Smart Spaces

Recerca en Interfícies Multimodals per afrontar els reptes de les tecnologies del futur

Josep R. Casas, Cristian Canton-Ferrer, Grup d'Imatge
Departament de Teoria del Senyal i Comunicacions, ETSETB-UPC

D'aquí no gaire, allà on anem, trobarem tecnologies que capten l'entorn, que hi actuen, que es comuniquen, raonen i interaccionen amb les persones amb la intenció de fer del món un lloc millor per viure i treballar[1]. Conforme es van estenent i, especialment, en augmentar les seves capacitats de comunicació, la complexitat intrínseca d'aquestes tecnologies genera reptes inèdits d'enginyeria de sistemes i problemes fonamentals en intel·ligència artificial. La interacció amb les persones requereix aproximacions multidisciplinars d'enginyeria, psicologia i ciències del coneixement, que centren el disseny en l'usuari canviant profundament les interfícies dels sistemes computacionals.

Entorns intel·ligents

Un 'Smart Space' (espai o ambient intel·ligent) és un entorn dotat d'elements amb capacitats sensores, que pensen, actuen, es comuniquen entre ells i interaccionen amb les persones. Llevat d'algunes excepcions, els ambients intel·ligents són avui espais destinats a la recerca, però aviat controlaran avions, gestionaran processos industrials, vigilaran les granges, els conreus, les reserves d'aigua i el clima. La miniaturització els farà menys visibles, més autònoms i més fàcils d'usar. La comunicació en xarxa permetrà integrar coneixements adquirits per a l'aprenentatge o per millorar els serveis de salut en el sector sanitari. Canviaran radicalment la manera com les persones interaccionen amb l'entorn, perquè compilaran informació complexa per ser compartida i usada a l'hora de prendre decisions. Seran a tot arreu, ajudant-nos en la nostra feina. S'adaptaran a les necessitats de les persones i esdevindran una extensió natural de les nostres capacitats.

Els ambients intel·ligents han de disposar d'una descripció de l'entorn i dels esdeveniments que s'hi produeixen ('environmental awareness'). També han de conèixer el seu propi estat, les seves capacitats i funcionalitats ('self-awareness'). Convé que tinguin la capacitat d'arxivar i recordar esdeveniments anteriors. Mantenir aquesta informació de manera continuada en el temps, els permet adaptar-se als canvis dinàmics de l'entorn i interaccionar amb les persones de forma natural, eficient i flexible, i estendre la interacció de la manera més adequada en l'espai i en el temps.

Figura 1. Canvi de paradigma: l'ordinador situat al centre de l'atenció de les persones (imatge esquerra), o proveint els seus serveis de la manera adequada en un entorn natural d'interacció humana (imatge dreta).

Repte per a la recerca

La comunitat científica reconeix els reptes que les necessitats anteriors comporten en moltes disciplines. En particular, la relació dels espais intel·ligents amb les persones és objecte d'especial atenció. Les "interfícies multimodals", entorns computacionals fàcils d'usar que es comuniquen de manera natural amb l'usuari (en diferents modalitats i llenguatges), aspiren a ser l'antídot contra "l'autisme" dels espais intel·ligents i, per tant, han esdevingut una prioritat en recerca[2].

Els ambients intel·ligents canviaran radicalment la manera com les persones interaccionen amb l'entorn, perquè compilaran informació complexa per ser compartida i usada a l'hora de prendre decisions.

Improved HCI versus Enhanced HHI

El paradigma actual de les interfícies computacionals (HCI) gira al voltant d'una pantalla i un teclat. L'usuari s'introdueix en el flux de treball de la màquina i ha de conèixer les tasques definides i programades per a operar-la explícitament amb comandes predeterminades. Això situa l'ordinador al centre d'atenció de les persones, desviant l'atenció de la tasca real que ens porta a emprar-lo. Tenim doble feina perquè, si volem que ens ajudi en el què volem fer, hem d'entendre la màquina que ens proveeix el servei per poder-la manipular. Hem de treballar per "l'ajudant", i això limita l'eficàcia del servei que ens proporciona.

La visió de futur allunya l'ordinador del centre d'atenció sense afectar els serveis

[1] De fet, aquestes tecnologies ja hi comencen a ser, però potser ni ens n'adonem. Hi ha portes que s'obren quan passem, l'ascensor ens demana que sortim quan hi ha excés de pes, el nostre mòbil es registra a la base més propera, el neteja-parabrisa s'engega amb un sensor de pluja i el rec automàtic decideix que avui no toca regar perquè aquest matí ha plogut. Encara fan tasques poc complexes, donat que una limitació fonamental és que no solen comunicar-se gaire en xarxa i, menys encara, amb les persones.

