

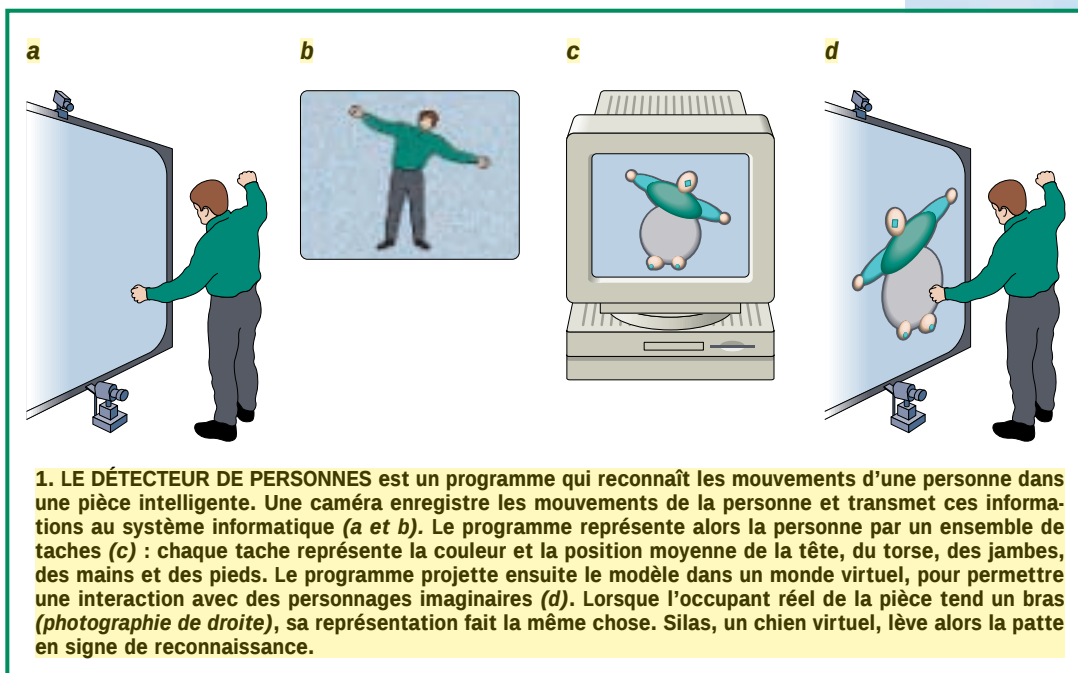
Une pièce douée de raison

ALEX PENTLAND

Afin d'élaborer des environnements domestiques et professionnels capables de nous aider dans nos tâches, les informaticiens étudient des systèmes informatiques qui identifient les occupants d'une pièce et les actions de ceux-ci.

Imaginez une maison qui surveillerait les enfants en l'absence des parents, un bureau qui déterminerait l'importance des réunions et qui éviterait les interruptions inappropriées, une voiture qui détecterait la fatigue de son conducteur et lui conseillerais de s'arrêter en cas de fatigue. Les informaticiens rêvent depuis longtemps à de tels systèmes, mais même les ordinateurs les plus perfectionnés ne peuvent remplacer les baby-sitters ou les secrétaires.

La difficulté tient à la surdité et à la cécité des ordinateurs : ils appréhendent le monde réel au travers d'un clavier ou d'une souris. Même les systèmes multimédias, qui traitent aussi bien les sons et les images que les textes, se contentent de transporter des données numériques. Ils ne comprennent pas la signification des signaux qu'ils transmettent. Les ordinateurs deviendraient bien plus utiles qu'ils ne le sont aujourd'hui s'ils étaient capables de voir et d'entendre. Et s'ils savaient identifier leur interlocuteur et deviner ses pensées...





2. SURPRIS, en colère, dégoûté, heureux, triste ou maussade? Irfan Essa a mis au point un système informatique qui reconnaît ces expressions faciales. Il a utilisé des données biomédicales pour modéliser la manière dont les tissus du visage bougent selon que l'on sourit, que l'on fronce les sourcils ou que l'on adopte une série d'autres expressions faciales (*ci-dessus*). Son système compare des images vidéo des utilisateurs de pièces intelligentes avec celles de modèles animés. Ici, la caméra a figé l'image du chercheur exprimant la surprise (*écran de gauche*) et le système informatique a choisi le modèle correspondant (*écran de droite*).



