

Reconhecimento de ambiente

Sistema desenvolvido com tecnologia de *videogame* transforma imagens em sons para simplificar o cotidiano de deficientes visuais

Dinorah Ereno

A partir de uma tecnologia disponível no mercado, o sensor Xbox Kinect, lançado pela Microsoft para jogos de *videogame*, pesquisadores da Universidade Estadual de Campinas (Unicamp) desenvolveram um sistema, ainda em fase experimental, que converte vídeos em informações sonoras para auxiliar deficientes visuais em suas tarefas cotidianas. A tecnologia Kinect possui duas câmeras, uma delas com um sistema que emite e captura luz infravermelha, reconhecido por um algoritmo que confere um valor de profundidade – ou distância da câmera – para cada ponto da imagem. Com base nesses dados é possível identificar objetos e usuários no ambiente sem se confundir com os seus contornos e variações de luzes e cores. No projeto conduzido na Unicamp em parceria com o pesquisador Dinei Florêncio, da Microsoft, foi montado um protótipo que teve como suporte um capacete de *skate*. Em cima do capacete foi colocado um sensor Kinect com as duas câmeras – os itens restantes que fazem parte do equipamento original

foram retirados – conectadas a um *laptop* com grande poder de processamento de informações, além de um giroscópio, um acelerômetro e uma bússola, dispositivos que registram, em conjunto, mudanças de direção da cabeça.

O *laptop* é carregado pelo usuário dentro de uma mochila. “Os dados do ambiente captados pela câmera são processados pelo computador, que retorna a informação para o usuário em áudio”, diz Siome Klein Goldenstein, professor do Instituto de Computação da Unicamp e coordenador do projeto, financiado por FAPESP e Microsoft dentro do Programa Pesquisa em Parceria para Inovação Tecnológica (Pite). O usuário conta ainda com fones de ouvido com tecnologia *bone conduction*, em que o áudio é transmitido aos ouvidos por meio da condução do som pelos ossos do crânio. Isso permite que os ouvidos fiquem livres para escutar outras fontes sonoras do ambiente, e não apenas os *feedbacks* de áudio que o sistema fornece, o que ocorreria caso fossem utilizados *headphones* comuns.

Siome explica que durante o projeto foram estudados vários cenários de uso

para a tecnologia. Uma das aplicações que já estão prontas é um módulo de detecção e reconhecimento de pessoas em um ambiente fechado. Para que o sistema faça o reconhecimento, é preciso primeiro criar um cadastro de pessoas. “Utilizamos uma técnica de localização de face e empregamos um algoritmo para classificação de cada uma das pessoas do cadastro”, explica Laurindo de Sousa Britto Neto, doutorando no Instituto de Computação da Unicamp e participante do projeto. Quando o usuário entra em um recinto com pessoas previamente cadastradas, o módulo de detecção e reconhecimento de face passa a informação para o módulo de áudio 3D, configurado para reproduzir um som na localização espacial de cada um dos presentes na sala. “Ao escutar o nome de cada uma das pessoas, o usuário saberá a posição exata em que ela se encontra”, explica Siome. “É como se o som saísse da cabeça da pessoa identificada pelo módulo de detecção.” O objetivo é fazer com que o deficiente possa virar o rosto na direção certa e se comunicar da forma mais natural possível.





Sensor com câmeras e equipamentos que registram mudanças de direção da cabeça do usuário faz parte do protótipo da Unicamp



A técnica de áudio 3D foi um capítulo à parte no desenvolvimento do projeto porque cada ouvido reage de forma diferente aos estímulos do mundo externo, segundo Siome. A pesquisa com áudio 3D foi conduzida por Felipe Grijalva Arévalo, equatoriano que chegou ao Brasil em 2012 para fazer o mestrado sob orientação do professor Luiz César Martini, da Faculdade de Engenharia Elétrica e de Computação da Unicamp, deficiente visual que trabalha com pesquisas direcionadas para esse público. Como cada cabeça tem uma anatomia

diferente, é preciso recorrer a um processo chamado de personalização para gerar áudio 3D. “A abordagem mais simples consiste em usar medidas já conhecidas de anatomia da orelha e, a partir daí, traçar um modelo de personalização”, relata Felipe. Ele pesquisou o tema na sua dissertação de mestrado, coorientado por Siome, e atualmente continua a trabalhar com áudio 3D em sua tese de doutorado. Os testes foram feitos inicialmente com pessoas vendadas; na próxima etapa, após aprovação pelo comitê de ética da Unicamp, serão

conduzidos com deficientes visuais. Após a realização de todos os experimentos necessários para a validação do novo sistema, ele poderá ser compartilhado com toda a comunidade.

ABORDAGEM SIMPLIFICADA

Outra aplicação que está em fase de finalização é um sistema de auxílio à navegação capaz de informar ao usuário onde há obstáculos à sua frente. “O Kinect posicionado na cabeça do deficiente irá captar os quadros do vídeo que serão segmentados, por meio de algoritmos, em diversos planos, como chão, porta, escada”, explica Laurindo Neto. As informações são então repassadas em áudio para o usuário. O projeto teve a participação de Vanessa Maíke sob orientação de Maria Cecília Baranauskas, do Instituto de Computação (IC), que trabalhou com experimentos na área de interfaces naturais, que envolvem formas naturais de interação entre as pessoas e os computadores, e de Anderson Rocha, também do IC, na parte de biometria.

A ideia de transformar imagens em som para deficientes visuais não é nova, mas as tecnologias utilizadas pelo grupo de pesquisa da Unicamp resultaram em uma abordagem de reconhecimento mais simples para o usuário, com reprodução do ambiente natural. O dispositivo mais conhecido no mercado com esse conceito é o The vOICE, criado por Peter Meijer, do Laboratório de Pesquisas Philips, na Holanda. O sistema consiste de uma câmera comum acoplada a óculos – também pode ser utilizada a do *smartphone* –, que captam as imagens do ambiente, convertidas em imagens por um computador e transmitidas por meio de fones de ouvido. Nesse caso, a câmera capta as cenas da esquerda para a direita e essas informações são recebidas pelo usuário no ouvido esquerdo e direito, alternadamente. ■

Projetos

1. Visão para o cego: traduzindo conceitos visuais 3D em informações de 3D de áudio (nº 2012/50468-6); **Modalidade** Pesquisa em Parceria para Inovação Tecnológica (Pite); **Pesquisador responsável** Siome Klein Goldenstein (Unicamp); **Investimento** R\$ 32.648,40 (FAPESP) e R\$ 32.648,40 (Microsoft).
2. Aprendizado de máquina em processamento de sinais aplicado ao áudio espacial (nº 2014/14630-9); **Modalidade** Bolsa de Doutorado; **Pesquisador responsável** Luiz César Martini (Unicamp); **Bolsista** Felipe Leonel Grijalva Arévalo (Unicamp); **Investimento** R\$ 136.000,80.