[2] Les "interfícies multimodals" són objectiu estratègic del 6è Programa Marc de la Unió Europea en l'àrea de Tecnologies de la Informació i la Societat (IST).

d'informació i comunicació que ens dóna. Per canviar el paradigma, cal evitar que les persones quedin atrapades en els procediments de les màquines. Ben al contrari: cal que els ordinadors observin, escoltin, entenguin i es comuniquin amb les persones en l'entorn natural d'interacció humana (HHI). Els sistemes d'informació han de proveir els seus serveis com ho faria un ajudant humà, molestant el menys possible, mantenint-se en un segon terme i intentant predir quan es generarà la necessitat.

Tecnologies d'Anàlisi i Síntesi Multimodals

En els espais intel·ligents, els ordinadors es relacionen amb nosaltres mitjançant senyals àudio-visuals. Han de tenir en compte com es comuniquen les persones i emprar tecnologies d'anàlisi multimodal que permetin la descripció completa i la comprensió dels senyals de comunicació humana en totes les modalitats (parla, gestos, imatges, text, sons, signes...). S'han de comunicar amb els usuaris en la modalitat més adient per cada situació (parlant-los a cau d'orella, si cal) i explotar diferents tecnologies de diàleg i interacció natural emprant "avatars" o sistemes de síntesi d'àudio i vídeo dirigits ('*targeted audio/targeted video*').

L'observació, l'estudi i la comprensió dels senyals de comunicació humana en espais intel·ligents requereix infrastructures adients que facin aquest problema de recerca més tractable. Les "Sales Intel·ligents" proporcionen l'entorn d'experimentació ideal. Cal que disposin de sensors, xarxes de comunicació i arquitectures de processament distribuït per a l'anàlisi de l'entorn; requereixen actuadors i equipament acústic i d'imatge per a la síntesi de senyals; i necessiten metodologies específiques per gestionar la complexitat i derivar models d'interacció correcta amb les persones.

El projecte CHIL

Tres grups de recerca de la UPC en imatge, veu i llenguatge natural dels departaments de TSC i LSI participen en el projecte CHIL, '*Computers in the Human Interaction Loop*'. CHIL investiga tecnologies d'anàlisi que generen descripcions de l'entorn i els esdeveniments que es produeixen en un espai intel·ligent, així com les eines imprescindibles per a la interacció amb les persones.

Per tal de canviar el paradigma i centrar l'atenció en l'usuari, els ambients intel·ligents han de disposar d'informació relativa al qui, on, quan, què, com i perquè en l'entorn d'interacció per tal d'actuar en aquest entorn i interaccionar correctament amb les persones. Els senyals dels sensors àudio-visuals són processats per sistemes de computació distribuïts en una arquitectura d'agents amb diferents nivells de complexitat. Al nivell inferior, hi ha la infrastructura de xarxa i els fluxos de senyals d'àudio i vídeo que alimenten les anomenades components perceptuals. En un nivell intermedi, aquestes components perceptuals són els elements bàsics que detecten el qui, on, quan, què, com i perquè. Models particulars de l'entorn, les persones i de les relacions entre ells permeten inferir la situació que es produeix. Al nivell superior, l'anàlisi de la situació detectada faculta al sistema per decidir com proveir el servei adequat en cada moment.

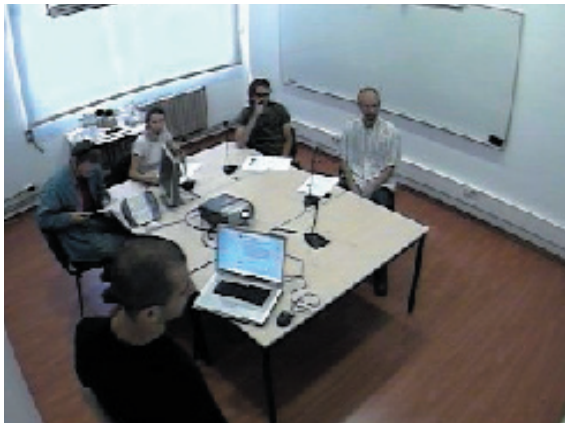

Figura 2. L'Espai Intel·ligent del Campus Nord de la Universitat Politècnica de Catalunya en configuració "sala de reunions"

CHIL desenvolupa prototips de serveis bàsics com a demostradors de la utilitat de l'anàlisi de l'entorn i de les interfícies multimodals. D'aquests serveis en podem destacar dos. El '*Memory Jog*' actua d'ajuda a la memòria, proveint informació pertinent de manera proactiva o reactiva. Per exemple, imaginem que en una reunió ens trobem amb una persona de qui no recordem el nom. El sistema pot proporcionar informació de forma automàtica de qui és aquesta persona. El '*Connector*' és un altre exemple de servei que ajuda a posar les persones en contacte a través del dispositiu adient en el moment adequat. Això evita la situació d'haver de fer múltiples trucades infructuoses i, sovint, inoportunes, per tal de trobar el moment per comunicar-se amb un interlocutor. El lector pot trobar més informació sobre els objectius d'aquests serveis a la pàgina del projecte CHIL. Els vídeos disponibles a l'apartat publicacions són força il·lustratius.

