

EMBARGO VENDREDI 4 SEPTEMBRE 2020 08H00 CET

Des algorithmes valaisans renseignent les clients d'un centre commercial finlandais

Les interactions des robots-humains manquent souvent de naturel, surtout en dehors des laboratoires. Aujourd'hui, les chercheurs de l'Idiap publient en libre accès les algorithmes qui ont permis d'utiliser un robot en conditions réelles dans un supermarché finlandais dans le cadre du projet européen Mummer.

Répondre aux questions de clients dans un centre commercial. Une activité banale et pourtant... pour qu'un robot puisse l'effectuer correctement, il doit pouvoir comprendre les questions qui lui sont posées, mais pas que. Afin de mener une conversation sans confusions, le robot doit être capable de détecter les personnes qui l'entourent, de distinguer celles qui s'intéressent à lui, celles qui parlent, séparer celles qui discutent entre elles de celles qui s'adressent à lui ou encore de vérifier si elles sont attentives. « Le but est de pouvoir développer des algorithmes qui analysent les signaux audio et vidéo captés par le robot pour en extraire les éléments de communication non-verbaux, » explique Jean-Marc Odobez, responsable du groupe de recherche Perception et Compréhension de l'Activité à l'Idiap. Permettre au robot d'interagir avec les humains dans des lieux du quotidien est un défi supplémentaire. C'est dans un centre commercial, à deux heures de route d'Helsinki, que chercheurs valaisans et européens ont testé et affiné leurs technologies. Grâce aux travaux de l'Idiap, l'expérience a montré que dans la vidéo captée par le robot, il est possible de reconnaître et de distinguer rapidement les interlocuteurs dans des conditions réelles d'éclairage.

Reconnaître un interlocuteur dans le temps et philosophie open source

« La confusion entre une personne et une autre et l'un des principaux écueils techniques à l'utilisation de robots dans l'espace public, où les interlocuteurs sont nombreux. Une seule erreur et le robot perd l'historique de la conversation pour recommencer l'interaction du début ou continue la conversation avec un autre, » détaille Jean-Marc Odobez. Pour évaluer et améliorer les performances du robot, les chercheurs ont enregistré de nombreuses interactions en conditions réelles. Ils ont ensuite créé une base de données où chaque interaction est décrite : qui regarde interlocuteur, à qui parle-t-il, etc. Finalement, les images des mêmes interactions ont été analysées par l'algorithme et le résultat a pu être comparé permettant d'affiner les algorithmes. Unique dans le domaine de l'interaction robot-humain, cette base de données ainsi que les algorithmes ont été mis à disposition de la communauté pour pouvoir évaluer les progrès futurs dans le domaine et font l'objet d'une publication et présentation dans le cadre de la conférence IEEE RO-MAN 2020 spécialisée dans ce domaine.

Des algorithmes multimodaux et qui apprennent avec moins d'informations

Le robot doit être capable de distinguer la voix humaine du bruit environnant et de la localiser. Ceci lui permet de tourner si nécessaire la tête vers la personne qui parle pour pouvoir ensuite mesurer la direction de son regard et déterminer si elle s'adresse à lui. La communication orale est donc essentiellement multimodale, impliquant bien plus que le seul contenu vocal. L'Idiap se spécialise en particulier dans l'analyse multimodale. « Grâce à l'utilisation combinée de la vision et de l'audition, la fluidité et la continuité des échanges est bien meilleure, » explique l'ingénieur recherche et

développement qui était en charge de suivre le robot à distance durant les trois mois de test grandeur nature en Finlande.

Une des prouesses de cet essai est liée au bon taux de réussite pour reconnaître et maintenir l'interaction avec les interlocuteurs. Pour y parvenir, les chercheurs de l'Idiap ont mis au point des techniques d'apprentissage avec des données dites faiblement étiquetées. « Habituellement, il faut fournir à l'algorithme un très grand nombre de données audio ou vidéos avec beaucoup d'informations associées pour que l'apprentissage automatique puisse se faire. Mais collecter ces informations automatiquement est difficile et leur annotation manuelle est fastidieuse et coûteuse, » précise Jean-Marc Odoñez. Réduire ces besoins en annotation est donc capital. Par exemple, pour apprendre à localiser d'où provient des sources sonores, il faut normalement enregistrer un ou plusieurs sons et annoter où se trouve chaque source. Le nouvel algorithme ne nécessite que d'indiquer le nombre de signaux sonores, pas leur localisation, ce qui est beaucoup plus aisé.

Plus d'informations

- Groupe de recherche Perception et Compréhension de l'Activité: <https://www.idiap.ch/en/scientific-research/perception-and-activity-understanding>
- Projet Mummer: <http://www.mummer-project.eu/>
- Conférence IEEE RO-MAN 2020 : <http://ro-man2020.unina.it/>
- Video illustrant le système de perception développé: <https://www.youtube.com/watch?v=Cfsc0zXAMVU>

Contact

Jean-Marc ODOBEZ, responsable du groupe de recherche Perception et Compréhension de l'Activité de l'Idiap Research Institute, odobez@idiap.ch, +41 27 721 77 26