Proposition de recherche doctorale

Détection et évaluation du stress dans la gestuelle dans le cadre d'échanges sociaux par l'analyse vidéo

Mots clés :

- **Directeur de thèse** : Severine Dubuisson
- **Co-encadrant(s)** :
- **Unité de recherche** : Laboratoire d'informatique de Paris 6
- **Ecole doctorale** : École Doctorale Informatique, Télécommunications, Électronique de Paris
- **Domaine scientifique principal**: Divers

Résumé du projet de recherche (Langue 1)

La communication homme-homme est un processus dynamique reposant sur l'échange et l'interprétation de signaux sociaux. Cette dynamique influence en particulier les tours de parole, l'engagement dans l'échange ou encore l'attention conjointe, et sa modélisation est identifiée comme un verrou majeur du traitement du signal social et de la robotique personnelle. Les approches proposées dans la littérature portent sur l'analyse de l'influence mutuelle des participants. Une interaction réussie se caractérise par une adaptation dynamique des comportements des interactants. Un des éléments jouant un rôle majeur dans la communication non verbale et donc dans l'adaptation mutuelle concerne tout mecanisme de compensation gestuelle liée aux émotions, telles que le stress. Dans cette thèse on s'intéresse à la détection et à l'estimation du stress dans la gestuelle d'une personne de manière non invasive par le traitement vidéo. En effet, si le stress d'un interlocuteur est perceptible dans sa voix (il existe en particulier de nombreux travaux en ayant étudié les caractéristiques pour y détecter du stress [2,3]) ou encore visible dans des signaux physiologiques [4,5], peu des travaux se sont interessés à l'étude et à l'analyse de la gestuelle et son lien avec le stress. Pourtant, recemment, des travaux [1] ont montré que l'analyse des gestes, en plus de celle de la voix, augmente significativement l'identification du stress. L'étude de la gestuelle comporte plusieurs avantages liés à sa nature directe, non invasive et universelle. D'une part, elle fait partie des signaux dits de communication non verbale (expression faciale, intonation de la parole, lien avec l'environnement, gestuelle et comportement), considérée comme étant plus fiable que la communication verbale car plus difficile à maîtriser. Par exemple, Zucherman et al. [6] ont montré que nous utilisons presque exclusivement les informations non verbales lorsque nous cherchons à déterminer si quelqu'un ment. D'autre part, l'étude de la gestuelle peut se faire à partir d'analyses vidéo, données acquises de manière moins intrusive que, par exemple, pour les signaux physiologiques acquis via des capteurs posés sur le corps. Enfin, la vidéo nous permet d'accèder aux comportements des personnes dans un très grand nombre d'environnements et de situations (i.e. de la vie de tous les jours), et en particulier quand la voix n'est pas le canal d'interaction privilegié, par exemple lors d'intéractions sociales coopératives homme-homme ou même homme-machine. Actuellement, il n'existe presqu'aucun travaux concernant la détection de stress dans la gestuelle, ou encore dans des cadres expérimentaux contraints, ce qui rend le sujet de cette thèse d'importance majeure pour la communauté scientifique travaillant sur le signal social ou, d'une manière plus générale, l'émotion.

Résumé du projet de recherche (Langue 2)