Smart Room/ Smart Classroom

El primer resultat visible del projecte CHIL a la UPC és la construcció d'un espai intel·ligent al Campus Nord (edifici D5). Està configurat com una sala de reunions, amb una taula central i cadires al voltant. Una de les parets és mòbil i permet convertir la sala en un aula amb el laboratori adjacent. S'hi ha instal·lat una xarxa de sensors àudio-visuals (càmeres i micròfons), equips de sincronització, i adquisició, una xarxa informàtica, ordinadors per al processament, projectors de vídeo, etc.

La visió de futur allunya l'ordinador del centre d'atenció sense afectar els serveis d'informació i comunicació que ens dóna.

La "Sala Intel·ligent" és una instal·lació imprescindible per als investigadors de la UPC que fan activitats de recerca en interfícies multimodals. Els senyals àudio-visuals adquirits permeten desenvolupar tècniques d'anàlisi dels senyals i experimentar amb demostradors dels serveis que es podrien oferir en les dues configuracions de sala de reunions i aula docent.

Anàlisi de l'activitat humana en espais intel·ligents

Els aspectes relacionats amb el llenguatge, com el reconeixement de la parla, són fonamentals en l'anàlisi de l'activitat humana per a la interacció en espais intel·ligents. Actualment, s'investiguen reconeixedors robustos de la parla amb

micròfons distants, de manera que no es destorbi a les persones fent-los portar a sobre cables i petaques. D'altra banda, les tecnologies visuals treballen en l'anàlisi de presència, localització i moviments de les persones, en el reconeixement de les cares, en la detecció de gestos, mirades i postures, en la detecció d'activitats, actituds i interaccions. Les tècniques de detecció, classificació i reconeixement basades en senyals de múltiples sensors —com ara localització visual i acústica, reconeixement de persones per la veu i la cara, o detecció d'activitat pel so i les imatges— prometen millorar la robustesa dels sistemes d'anàlisi actuals.

Si ens centrem en les tecnologies d'anàlisi visual, reconèixer l'activitat humana no resulta fàcil. Les persones oferim una imatge molt variable als sensors: el nostre cos és dinàmic, articulat i deformable, l'usem per a actuar en l'entorn, per a expressar-nos i interaccionar amb els altres, ens agrada cobrir-nos amb teixits diferents i objectes diversos i, sovint, solem aparèixer en grup més que no pas aïlladament, generant oclusions que dificulten la visió. Tot i els reptes esmentats, el processament de les imatges que "veuen" les càmeres situades en l'entorn, pot proporcionar al sistema informació rellevant per "entendre" l'escena.

El llenguatge del cos: gestos, mirades, postures i actituds

Com a exemple pràctic de l'anàlisi visual de l'escena, mostrarem com l'extracció de dades sobre la posició, actitud o gestos de les persones permet extraure informació d'alt nivell semàntic de l'entorn observat: des de saber si una persona es troba dreta o asseguda, fins a fer el recompte de vots (a mà alçada) en una votació o detectar l'activitat que s'està desenvolupant a la sala.

A la '*Smart Room*' de la UPC es genera una reconstrucció virtual 3D de l'escena a partir de les imatges de múltiples càmeres. L'anàlisi de la reconstrucció 3D permet detectar-hi persones (per diferenciar-les d'una cadira, per exemple) i analitzar la seva estructura ajustant-hi un model jeràrquic del cos humà. En funció de la complexitat del model, es poden obtenir informacions amb diferent nivell de detall sobre l'actitud postural de la persona (Figura 3), o es pot fer un anàlisi semàntic de més alt nivell per detectar interaccions entre individus, com ara la detecció on es concentra l'atenció dels assistents a una reunió en funció d'on està mirant cada persona (Figura 4).

Conclusió

És evident que encara som molt lluny del dia que podrem confondre el comportament d'una màquina amb el d'una persona; el dia que la màquina ens entendrà i es comportarà "naturalment". En el camí cap aquest dia, estem convençuts que s'avançarà cap l'objectiu de fer-nos la vida més fàcil i còmoda. Infrastructures com les sales intel·ligents permeten als investigadors treballar en tecnologies d'interfícies multimodals per als sistemes d'interacció natural. Sistemes que processen els senyals que "veuen" i "senten" els sensors, que "entenen" el seu entorn sense esperar què algú els digui quin és. Ordinadors-ajudants que trien adequadament els senyals que "mostren" i "diuen" a les persones. Sistemes, en definitiva, capaços de gestionar la complexitat de la interacció humana per donar resposta a les nostres necessitats d'informació i comunicació.

