COGNITO – Captura, reconhecimento e visualização de atividades manuais complexas

Gustavo Maçães Hugo Domingues Luis Almeida
Centro de Computação Gráfica
Campus Azurém, Guimarães
{gustavo.macaes,hugo.domingues,luis.almeida}@ccg.pt

Luis Paulo Santos
Dep. Informática, Universidade do Minho
Campus Gualtar, Braga
psantos@di.uminho.pt

Resumo

Neste artigo curto apresenta-se um sistema capaz de automaticamente capturar, reconhecer e visualizar atividades motoras humanas, em diferentes contextos, mas com aplicação prática, por exemplo, na implementação de manuais virtuais 3D ou em vídeo-jogos de nova geração, passando pelos simuladores de treino. Este trabalho tem vindo a ser desenvolvido em consórcio internacional, no contexto de um projeto apoiado pela comissão europeia (COGNITO), e tem-se centrado na captura, análise, armazenamento e visualização 3D, com recurso a tecnologias de realidade virtual e aumentada, de tarefas manuais complexas, executadas em ambiente industrial. O sistema é composto por quatro módulos principais: uma rede de sensores colocados no corpo, uma unidade de captura dos movimentos e ferramentas utilizadas, uma componente de aprendizagem não supervisionada e uma componente gráfica capaz de fazer a apresentação de informação ao utilizador através de um módulo de realidade aumentada (RA). Este artigo apresenta o sistema global e a sua arquitetura, referindo com mais detalhe os desenvolvimentos efetuados para a componente gráfica.

Palavras-Chave

Realidade Aumentada, Realidade Virtual e Visão por computador.

1. INTRODUÇÃO

A captura, reconhecimento e representação gráfica da atividade humana tem potencial para gerar diferentes tipos de aplicações desde manuais virtuais, simuladores 3D, vídeo-jogos de nova geração ou automatização de robôs. Contudo, os sistemas de captura de atividade humana existentes no mercado focam-se essencialmente na captura dos dados alinhando os mesmos com um esqueleto predefinido sem preocupação com o tipo de atividade executada. Esta abordagem traz problemas na criação de manuais virtuais ou simuladores porque normalmente estes ficheiros são extensos e implicam pós processamento por parte de um técnico. Nos últimos anos têm sido obtidos avanços significativos no campo da RA, tanto a nível de software como do hardware [Teichrieb07]. No campo do hardware, estão disponíveis novos HMDs com ecrãs e câmaras com alta resolução, novos sistemas de visão por computador e sensores que permitem um *tracking* mais robusto e a implementação de aplicações mais apelativas. Em particular, os sistemas com *tracking* baseado em modelos [Drummond02][Pupilli06], combinam visão por computador e sensores inerciais [You99] e mais recentemente a técnica SLAM - Simultaneous Localization and Mapping [Davison07]. No campo do software para desenvolvimento aplicacional tem-se assistido ao aparecimento de múltiplas soluções, tais como, por exemplo, o NyARToolkit [NyARToolkit08], StudierStube [StudierStube08], SLARToolkit [SLAR10] e PTAM [Klein07].

Todas estas tecnologias permitem o surgimento de projetos para a implementação de sistemas avançados, que delas tirem partido, como é o caso do projeto MATRIS [Comport06], AMIRE [Grimm02], ARVIKA [Friedrich02] e DWARF [Bauer01].

O projeto COGNITO [COGNITO10] surge nesta sequência e propõe-se especificar e implementar o protótipo de um sistema que irá permitir a criação de um assistente pessoal capaz de dar suporte ao utilizador na realização de atividades e na manipulação de ferramentas para a realização de tarefas industriais complexas. Para isto ser possível é necessário começar por capturar um conjunto de informação relativa aos movimentos do indivíduo e dos objetos da cena (ferramentas que estão a ser utilizadas, etc), utilizando um conjunto de sensores de movimento e de imagem, devidamente colocados no corpo, ao que se segue o processamento necessário à identificação

e segmentação da atividade em ações atómicas. O resultado desta análise é exportado em formato XML, verificado e filtrado por um técnico experiente que identificará as ações atómicas reconhecidas pelo sistema com apoio de um editor desenvolvido para o efeito. Após este tratamento, a informação armazenada está preparada para ser visualizada na componente de RA, durante a execução de tarefas por qualquer utilizador. A secção seguinte apresenta com mais detalhe o sistema COGNITO. A secção três apresenta o estado atual dos elementos constituintes da componente gráfica. Na secção quatro são descritos os cenários de teste e por fim, na secção cinco, são apresentadas as principais conclusões e indicações relativas ao trabalho futuro.