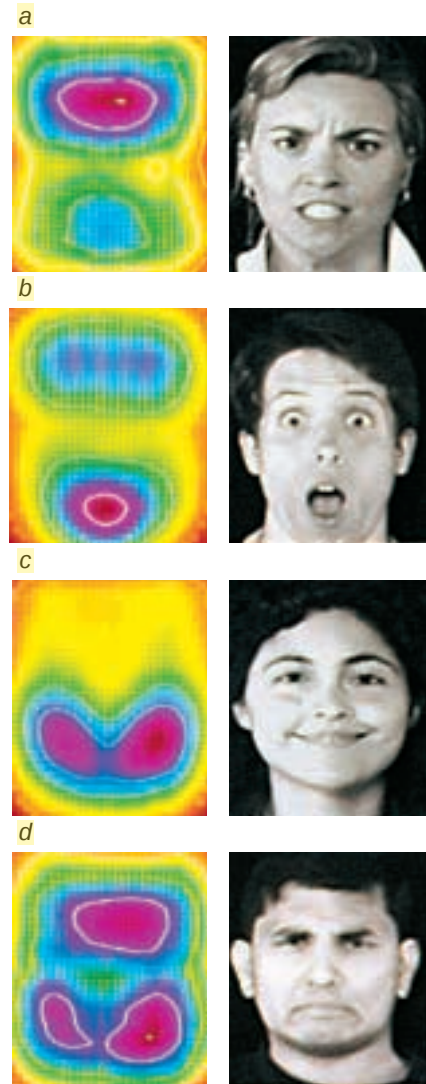
À l'Institut de technologie du Massachusetts, nous venons de mettre au point un ensemble de systèmes informatiques qui reconnaissent les visages, les expressions et les gestes. À l'aide de ces systèmes, nous avons réalisé des environnements qui s'approchent de la maison, du bureau ou de la voiture décrits précédemment. Ces «pièces intelligentes» sont équipées de caméras et de microphones dont les enregistrements sont transmis à un réseau d'ordinateurs. Ces derniers déterminent ce que font et disent les occupants de la pièce intelligente, qui peuvent ainsi utiliser leurs gestes, leurs voix et leurs expressions faciales, au lieu d'un clavier ou d'une souris, pour interagir avec des programmes informatiques, pour parcourir des bases de données multimédias ou pour voyager dans le royaume de la réalité virtuelle.

La première pièce intelligente fut construite dans notre laboratoire en 1991, et il existe aujourd'hui cinq pièces de ce genre à travers le monde, toutes reliées par le réseau téléphonique : trois à Boston, une au Japon et une au Royaume-Uni. Des installations similaires sont en projet à Paris, à New York et à Dallas.

Chaque pièce contient plusieurs micro-ordinateurs, chargés chacun d'une tâche spécifique. Quand une pièce intelligente doit analyser les images, les sons et les gestes, nous l'équiperons de trois ordinateurs, un par type d'interprétation. Si elle doit avoir des capacités supérieures, nous ajoutons des machines au réseau. Les ordinateurs utilisent tous la même méthode statistique, nommée analyse du maximum de vraisemblance : ils comparent les informations qu'ils reçoivent avec des modèles stockés en mémoire ; ils calculent la probabilité qu'un modèle corresponde aux données reçues, et choisissent le modèle le plus probable. Ces comparaisons permettent aux ordinateurs de répondre à une série de questions sur leurs utilisateurs, y compris déterminer leur identité et, parfois, ce qu'ils veulent.

Où?

Avant de chercher à savoir ce que font ses occupants, la pièce doit les localiser. Aussi avons-nous mis au point un détecteur de personnes qui suit les mouvements d'un individu à travers la pièce. Nous l'avons vu, ce détecteur utilise l'analyse du maximum de



3. DANS LES CARTES DE MOUVEMENTS, des couleurs chaudes montrent les parties du visage qui ont le plus bougé. Certaines expressions produisent des motifs de couleur typiques. Lorsque quelqu'un est fâché, par exemple, c'est surtout la périphérie des sourcils qui se modifie (a). La surprise se reconnaît aux sourcils et à la bouche (b), la joie s'exprime surtout par la bouche (c), et le dégoût par le visage tout entier (d).

vraisemblance. Tout d'abord, il crée un modèle de la personne enregistrée par la caméra sous forme de sept taches de couleur reliées (deux pour les mains, deux pour les pieds, et trois taches distinctes pour la tête, le torse et les jambes). Il décrit chacune des taches de deux manières : sous la forme d'une série de valeurs pour la couleur et l'emplacement de la tache ; et sous la forme d'une «carte de support», une liste qui indique les pixels de l'image qui appartiennent à la tache (les pixels sont les éléments d'une image, semblables aux points qui composent l'image de votre téléviseur). Ensuite, pour modéliser