

VISÃO COMPUTACIONAL COMO FORMA DE GERAR INTERATIVIDADE EM UMA PERFORMANCE ROBÓTICA

Julia Ghorayeb Zamboni, juliaghorayeb@yahoo.com.br
Instituto de Artes, UnB

Gabriel Gomes Gaspar, gabriel.g.gaspar@hotmail.com
Flávio de Barros Vidal, fbvidal@cic.unb.br
Dep. Ciência da Computação, UnB

André Luiz Gonçalves Paiva, apaiva90@gmail.com
Instituto de Geociências, UnB

João Gabriel Gomes, joao_unb_enm@hotmail.com
Nathan Costa Alves Souza, nathansouza110@gmail.com
Paulo Henrique Marques de Oliveira, paulohenriquemo@gmail.com
Jones Yudi Mori Alves da Silva, jones@unb.br
Dianne Magalhães Viana, diannemv@unb.br
Dep. Eng. Mecânica, UnB

Maria Luiza Pinheiro Guimarães Fragoso, malufragoso@gmail.com
Dep. Comunicação Visual, EBA, UFRJ

Resumo. *Ao construir robôs capazes de interagir com o mundo, é preciso utilizar metodologias para extração de informações sobre seus ambientes. A capacidade de perceber não só a sua própria condição, mas também a condição externa são essenciais para performances robóticas, seja em um caráter funcional ou artístico. A maioria dos sensores usados em robôs foram inspirados nas habilidades sensoriais de animais, tais como visão, audição, tato e senso de orientação espacial. Seus correspondentes robóticos são as câmeras de vídeo, microfones, sensores de toque, giroscópio e acelerômetros. O presente trabalho aborda o desenvolvimento do modelo de visão computacional de detecção de faces para a performance robótica Cacotecnia. Na exposição, que ocorreu em janeiro de 2013, o robô Caco atuou como bailarina em uma caixa de música em grande escala. Foram escolhidos dois grupos principais de movimentos para a produção do comportamento: a simulação da respiração, que refletia os estados emocionais do robô, e a habilidade interativa de olhar na direção dos espectadores e acompanhá-los com os olhos e pescoço. Os requisitos do projeto artístico foram as definições dos estados emocionais do robô: sono, medo, calma, despeito e delírio, e a probabilidade de cada estado ocorrer. A integração da visão computacional forneceu dados sobre o posicionamento das pessoas no ambiente para possibilitar o comportamento interativo do robô. Para executar esta tarefa, foi necessário o desenvolvimento de três fases: reconhecimento de estímulos ambientais, interpretação dos estímulos e codificação das respostas.*

Palavras-chave: *performance robótica, visão computacional, interatividade.*

Abstract. *In order to build robots capable of interacting with the world, we need to provide them with methods for extracting information about their environment. The ability to perceive not only one's own condition, but also the external condition is essential for robotic performance, either in a functional or artistic way. Most sensors used in robots are inspired by animal's sensory abilities, such as vision, hearing, touch and sense of spatial orientation. Their robotic matches are video cameras, microphones, touch sensors, gyroscope and accelerometers. This article is about the development of the computer vision model for face detection in Cacotecnia robotic performance. In the exhibition, which took place in January 2013, Caco, the robot, performed as a ballerina in a large scale music box. Two main groups of motions were chosen for the production of behavior: breath simulation, which reflected the robot's*

emotional states, and the interactive ability to look towards the spectators and accompany them with the eyes and neck. The requirements of the artistic project were the robot's emotional state definitions: sleep, fear, calm, resentment and delirium, and the probability of occurring each state. Computer vision integration provided data on people's place in the environment in order to allow the robot's interactive behavior. To perform this task, three phases were necessary: the recognition of environmental stimuli, stimuli interpretation and response codification.

Key words: *robotic performance, computational vision, interactivity.*

1. Introdução

O empenho da humanidade em produzir objetos que replicam seres vivos é percebido pelo vasto legado cultural destas obras. Tradicionalmente, as obras de arte performáticas e visuais, além dos ídolos, imagens funerárias, bonecas e autômatos, reproduzem os fenômenos da vida por meio da imitação da aparência e dos movimentos dos animais, especialmente dos seres humanos. As figuras animadas possuem aspectos de vontade própria e autonomia. A animação de aparatos pode ocorrer por meio da manipulação humana direta, como ocorre no teatro de formas animadas, ou pela criação de sistemas autônomos (BURNHAM, 1973).