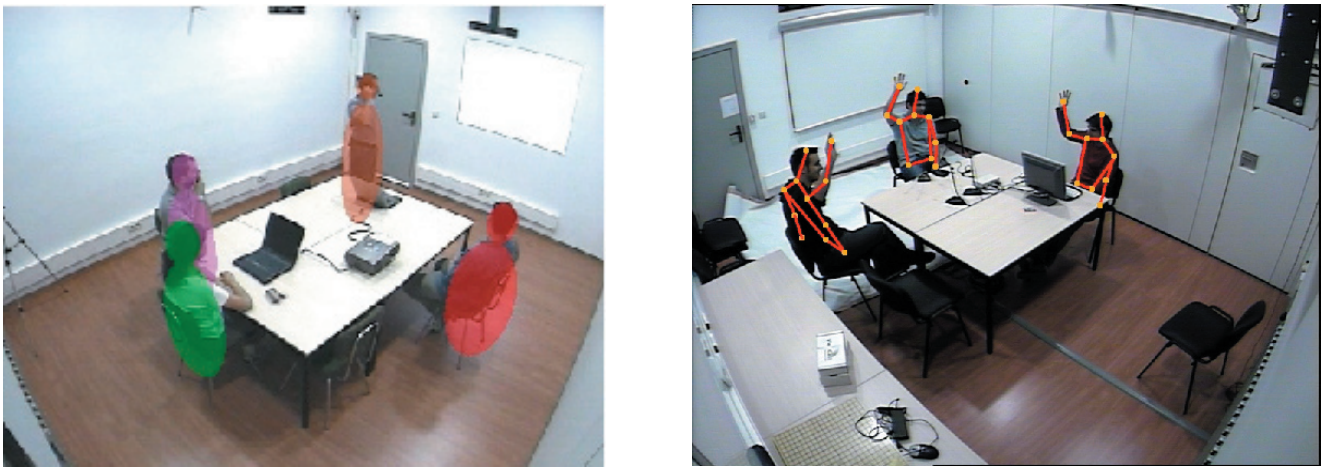

Figura 3: Un model simple del cos humà basat en dos el·lipsoids (imatge esquerra) permet detectar si una persona està asseguda (oients) o dreta (presentador). Un model més complex basat en nodes i segments (imatge dreta) permet analitzar si una persona aixeca el braç en una votació, permetent així al sistema fer un recompte dels vots.

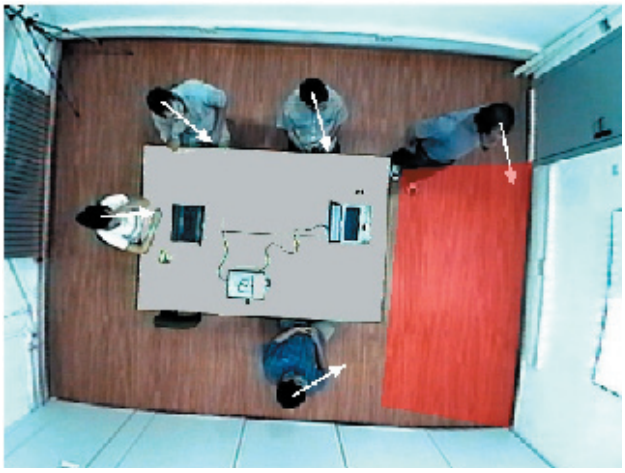

Figura 4. Vista zenital de la 'Smart Room' on s'ha representat el vector que indica la orientació del cap de les persones. A partir d'aquesta informació, es pot detectar on es concentra l'atenció dels assistents (àrea en vermell). En aquest cas, és la pantalla on es projecten les transparències.

Referències

Projecte CHIL: http://chil.server.de
The NIST Smart Space Laboratory Web Site: http://www.nist.gov/smartspace
MIT - Smart Rooms: http://vismod.media.mit.edu/vismod/demos/smartroom
CSIRO - Smart Spaces: http://www.smartspaces.csiro.au
TsingHua Smart Classroom: http://media.cs.tsinghua.edu.cn/~pervasive/introduce.html
Perception, recognition & integration for interactive environments: http://www-prima.imag.fr
Sunnys.edu - Smart Rooms: http://www.cs.sunysb.edu/~tony/392/smartrooms/smartrooms.html
Computerworld - Smart Rooms (August 4, 2003): http://www.computerworld.com