2. ARQUITETURA COGNITO

O sistema COGNITO é complexo e inclui hardware integrado que vai gerar uma grande quantidade de dados, a uma velocidade elevada. O sistema terá de lidar com fluxos de dados heterogéneos da rede de sensores e ainda disponibilizar informação, na forma de RA, em tempo real, ao utilizador. Para além disso, o sistema deve permitir a mobilidade do utilizador, sendo por isso leve e portátil.

A figura 1 apresenta uma representação gráfica do sistema e respetivos módulos constituintes**.**

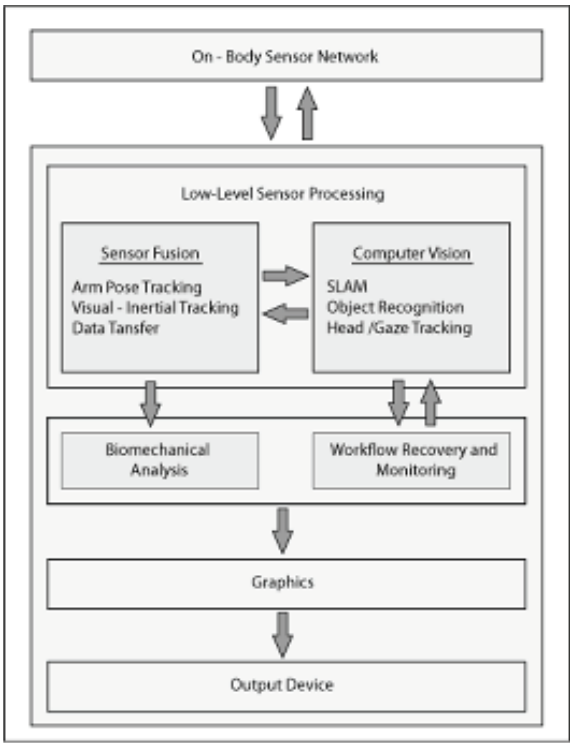

Figura 1: Arquitetura COGNITO

O sistema COGNITO pode ser divido em 4 módulos principais:

- Rede de Sensores
- Sistema de Captura de Dados
- Componente de Aprendizagem
  - Análise biomecânica
  - Identificação de workflow
- Componente gráfica

A rede de sensores é composta por seis unidades de medição inercial (IMU), que detetam os movimentos do tronco e braços do utilizador. Inclui também uma câmara que fica colocada no peito do utilizador com uma lente olho de peixe que permite visualizar a tarefa a 180º e guardar as imagens captadas para posterior identificação e análise. A rede de sensores fica completa com uma câmara RGB-D que dá informação relativa a profundidade da área de trabalho que ajuda a identificar as ferramentas que o utilizador manipula (fig. 2).

O módulo de captura de dados é responsável por guardar e integrar os diferentes fluxos de dados. Este processamento é feito por uma unidade de visão por computador, que trata das imagens provenientes das câmaras, e por uma outra unidade responsável por adaptar todos os dados ao mesmo sistema de coordenadas (referencial).

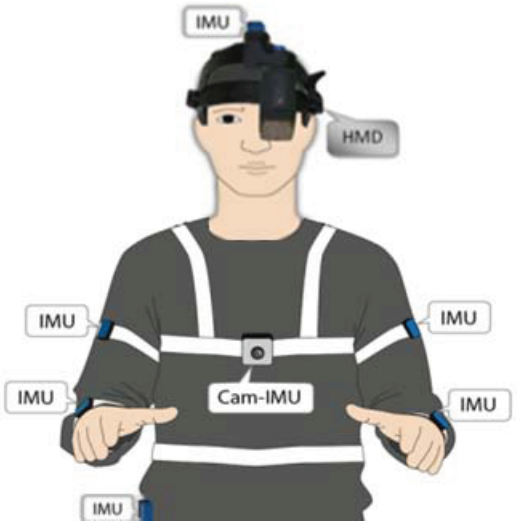

Figura 2: Rede de sensores

Em seguida os dados serão tratados pelo módulo de aprendizagem que analisará os movimentos e objetos manipulados, efetuando uma segmentação da atividade em tarefas atómicas e registando o tempo inicial e final das mesmas.

Por fim, é feito o registo da informação processada, em formato XML, utilizando para o efeito um modelo de ações definido especificamente para o COGNITO. A informação gerada neste processo, será utilizada num editor, com o qual técnicos experientes analisam e filtram as tarefas atómicas definidas e acrescentam informação adicional que posteriormente será usada pelo visualizador RA. Quando todas as ações se encontram identificadas o template de ações é armazenado para ser utilizado pelo assistente (visualizador de RA) que irá guiar o utilizador na execução das tarefas.

3. COMPONENTE GRÁFICA

Esta componente pode ser divida em duas aplicações principais: um editor da informação capturada e processada na forma de um fluxo de ações elementares e um visualizador de RA para apresentação de toda a informação, utilizando um HMD.