A criação destes objetos foi desenvolvida junto com as mudanças tecnológicas de cada tempo. Em 1642, Pascal inventou a máquina de calcular aritmética, um dos primórdios do computador, com a capacidade de resolver problemas matemáticos. O trabalho de Pascal foi levado adiante no século seguinte pelo matemático e filósofo Gottfried Wilhelm von Leibniz, considerado prenunciador do desenvolvimento das técnicas de inteligência artificial. A partir dessa época, além do objetivo tradicional de construir réplicas de humanos e animais, a produção dos autômatos também objetivava o desenvolvimento de sistemas inteligentes capazes de resolver problemas matemáticos (PATTERSON, 2005).

No século XX, surgiram os primeiros robôs, desenvolvidos pelos princípios da Cibernética (WIENER, 1970). Os robôs desempenham operações análogas às do sistema nervoso dos organismos vivos e podem ser providos de memória, aprendizagem e órgãos sensoriais por meio de termômetros, câmeras de vídeo e sonares. Eles são capazes de estimar dados relativos ao seu estado interno e sobre as condições externas do meio, de forma que podem desempenhar ações pré-estabelecidas sobre o ambiente com comportamento reativo e autônomo.

Junto com o advento dos primeiros robôs, a arte robótica foi concretizada propriamente nas décadas de 50 e 60. Nicolas Schoffer (BURNHAM, 1973) foi um dos primeiros artistas a utilizar o potencial de aparatos robóticos na arte, com a obra CYPSP 1 (nome derivado da conjunção das iniciais de *cybernetic* e *spatiodynamic*) de 1956. Essa obra foi projetada como uma escultura capaz de reagir às diferenças ambientais. Constituída de uma estrutura articulável presa a uma base fixa, possuía alguns sensores analógicos que, ao detectarem a presença de observadores, produziam diferentes movimentos em sua parte articulável. Se o ambiente estivesse escuro e silencioso, a escultura era animada; se o ambiente estava iluminado e ruidoso, ela ficava parada. Estímulos ambíguos produziam comportamentos imprevisíveis. Nas palavras de Eduardo Kac: “ao passar do eletromecânico ao eletrônico, a obra de Schoffer criou uma ponte entre a arte cinética e a arte robótica” (KAC, 1997).

Os órgãos sensoriais dos robôs geram informações sobre o estado de seu próprio sistema e o estado do ambiente (BEKEY, 2005). Estas informações estão diretamente

relacionadas com a sua capacidade de cumprir tarefas. Os sensores podem classificados como exteroceptivos e propioceptivos. Sensores exteroceptivos adquirem informações sobre o ambiente de atuação do robô (exemplo: distância entre os objetos, intensidade de luz do ambiente,...). Os sensores propioceptivos fazem a medição dos elementos internos do robô (ex.: nível da bateria, posição e orientação dos dispositivos de locomoção). A maior parte dos sensores utilizados em robôs foi inspirada nas capacidades sensoriais dos animais, tais como: a visão, a audição, o tato, o senso de orientação espacial. Seus correspondentes robóticos são as câmeras de vídeo, os microfones, os sensores de tato, o giroscópio e acelerômetros o ultrassom e o sonar. No entanto, existem sensores que não foram biologicamente inspirados, tais como os detectores de intensidade luminosa, por exemplo, o LDR (Light Dependent Resistor), que mede a intensidade da luz no ambiente (muito usado na automatização de iluminação pública); o sensor *Geiger*, para medição de radiação alfa e gama; e os sensores de efeito *Hall* para campos magnéticos. (BEKEY, 2005).

A quantidade de informação disponibilizada pelos sensores pode ser mínima, com valores restritos a “sim” ou “não” (conhecidos como sensores booleanos) – como, por exemplo, os sensores de presença que identificam se algo passou por determinado lugar ou não, mas não é capaz de identificar mais nenhum dado, como a posição contínua no espaço do objeto que passou por ele. Outros sensores, como câmeras, geram dados mais complexos, como a distância relativa de um objeto que se aproxima ou informações sobre sua forma e/ou aparência.

