ainda componentes de hardware especializados como câmeras infravermelhas de rastreamento [Fischer 2006]. King et. al. [2000] apresentam a integração de marcadores fixados ao corpo em pontos definidos. Em [Zhao 2007], foi apresentado um método de rastreamento 3D da cabeça direcionado para movimentos “largos” inclusive no eixo Z, usando algoritmos de registro baseados em *Scale Invariant Feature Transform* (SIFT). Primeiro os pontos de saliência dos dois quadros são detectados, obtendo-se os pontos 3D usando a correspondência entre os *pixels* na intensidade da imagem e valores de profundidade dos dois quadros e, finalmente, é estimada a posição da cabeça usando o algoritmo *Random Sample Consensus* (RANSAC) [Fischler and Bolles 1981]. Este algoritmo permite que sejam definidos limiares para limitar o erro de re-projeção cometido nas estimativas de pontos e câmeras.

Dentro desse panorama, o presente projeto de iniciação científica tem como foco o desenvolvimento de um protótipo que nessa fase reconhece a face do paciente usando técnicas de Visão Computacional, na etapa de captura e identificação dos elementos da imagem/vídeo, e técnicas de Realidade Aumentada no registro e na composição do vídeo com um rosto sintético. Esse projeto é parte de um projeto maior de doutoramento que tem como foco a criação de um modelo computacional para visualização de dados médicos, como exames ou próteses, compostos com a imagem em tempo real do paciente, voltado para a área de bioengenharia craniofacial. Tal modelo deve ser capaz de: a) identificar a face do paciente em um fluxo de vídeo; b) compor, em tempo real, os dados médicos a serem analisados a esse fluxo de vídeo; e c) acompanhar os movimentos do paciente mantendo consistente a composição do fluxo de vídeo e dos dados médicos. Para isso, técnicas de Visão Computacional deverão ser aplicadas na etapa de captura (a) e técnicas de Realidade Aumentada nas etapas de registro e composição dos dados (b) e (c).

## 2 FUNDAMENTAÇÃO TEÓRICA

As técnicas de realidade aumentada e de realidade virtual foram desenvolvidas a partir da convergência de uma série de fatores incluindo pesquisas voltadas para área, disponibilidade de produtos e os custos mais baixos. Sendo assim, a linguagem de modelagem VRML – *Virtual Reality Modeling Language* para aplicação em realidade virtual e a biblioteca ARTOOLKIT para realidade aumentada são de fundamental importância para criar e dar continuidade a um bom trabalho que envolva RA e RV. Entretanto, atualmente, há inúmeros dispositivos e produtos disponibilizados por pesquisadores e por empresas como:

- ARTOOLKIT PLUS [http://studierstube.icg.tugraz.ac.at/handheld\\_ar/artoolkitplus.php](http://studierstube.icg.tugraz.ac.at/handheld_ar/artoolkitplus.php)
- ARTAG – <http://www.artag.net>
- DART – [www.gvu.gatech.edu/dart](http://www.gvu.gatech.edu/dart)

- OSGART- [www.osgart.org](http://www.osgart.org)

O sucesso de uma aplicação de realidade aumentada está diretamente relacionado com o uso correto dos diversos recursos e métodos na entrada, no processamento e na saída de imagens para o usuário. A entrada da aplicação feita por rastreadores (equipamentos muito utilizados em aplicações médicas) fornece um bom posicionamento em tempo real e orientações de movimento de objetos ou pessoas, explorando diferentes tecnologias.

Os rastreadores são classificados em mecânico, ultrassônicos, luminosos e óticos [Nunes et. Al 2007]. Quando acoplados a instrumentos de manipulação cirúrgica esses rastreadores permitem ao usuário, precisão no controle dos seus movimentos e ao mesmo tempo geram resistência fazendo com que as aplicações se aproximem ao máximo do mundo real. Um exemplo de rastreador são as luvas que capturam os movimentos das mãos e dos dedos, tal dispositivo faz com que as respostas do ambiente sejam compatíveis com movimentos da extremidades dos membros anteriores humanos [Machado 2007].

O objetivos dessas estruturas é permitir que os sistemas realizem operações similares ao utilizados pelo homem como reconhecimento de tamanho relativo, altura, perspectiva, oclusões para estimar o tamanho, orientação e posicionamento relativo de objetos visíveis, usando, para tanto, câmeras de vídeo associadas a métodos de visão computacional.