A função principal do editor centra-se na identificação, por parte de um utilizador experiente, das tarefas atómicas que o sistema deteta e guarda no modelo, associando-lhes toda a informação que deverá ser apresentada na forma de RA (modelos virtuais 3D, designação e descrição das tarefas, etc). O Editor (fig. 3) é composto por três áreas principais: na parte lateral esquerda é feita uma representação gráfica do template de ações que contém as tarefas atómicas; na parte lateral direita é apresentado o vídeo ilustrativo da ação atómica selecionada (captado em tempo de recolha de dados); a parte inferior da aplicação é reservada à introdução da informação correspondente a cada ação atómica.

O processo de edição de tarefas atómicas consiste, além da atribuição da designação inteligível, na inclusão de informação adicional que serve para alimentar o visualizador de RA. Esses elementos podem ser objetos virtuais 3D, imagens, vídeos, sons e ainda descrição mais detalha acerca da atividade que é depois fornecida ao utilizador final recorrendo a um conversor de texto para áudio(TTS).

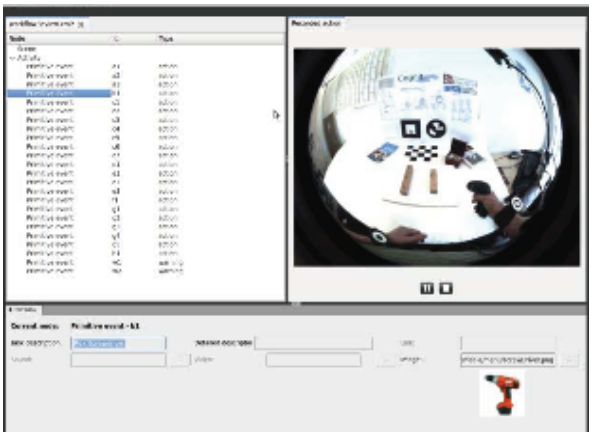

Figura 3: Componente Gráfica - Editor

Após o processo de edição, o modelo de ações fica atualizado para ser lido pelo visualizador RA. A visualização em RA destina-se a utilizadores finais que vão executar tarefas manuais e com necessidade de instruções/ilustrações de procedimento passo a passo (em termos de tarefas atómicas). Esta abordagem imersiva recorrendo a RA pretende manter o utilizador focado na sua tarefa.

O visualizador de informação RA pode ser utilizado de duas formas distintas: integrado no sistema ou de modo autónomo. Quando integrado no sistema, é controlado pelo componente de aprendizagem, a qual, recebendo a informação dos sensores e sabendo quais as ferramentas que estão a ser usadas, quando compara esses dados com as atividades atómicas previamente analisadas, consegue transmitir instruções ao módulo de RA para apresentar ao utilizador a tarefa corrente e os objetos empregues (fig.4).

Figura 4: Componente Gráfica - Visualizador RA

No modo autónomo é o utilizador que controla a aplicação através de comandos de voz.

Atualmente, as combinações funcionais que o visualizador RA disponibiliza, em termos de informação apresentada e controlo, são as seguintes:

- Texto + Instrução áudio
- Texto + Vídeo Ilustrativo + Instrução áudio
- Texto + Modelos 3D + Instrução áudio

O utilizador pode navegar nas atividades atómicas podendo avançar, voltar atrás ou pedir para repetir a instrução. A informação de RA apresentada ao utilizador pode variar consoante as suas preferências ou o local onde se encontra a trabalhar. A informação textual e o progresso da atividade encontram-se sempre disponíveis. Além do texto, as descrições alargadas das tarefas atómicas podem ser apresentadas na forma de áudio através da funcionalidade de TTS. Associado à tarefa atómica existe também um pequeno vídeo representativo da tarefa, o qual ilustra o procedimento a executar. O recurso à visualização de modelos virtuais 3D em sobreposição contextualizada à imagem real, é, no entanto, a principal funcionalidade de RA disponível no visualizador, com os quais é possível exemplificar os movimentos a realizar, as ferramentas a empregar, indicar os pontos do espaço físico onde a ação deve ser executada, entre muitas outras representações gráficas possíveis.

4. CENÁRIOS DE TESTE

Atualmente o sistema COGNITO tem vindo a ser testado com sucesso em dois cenários distintos, os quais, não sendo o definitivo (complexo), permitem cumprir o objetivo de realizar testes que validem os vários módulos do sistema e a comunicação entre eles. O primeiro consiste numa tarefa aparentemente simples que instrui o utilizador para posicionar um taco de madeira numa região previamente marcada, fixá-la com três pregos usando um martelo; de seguida, o utilizador deve pegar noutro taco ligeiramente mais pequeno e aparafusar em três locais previamente estabelecidos, usando uma parafusadora. O

segundo cenário testado consiste no empacotamento de embalagens de sabão líquido. Nesta sequência de atividades o utilizador tem que pegar numa embalagem e etiquetá-la; posteriormente coloca-a numa caixa que é fechada e escreve a morada para onde a mesma deve ser enviada.