Este artigo apresenta o desenvolvimento da visão computacional de um robô capacitando-o a identificar e interagir com o público, no escopo da exposição Cacotecnia. O sensor utilizado neste projeto é a câmera de vídeo que permite ao robô identificar faces no ambiente e escolher aquela com a qual irá interagir.

2. Exposição Cacotecnia

Cacotecnia:

s. f. || falta de arte. F. gr. Kakos (mau)+tekhne (arte)+ia.
Dicionário Aulete

Ao entrar na galeria, o espectador se depara com uma grande caixa de madeira ornamentada. Na parede central há uma figura com uma indicação do que se encontra em seu interior, com o perfil de uma cabeça sobre uma estrutura mecânica que alude às vértebras e olhos humanos, Fig. 1.



Figura 1: Detalhe do biombo.

A caixa apresenta duas entradas laterais, que levam o espectador para seu interior. Dentro, o ambiente é mais escuro e intimista, forrado com carpete azul escuro, onde o espectador ouve uma música metálica, típica de caixas de música, mas com a melodia pouco harmoniosa e com notas arrastadas e distorcidas. A Figura 2 mostra o ambiente externo da exposição.



Figura 2: Ambiente “Caixa de Música”.

No interior da caixa, sobre um palco com fundo ornamentado e debaixo de uma luz focal, encontra-se Caco, uma bailarina robótica. Ao contrário das bailarinas tradicionais das ornamentadas caixas de músicas, que apresentam uma rotina de movimentos tecnicamente precisos e delicados, Caco possui comportamento disfuncional, que se resume a alguns conteúdos fisiológicos e emocionais ordinários, como sono e medo. Além disso, o desempenho de Caco é imperfeito do ponto de vista técnico. Ele executa um repertório de movimentos imprecisos e barulhentos, de forma que seu caráter maquinal é bastante evidenciado. Seu comportamento não remete às melhores qualidades humanas, tão pouco, às melhores qualidades maquinais. A bailarina dorme em grande parte do tempo e, com sua falta de desenvoltura, fica tímida perante a audiência. A Figura 3 mostra o robô no interior do ambiente.

A respiração e o foco do olhar foram definidos como ações principais do robô. Estes dois movimentos foram escolhidos porque, para muitos bonequeiros, o movimento da respiração é o mais importante para denotar sinal de vida em um objeto e é normalmente o primeiro movimento a ser realizado para indicar a sua animação. A respiração reflete o estado emocional do personagem. Para “ampliar” o movimento da respiração, o som é sincronizado à movimentação do robô. Na obra *Cacotecnia*, o som da respiração é sugerido pelo ruído dos motores que geram estes movimentos.



Figura 3: Robô Caco.

O foco do olhar também é considerado uma das ações fundamentais dos bonecos, pois representa o estado mental do personagem. A visão é um dos sentidos pelo qual se percebe o mundo e é justamente a percepção do mundo externo que permite que o robô mantenha contato com o espectador. Além disso, o foco do olhar evidencia a interatividade do robô, quando o espectador compreende que Caco está percebendo e reagindo à sua presença.

A animação da respiração utilizada foi adaptada do modelo apresentado por Zordan *et al.* (2004) da Universidade da Califórnia. Este modelo foi desenvolvido para ser aplicado em animações computacionais 3D. Nesta obra, os autores fazem uma análise do movimento respiratório e dividem os tipos de respiração em lenta e profunda, ansiosa, calma e expiração forçada. O artigo ressalta que o movimento muscular tende a ter um amortecimento, uma força que resiste ao movimento, diminuindo a sua velocidade até o final de cada movimento. Os autores explicam que o efeito visual da respiração humana deriva da atividade de dois grupos musculares, o diafragma e os músculos intercostais, presos às costelas. Estes dois grupos musculares causam a movimentação dos ombros, do peitoral, dos braços, do abdômen, da espinha dorsal, até do movimento involuntário da cabeça. Durante a inspiração, enquanto o diafragma se move, o abdômen se projeta para frente e para os lados. Na expiração, o diafragma relaxa e o abdômen volta à sua posição. A respiração em repouso é mais perceptível na região do abdômen, apresentando movimentos controlados, com eventuais inspirações lentas e profundas. Esse tipo de respiração apresenta uma frequência entre 13 e 20 vezes por minuto em um ser humano adulto. A segunda respiração mais enérgica caracteriza-se por constantes e bruscos movimentos do tórax, onde são percebidos movimentos de expansão laterais e frontais das costelas e do diafragma. Nesse tipo de respiração, a taxa de respiração atinge, normalmente, valores entre 30 e 50 vezes por minuto, podendo ultrapassar um pouco esses valores.