O registro ou o alinhamento dos objetos virtuais com a cena real é realizado com auxílio de rastreadores ou *trackers*. Mas, o problema aqui é saber em qual posição e com qual orientação os objetos virtuais deverão ser colocados em relação à cena real. Para identificar este posicionamento, é necessário relacionar as coordenadas dos objetos virtuais com as da cena real [Silva 2006]. Portanto, também se torna necessário saber qual é a posição e orientação do usuário.

O registro é um problema crucial em sistemas de Realidade Aumentada, uma vez que objetos mal posicionados passam a impressão de estar “flutuando” ou, em alguns casos, sumirem atrás de outros objetos. O registro acurado se torna ainda mais fundamental em aplicações que demandam alta precisão, como por exemplo, para guiar exames e cirurgias [Silva 2006]. Para obter esta informação, usam-se os equipamentos rastreadores.

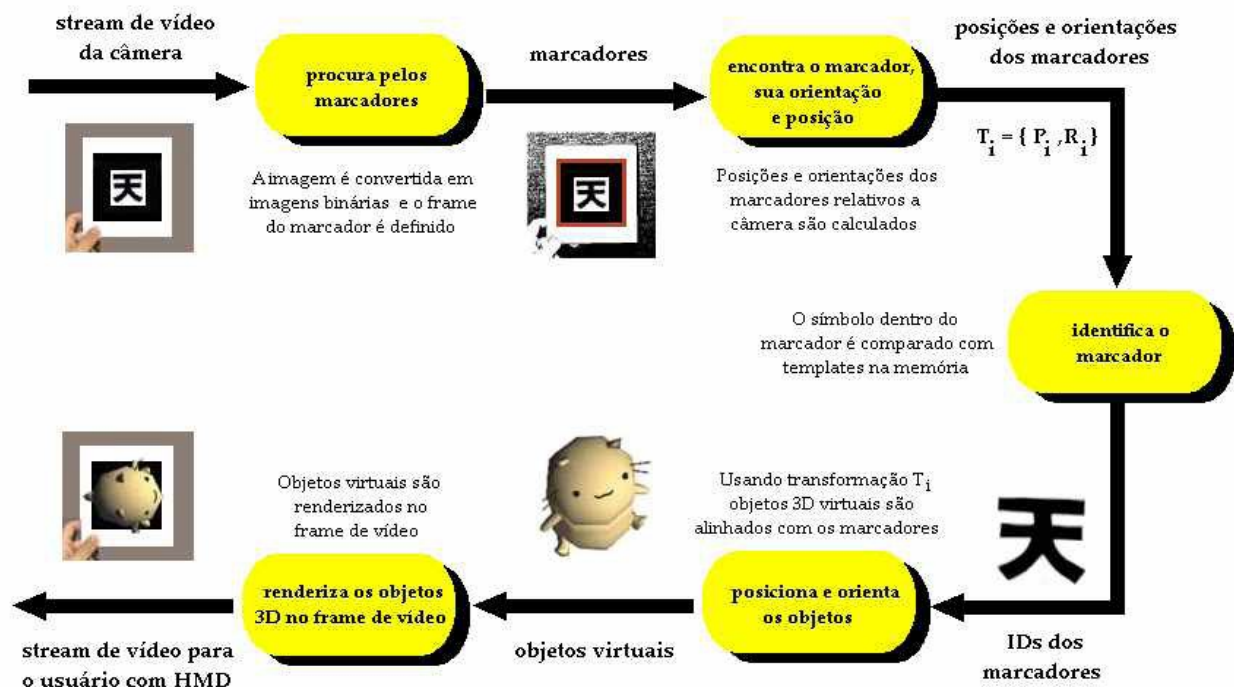
Características sobre rastreadores foram discutidas em [Navab, Bani-Hashemi, and Mitschke 1999], que foram resumidas por Silva [Silva 2006] em precisão, resolução, retardo, taxa de atualização, infra-estrutura, faixa de operação, interferência e preço.