Estes cenários de teste, apesar de simplificados permitiram já validar todo o sistema COGNITO, com as suas várias componentes integradas e comunicantes, prevendo-se que o último cenário a ser utilizado, no fim do projeto, consista num ambiente industrial real.

5. CONCLUSÕES E TRABALHO FUTURO

O sistema COGNITO, aqui apresentado, é capaz de aprender ações executadas por um utilizador, de forma autónoma, através de um conjunto de sensores de inércia colocados no corpo e recorrendo a tecnologias de visão por computador. A componente de RA apresenta instruções multimédia, devidamente contextualizadas no espaço físico, para utilizadores não experientes, instruindo a realização de tarefas passo a passo. O uso de realidade aumentada e virtual apresenta-se como uma solução interessante para ajudar pessoas na execução de tarefas, sendo mais eficazes que os tradicionais manuais em papel [Tang03]. Em termos de trabalho futuro, há ainda evoluções a realizar, como por exemplo o cálculo detalhado da posição dos elementos 3D, a partir do reconhecimento de ações em tempo real.

O estudo e melhoramento das representações (texto, elementos gráficos, elementos 3D, etc) e da funcionalidade de reconhecimento de voz no visualizador de RA serão efetuados em conjunto com o parceiro industrial.

A aplicação e teste do sistema em cenários mais complexos é objetivo final do projeto.

6. AGRADECIMENTOS

Este trabalho tem o apoio da Comissão Europeia através do sétimo programa quadro, sob o contracto nº COGNITO FP7-ICT248290. Um agradecimento à Doutora Elizabeth Carvalho pela colaboração no projeto.

7. REFERÊNCIAS

[Bauer01] Martin Bauer, Bernd Bruegge, Gudrun Klinker, Asa Williams, Thomas Reicher e Martin Wagner Distributed Wearable Augmented Reality Framework (DWARF) Design and Implementation of a Module for the Dynamic Combination of Different Position Tracker, Augmented Reality IEEE, 2001.

[COGNITO10] COGNITO Project, FP7, ICT-2009.2.1 Cognitive Systems and Robotics, ICT – Information and Communications Technologies, number ICT–24829, <http://www.ict-cognito.org/index.html>

[Comport06] Andrew Comport, Eric Marchand, Muriel Pressigout e François Chaumettex. "Real-time markerless tracking for augmented reality: the virtual visual servoing framework," *Visualization and Computer Graphics, IEEE* 2006.

[Davison07] Andrew Davison, Ian Reid, Nicholas Molton e Olivier Stasse, "MonoSLAM: Real-Time Single Camera SLAM", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007.

[Drummond02] Tom Drummond e Roberto Cipolla, "Real-time visual tracking of complex structures", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002.

[Friedrich02] Wolfgang Friedrich "ARVIKA-augmented reality for development, production and service," Mixed and Augmented Reality, 2002. ISMAR 2002.

[Grimm02] Paul Grimm, Michael Haller, Volker Paelke, Silvan Reinhold, Christian Reimann e Jurgen Zauner "AMIRE - authoring mixed reality" *IEEE* 2002.

[Klein07] Georg Klein e David Murray "Parallel Tracking and Mapping for Small AR Workspaces," *Mixed and Augmented Reality, 2007. ISMAR 2007.*

[NyARToolkit08] NyARToolkit (available at <http://nyatla.jp/nyartoolkit/wiki/index.php?FrontPage.en>, visited on 31-5-2012).

[Pupilli06] Mark Pupilli e Andrew Calway, "Real-time Camera Tracking Using Known 3D Models and a Particle Filter" 18th International Conference on Pattern Recognition 2006.

[SLAR10] SLARtoolkit (available at <http://slartoolkit.codeplex.com/>, visited on 31-5-2012).

[StudierStube08] StudierStube Project (available at <http://studierstube.icg.tugraz.at/main.php> visited on 31-5-2012).

[Tang03] Arthur Tang, Charles Owen, Frank Biocca e Weimin Mou, Comparative effectiveness of augmented reality in object assembly, SIGCHI 2003.

[Teichrieb07] Veronica Teichrieb, João Lima, Eduardo Apolinário, Thiago Farias, Márcio Bueno, Judith Kelner e Ismael Santos, "A Survey of Online Monocular Markerless Augmented Reality", 2007.

[You99] Suya You, Ulrich Neumann e Ronald Azuma, "Hybrid inertial and vision tracking for augmented reality registration", Proceedings of Virtual Reality IEEE 1999.