Utilizando a linguagem *script* do circuito Mini Maestro¹, é possível animar e deixar gravada a animação da respiração, de acordo com o modelo descrito no trabalho de Zordan *et al.* (2004).



Figura 4: Interação do robô com o público.

¹ Os Mini Maestros são controladores de servomotores.

3. Integração da Visão Computacional

A integração da visão computacional no projeto teve como objetivo fornecer dados sobre o posicionamento de pessoas no ambiente de acordo com o campo visual e gerar o comportamento interativo do robô. Para a realização desta tarefa, foi necessária a implementação de três etapas em um procedimento computacional: a detecção dos indivíduos no ambiente, a interpretação das detecções realizadas e a codificação das reações.

Os requisitos estabelecidos no projeto artístico consistiam nos seguintes estados emocionais, a saber: sono, medo, calma, despeito e delírio, e a probabilidade de cada estado ocorrer, conforme estruturado e apresentado na Tabela 1, da definição do comportamento esperado do robô.

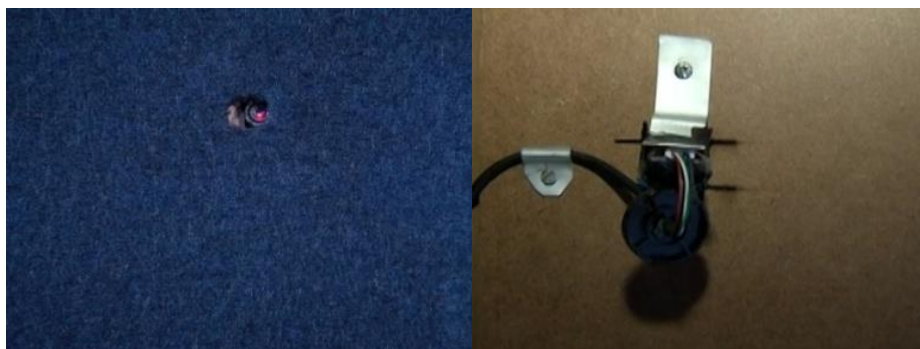
Entre as possíveis ações disponíveis para o robô executar, a ação desempenhada é escolhida por regras probabilísticas simples e ocorrem em função dos dados situacionais. Uma regra probabilística pode ser exemplificada como: Se não houver nenhuma pessoa no ambiente, o robô tem 40% de chance de dormir e 60% de chance de adotar o comportamento padrão. Quando uma ou mais pessoas são detectadas pelo robô ele também pode assumir um comportamento de medo, despeito ou delírio, definidos pela frequência da respiração e movimentos dos olhos e do pescoço (as combinações de ações esperadas são apresentadas nas Tabelas 1 e 2).

Tabela 1: Comportamento do robô.

Ação	Respiração	Olhos	Pescoço
Padrão	Calma. Este movimento alcança todo o sistema respiratório abdominal, com inspiração completa com o máximo de capacidade de ar e completa expulsão do ar.	Na direção do espectador. Eventuais piscadas com as pálpebras.	Na direção do espectador. Eventuais olhadas para os lados, para cima e para baixo.
Dormir	Lenta e profunda. Quantidade considerável de movimento abdominal devido à importante contribuição do diafragma.	Fechados. Inicia com movimentos lentos de abrir e fechar.	Cabeça caída. Com movimento de levantar, olhar para os lados e voltar a cair.
Despeito	Respiração lenta, com pausas entre a inspiração e a expiração. Ritmo menos contínuo. Algumas pausas maiores, seguidas de expiração acelerada.	Olhos parados na posição neutra.	No primeiro momento, olhar para o espectador por 3 segundos. Depois, virar pescoço na posição contrária à do espectador.
Medo	Ansiosa. Pequenas e rápidas inspirações e expirações. A maior parte do movimento é percebida no peito e na parte superior do torso.	Na direção do espectador (mais de um espectador, o olhar pode alternar entre eles). Pálpebras bem abertas.	Na direção do espectador (mais de um espectador, o olhar pode alternar entre eles).
Delírio	Lenta e profunda.	Olhos abrindo e fechando lentamente.	Pescoço com movimentos laterais, calmos e ritmados.