Por sua vez, a saída pode ser feita, por exemplo, usando HMD (*Head-mounted display*), uma espécie de capacete com câmera que gera uma visão estereoscópica (também conhecido como capacete), mesas de projeção, que integram duas telas verticais em formato de “L”; monitor comum; e a caverna ou Cave, que é uma pequena sala onde as imagens são projetadas sobre paredes translúcidas. Os HMD são utilizados em variados tipos de aplicações, desde tratamento de fobias e problemas neuropsiquiátricos até jogos colaborativos. A CAVE possui sensores de posição que capturam os movimentos dos usuários e atualizam as imagens projetadas.

## 2.1 Pipeline tradicional do ARTOOLKIT

As técnicas de Realidade Aumentada normalmente são utilizadas em conjunto com as técnicas de visão computacional, pois estas fazem a captura e reconhecimento dos marcadores na imagem ou vídeo obtido por meio de câmeras.

O processo de reconhecimento dos marcadores realizado atualmente pelo ARToolkit acontece num *loop*, onde é feita a varredura de cada quadro capturado pela *webcam* em busca de figuras em forma de quadrado. Encontrando algum quadrado, o seu conteúdo interno é analisado e comparado a um padrão carregado na memória. Com esses procedimentos, o ARToolkit é capaz de responder às ações do usuários de acordo com o marcador detectado.



**Figura 1: PIPELINE do ARTOOLKIT**

Em suma, três componentes básicos são necessários para a existência da Realidade Aumentada:

- Objeto real com algum tipo de marca de referência, que possibilite a interpretação e criação do objeto virtual;
- Câmera ou dispositivo capaz de transmitir a imagem do objeto real;
- Software capaz de interpretar o sinal transmitido pela câmera ou dispositivo.

O passo a passo para visualização do modelo virtual no ARTOOLKIT, conforme a figura 1 é o seguinte:

- Coloca-se o objeto real em frente à câmera, para que ela capte a imagem e transmita ao equipamento que fará a interpretação. Algumas câmeras precisam ser calibradas, pois a lente pode distorcer a imagem, mas o ARToolkit possui um software responsável por essa tarefa.
- A câmera “enxerga” (visão computacional) o objeto e manda as imagens, em tempo real, para o software que gerará o objeto virtual.
- O software já estará programado para retornar determinado objeto virtual, dependendo do objeto real que for mostrado à câmera. Esse objeto também pode ser animado e combinado com sons que são emitidos quando ele é renderizado, para obter um realismo maior por parte do objeto virtual.
- O dispositivo de saída (que pode ser uma televisão ou monitor de computador) exibe o objeto virtual em sobreposição ao real, como se ambos fossem uma coisa só. Outros dispositivos podem ser utilizados, tais como óculos especiais que permitem visão estereoscópica do objeto, fazendo com que ele pareça mais real.

No protótipo desenvolvido, em termos de hardware, é necessária apenas uma câmera para capturar a imagem em tempo real e o monitor do próprio computador para mostrar o resultado da junção da imagem real com o rosto sintético. Em termos de usabilidade, ao invés de utilizar um marcador fiducial, o software realiza as técnicas de visão computacional voltadas para detecção de face, que será vista na próxima seção.

## 2.2 Técnicas de Visão Computacional para Detecção de Face

Como normalmente as ferramentas como ARTOOLKIT utilizam marcadores fiduciais para reconhecimento, o protótipo desenvolvido nesse projeto de iniciação científica teve como foco possibilitar a composição de um rosto real capturado por uma câmera de vídeo com um rosto sintético. Para tanto, ao invés de utilizar marcadores fiduciais comuns, foram usados algoritmos de Visão Computacional como PCA e LDA para detectar o rosto, na etapa de captura e identificação dos elementos da imagem/vídeo, e técnicas de Realidade Aumentada no registro e na composição do vídeo com um rosto sintético.

Em geral, o primeiro passo é a localização de regiões com cores semelhantes a cores de pele, definidas nos classificadores e as comparações dessas regiões com identificadores predefinidos para realizar a identificação e localização da face do paciente.

O algoritmo PCA é um método estatístico para reduzir a dimensão do conjunto de dados e enquanto mantém a maioria das variações presentes no conjunto de dados. O PCA é considerado referência e *benchmarks* em comparação com outros algoritmos pela facilidade de implementação e razoável desempenho.

Esse algoritmo possui três importantes propriedades: 1) é possível comprimir um conjunto de vetores de dimensão elevada em um conjunto de vetores de menor dimensão e em seguida reconstruí-lo; 2) o modelo de parâmetros pode ser calculado diretamente a partir dos dados, “diagonalizando” a covariância da amostra; 3) realiza operações de compressão e descompressão. É uma técnica clássica que pode fazer aplicações com modelos lineares, como processamento de sinais e imagens, sistema de teoria de controle, entre outros. Já o LDA tem como objetivo classificar objetos (pessoas, animais, clientes, face/não-face) em um de dois ou mais grupos com base em um conjunto de características que descrevem esses objetos (como exemplo, sexo, idade, entre outros). Os objetos podem ser atribuídos a uma série de grupos pré-determinados com base em observações feitas sobre um determinado objeto, no quais esses grupos são conhecidos ou pré-determinados e não são disponibilizados em ordem (ou seja, escala nominal). Esse algoritmo utiliza o critério de classificação para minimizar o erro total (TEC), que tende a mudar a proporção do objeto. O TEC é a regra de desempenho a “longo prazo”, numa amostra aleatória de objetos [Yuriy 2007] e [De La Torre et. Al, 2001].

No próximo tópico será mostrada a implementação do protótipo e como é realizado o processo de detecção da face. Para tanto, o algoritmo utiliza os chamados classificadores para obter, através de um processo de “treinamento” do software, as imagens para rastrear traços de rostos, para diminuir a possibilidade de erro e que alguma imagem seja indevidamente detectada como face.

### 3 O PROTÓTIPO

O protótipo desenvolvido nesse projeto de iniciação científica teve como foco possibilitar a composição de um rosto real capturado por uma câmera de vídeo com um rosto sintético. Para tanto, ao invés de utilizar marcadores fiduciais comuns, foram usados classificadores, utilizados pelos algoritmos de Visão Computacional PCA e LDA para detectar o rosto, na etapa de captura e identificação dos elementos da imagem/vídeo, e técnicas de Realidade Aumentada no registro e na composição do vídeo com um rosto sintético.

Basicamente o processo de detecção de face no modelo implementado é o seguinte:

- a) O paciente se posiciona em frente à câmera, para que ela capte a imagem e transmita ao equipamento que fará a interpretação, conforme figura 2.
- b) A câmera “enxerga” (visão computacional) o paciente e manda as imagens, em tempo real, para o software que gerará o objeto virtual. Nesse momento, os algoritmos varrem a imagem em busca de padrões semelhantes aos dos classificadores cadastrados.
- c) O software já estará programado para posicionar uma figura de um rosto sintético em alinhamento com a face detectada, conforme figura 3.

Para desenvolver o protótipo, foram usados os SDKs do Windows e do DirectX. Este segundo provê classes para captura do vídeo da *webcam*, comparação entre a imagem e os classificadores, bem como a exibição da imagem virtual.



**Figura 2. Composição da imagem real com o rosto sintético**

#### 4 CONSIDERAÇÕES FINAIS

Da forma como as ferramentas apresentadas de realidade aumentada trabalham atualmente (exemplo do AR-TOOLKIT), a composição em tempo real do rosto real do paciente com o modelo sintético seria possível a um custo baixo apenas com a colagem de marcadores fiduciais no rosto do paciente. Como o objetivo é melhorar a usabilidade e automatizar o processo, um protótipo foi desenvolvido usando a detecção da face como um marcador e compondo a imagem real com o rosto sintético usando algoritmos como PCA e LDA.

Assim, o protótipo apresentado só faz a detecção do rosto e composição com uma imagem de uma máscara. O próximo desafio é descobrir a orientação e a posição do rosto do paciente em relação à câmera para o alinhamento e inclinação correta e proporcional ao modelo sintético. Em paralelo a isso, uma importante contribuição é modificar o código fonte do ARTOOLKIT para reconhecimento da face como um marcador.

Além de aperfeiçoar o programa para acompanhar a orientação da face detectada, serão testados melhoramentos do método usando PCA (Análise do Componente Principal), ICA (Análise do Componente Independente) e LDA (Análise Discriminante Linear).

O protótipo será utilizado principalmente para avaliar o modelo computacional para visualização de dados médicos, como exames ou próteses, compostos com a imagem em tempo real do paciente, voltado para a área de bioengenharia craniofacial que : a) identificará a face do paciente em um fluxo de vídeo; b) comporá, em tempo real, os dados médicos a serem analisados a esse fluxo de vídeo; e c) acompanhará os movimentos do paciente mantendo consistente a composição do fluxo de vídeo e dos dados médicos.