Tabela 2: Probabilidades de cada comportamento.

Variáveis do espectador	Probabilidade de ocorrer cada ação	Duração (segundos)
Presença de espectador no ambiente	Dormir: 20%	15
	Calmo: 20%	15
	Despeito: 30%	10
	Delírio: 10%	10
	Medo: 20%	15
Ausência de espectador no ambiente	Dormir: 40%	20
	Padrão: 60%	20



Figuras 5 e 6 : Câmera de vídeo.

Para se trabalhar com a programação para o processamento das imagens e vídeos, optou-se por utilizar as linguagens de alto nível C/C++, a partir da biblioteca OpenCV².

A biblioteca OpenCV possui rotinas para detecção facial utilizando o algoritmo de Viola-Jones com classificador *Boosted Cascade*. Este algoritmo utiliza a imagem captada pela câmera, detecta a região da face e delimita esta região por retângulos, mostrando o resultado na tela (Fig. 7). Esse classificador é, em essência, um conjunto de imagens que servirão de base para a busca, atualizando-se constantemente para se adaptar às variações de iluminação e tamanho das faces a serem detectadas (vide Fig. 7).

Desta forma, a partir das informações captadas pela câmera, o sistema de processamento faz a interpretação de estímulos e codifica as reações do robô, ou seja, a unidade processa os dados enviados pelo sensor e comanda as ações dos atuadores conectados ao robô. O processamento de imagens ocorre em tempo real³. A interpretação de estímulos é realizada por meio do aplicativo desenvolvido, que é capaz de escolher uma resposta dentre as possíveis ações programadas para aquele padrão identificado, gerando uma instrução (comando) para o robô. As instruções são codificadas e enviadas ao micro-controlador⁴ conectado ao computador principal,

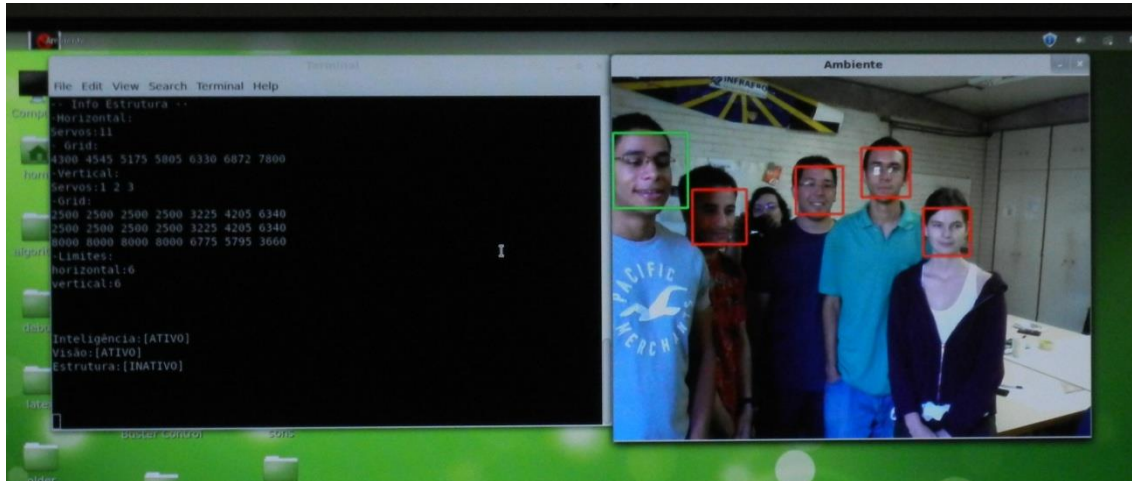
² Disponível em <http://opencv.willowgarage.com/wiki/Welcome>

³ Programas em tempo real devem garantir resposta dentro de restrições tempo estabelecidas. Muitas vezes, os tempos de resposta do sistema ocorre em milissegundos.

⁴ Os microcontroladores são unidades de processamentos completos em um chip; ou seja, ele possui os elementos principais de um computador (CPU, memória RAM e um processador). Normalmente, microcontroladores são usados para aplicações que necessitam de pouco processamento, como controle de motores e câmeras. Quando a limitação de memória ou tempo de processamento são um problema, normalmente são usados microcontroladores de maior potência ou ele é combinado com um computador.

usado para acionar os servomotores de forma apropriada. Deste modo, as ações podem ser codificadas em reações mecânicas, tais como a alteração no padrão de respiração e movimentos da cabeça e dos olhos

Figura 7: Detecção de faces.