#### REFERÊNCIAS

- Azuma, R. (2001). Baillot, Y. Behringer, R. Feiner, S. Julier, S. & MacIntyre, B. *Recent Advances in Augmented Reality*. Computer Graphics and Applications, v. 21, n. 6, pp. 34–47.
- Campar. (2008). Research in Medical Augmented Realit., TUM. <http://campar.in.tum.de/Chair/ResearchIssueMedAR>.
- De la Torre, Fernando and Blacky, Michael J. (2001). Robust Principal Component Analysis for Computer Vision. Int. Conf. on Computer Vision (ICCV'2001), Vancouver, Canada.

- Fischer, Jan. (2006). Rendering Methods for Augmented Reality, Dissertation, Universität Tübingen.
- Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.
- Kirner, C. and Siscouto, R. (2007). Fundamentos de Realidade Virtual e Aumentada. In: Kirner, C.; Siscouto, R.. (Org.). *Realidade Virtual e Aumentada: Conceitos, Projeto e Aplicações*. 1 ed. Porto Alegre - RS: Sociedade Brasileira de Computação – SBC, v. 1, p. 2-21.
- Machado, Liliane dos Santos. Dispositivos Hápticos para Interface de Realidade Virtual e Aumentada. In KIRNER, C. (Org.); SISCOUTO, R. (Org.). *Realidade Virtual e Aumentada: Conceitos, Projeto e Aplicações*. 1. ed. Porto Alegre - RS: Sociedade Brasileira de Computação - SBC, 2007. pp. 152-167.
- Milgram e Kishino. (1994). A Taxonomy of Mixed Reality Visual Displays. [http://vered.rose.utoronto.ca/people/paul\\_dir/IEICE94/ieice.html](http://vered.rose.utoronto.ca/people/paul_dir/IEICE94/ieice.html), (Acessado em 19/01/2010).
- Navab, N. Bani-Hashemi, A. and Mitschke, M. (1999). Merging Visible and Invisible: Two Camera-Augmented Mobile C-arm (CAMC) Applications. In *Proc. of IEEE and ACM International Workshop on Augmented Reality (IWAR)*, pages 134–141.
- Nunes, Fátima. Costa, Rosa, et. al. (2007). Aplicações Médicas usando Realidade Virtual e Realidade Aumentada. In KIRNER, C. (Org.); SISCOUTO, R. (Org.). *Realidade Virtual e Aumentada: Conceitos, Projeto e Aplicações*. 1. ed. Porto Alegre - RS: Sociedade Brasileira de Computação - SBC. pp. 223-255.
- Sielhorst, Tobias. (2007). New Methods for Medical Augmented Reality. Dissertação. Technische Univesistat Munchen, Fakultat fur Informatik.
- Silva, Rodrigo Luis de Souza. (2006). Um Modelo de Redes Bayesianas Aplicado a Sistemas de Realidade Aumentada. Phd thesis sob orientação de Luiz Landau e Gilson Giraldi, COPPE/Universidade Federal do Rio de Janeiro. Rio de Janeiro.
- Siemens. <http://www.siemens.com.br/templates/columa1.aspx?channel=2129> acesso em 23 de agosto de 2008.
- Vallino, James R. (1998). Interactive Augmented Reality. PhD Dissertation, Departament of Computer Science, University of Rochester, Rochester, NY, 1998. [http://www.ai.mit.edu/projects/medical-vision/Grupo de visão médica](http://www.ai.mit.edu/projects/medical-vision/Grupo%20de%20vis%C3%A3o%20m%C3%A9dica).
- Yuriy, Chesnoloy. (2007). Face Detection C++ Library with Skin and Motion Analysis. Biometrics AIA 2007 TTS. 22 November 2007, Moscow, Russia. [http://www.codeproject.com/KB/audio-video/face\\_detection.aspx](http://www.codeproject.com/KB/audio-video/face_detection.aspx) (Acessado em 19/02/2010).
- Zhao, Gangqiang, Chen, Ling, Song, Jie and Chen, Gencai (2007). “Large Head Movement Tracking Using SIFT-Based Registration”. *MM’07*, Augsburg, Bavaria, Germany. ACM 978-1-59593-701..