4. Resultados

Neste artigo foi apresentada a estratégia utilizada com base na visão computacional como forma de gerar um robô interativo. O objetivo do trabalho foi produzir um robô capaz de identificar o posicionamento das pessoas, escolher uma delas, a seguir com movimentos da cabeça e dos olhos e reagir à sua presença por meio de respostas com conteúdos emocionais pré-definidos. Destarte, esta pessoa poderia acreditar que o robô Caco possuía vida, embora ele evidentemente fosse uma máquina. Isto ocorre quando o espectador passa por um processo de *suspension of disbelief*, ou seja, a suspensão de seus julgamentos sobre as limitações do robô em se aparentar e se comportar como um humano, aceitando a premissa da vida do robô (Tillis, 1992).

A exposição Cacotecnia foi apresentada na galeria Espaço Piloto, da Universidade de Brasília. A interatividade com o público foi verificada por meio da reação das pessoas aos movimentos realizados pelo robô. A exposição atendeu às expectativas e o projeto foi concluído com sucesso. Mais de 50% das pessoas que participaram da *performance* agiram como se o robô estivesse vivo e sujeito à reações emocionais em razão de suas interferências. A obra foi considerada em parte disfuncional e estranha, em parte familiar e fascinante por muitos visitantes.

Referências

AUSLANDER, P. **From Acting to Performance: Essays in Modernism and Postmodernism**. New York, Routledge. 1997.

AUSLANDER, P. Humanoid Boogie: Reflections on Robotic Performance. In: **Staging Philopshy: Intersections of Theatre, Performance, and Philosophy**, Editado por David Krasner e David Z. Saltz, Ann Arbor. University of Michigan Press, 2006.

BEKEY, G. A. **Autonomous Robots, From Biological Inspiration to Implementation and Control**. Cambridge, Massachusetts. The MIT Press. 2005.

BURNHAM, J. **Beyond Modern Sculpture: The Effects of Science and Technology on the Sculpture of This Century**. New York. George Braziller. 1973.

DEMERS L. P. Machine Performers: Neither Agentic nor Automatic. Paper presented at the Workshop "What do Collaborations with the Arts Have to Say About Human-Robot Interaction?" Held at the 2010 Human- Robot Interaction Conference, in Osaka, Japan. 2010.

KAC, E. Foundation and Development of Robotic Art. In: **Art Journal**, Vol. 56, nº3, Digital Reflexions: The Dialog of Art and Technology. 1997.

MATARIC, M. **The Robotics Primer**. Cambridge, Massachusetts. The MIT Press. 2007.

PATTERSON D.A., HENNESSY, J.L., **Organização e Projeto de Computadores – A Interface Hardware/Software**, terceira edição, Editora Campus. 2005.

VIOLA,P. & JONES, M. Rapid Object Detection using a Boosted Cascade of Simple Features. Conference on Computer Vision and Pattern Recognition (CVPR), pp. 511-518. 2001.

TILLIS, S. **Toward an Aesthetics of the Puppet. Puppetry as a Theatrical Art**. New York. Greenwood Press. 1992.

WIENER, N. **Cibernética: ou Controle e Comunicação no Animal e na Máquina**. Traduzido por Gita K. Ghinzberg. São Paulo. Editora Polígono. 1970.

VORN, B. Adaptive Machines for Interactive Robotic Art Installations. In: **Machines Agency: Artistic Perspectives**. University of the Arts (Bremen, Germany), eds. Christoph Lischka, Andrea Sick, pp.180-190. 2007.

ZAMBONI, J. G. **Performance Robótica: Aspectos Expressivos e Experimentais**. 94 p. Dissertação de Pós-Graduação em Artes. Universidade de Brasília – UnB, 2013.

ZORDAN, V. B., CELLY, B., CHIU, B. & DILORENZO P. C. **Breathe Easy: Model and Control of Simulated Respiration for Animation**. Editors R. Boulic, D. K. Pai. Eurographics/ACM SIGGRAPH Symposium on Computer Animation. 2004.