Le sujet de la thèse proposée concerne le développement d'un cadre théorique utilisant l'apprentissage artificiel pour la caractérisation et la modélisation de la dynamique du stress émis et induit. On s'intéressera aussi bien à la reconnaissance du stress à partir des gestes (méchanisme de compensation) qu'aux gestes induisant du stress chez l'interlocuteur. Plusieurs situations d'interaction seront envisagées et on préviligiera celles menant à des situations coopératives entre deux humains. En particulier, une première base de données dont nous disposons permettra de reconnaître et d'évaluer le stress d'une personne dans une situation de conduite routière, base pour laquelle nous disposons d'une vérité terrain (i.e. les images des séquences sont étiquetées selon que la personne est stréssée ou pas). Ce travail se décompose en plusieurs parties fortement liées les unes aux autres, qui sont détaillées ci-dessous. 1) Etude bibliographique des approches invasives et non invasives pour la détection de stress. Cette étude permettra de rendre compte de l'état de l'art actuel, et des techniques les plus abouties à ce jour pour la détection et l'analyse du stress chez l'être humain. On focalisera en particulier sur l'étude du stress né de l'échange social ou de situations de la vie de tous les jours. Ici le stress ne sera étudié que par sa caractéristion en tant que signal non verbal : expression faciale, intonation de la parole, lien avec l'environnement, gestuelle et comportement. 2) Extraction de descripteurs pertinents pour caractérisation de gestes normaux et de gestes stressés. Dans un premier temps, on identifiera les gestes traduisant la présence de stress chez un individu. On s'appuiera en particulier sur des analyses d'experts psychologues (Pitié-Salpétrière, LUTIN), mais aussi on pourra s'inspirer des analyses de signaux physiologiques ayant produit dans l'état de l'art des résultats fiables. Le stress de la gestuelle se caractérisant dans les images à la fois de manière spatiale et temporelle (l'amplitude du geste et sa latence, par exemple), mais étant également fortement lié au contexte (scène, objets de l'environnement, comportement du partenaire lors de l'échange, etc.), nous proposons de réfléchir à des descripteurs spatio-temporels et contextuels le caractérisant. On étudiera ensuite la fiabilité de leur extraction et leur caractère discriminant en testant plusieurs techniques de reconnaissance de gestes existantes [8]. Eventuellement, on étudiera l'apport de l'utilisation d'autres signaux non verbaux, tels que l'expression faciale comme information supplémentaire, en s'appuyant sur des travaux existants [7,9]. 3) Modélisation spatio-temporelle du stress. Cette partie est le coeur même de la thèse. Il s'agira ici de proposer un modèle de représentation spatio-temporelle du stress. On s'attachera en particulier à intégrer une continuité temporelle dans le modèle, qui nous permettra de répondre à des questions plus générales, concernant l'origine du stress (i.e. quel lien avec l'environnement a causé le stress ?) et ses conséquences (i.e. quelle interaction entre la personne stressée et son l'environnement a découlé du stress ?). Afin de proposer un modèle fiable et robuste, nous souhaitons intégrer de manière conjointe des informations concernant la gestuelle, mais également des connaissances extérieures, telles que des modèles psychologiques qui seront donnés par les praticiens psychologues. Concernant la gestuelle, nous envisageons d'étudier plus particulièrement des caractéristiques telles que l'amplitude des gestes, leur côté saccadé, leur latence ou encore leur vitesse d'execution. Pour cela, il faudra mettre en place une technique de suivi capable de prendre en compte des mouvements très rapides et erratiques, typiques de mouvement liés au stress. Le filtrage particulaire [10] sera très probablement l'outil utilisé, puisqu'il est (re)connu pour gérer des sytèmes à dynamique instables. En nous appuyant sur de précédents travaux [11,14], nous y intégrerons des connaissance extérieures qui permettront d'affiner le modèle de prédiction, et les observations correspondront aux descripteurs extraits des images de la séquence afin de rendre le suivi plus robuste. A partir des trajectoires extraites, nous pourrons calculer des caractéristiques liées à la gestuelle, telles que celles énumérées plus haut. Une étude comparative entre les caractéristiques calculées sur des trajectoires gestuelles stressées et celles que des trajectoires gestuelles non stressées permettra de mettre en avant des différences comportementales. De ces dfférences, sera proposé un modèle théorique de la situation dite de stress pour pouvoir ensuite mieux la caractériser et la reconnaître, et améliorer en particulier le processus de détection, comme c'est décrit ci-dessous. 4) Exploitation des caractéristiques extraites pour déterminer s'il y a présence de stress. Pour cette partie de prise de décision, il faudra prendre en compte les particularités de la communication non verbale : son ambiguité, sa continuité dans le temps et son aspect multi-canaux. C'est pourquoi on étudiera la possibilité d'intégrer ces connaissances, voire de modèles psychologiques, dans un algorithme apprentissage. Comme cadre formel pour représenter les connaissances, ici numeriques-symboliques, nous utiliserons la modélisation par logique floue. Elle nous sera utile, pour ce projet, à trois niveaux. D'une part elle nous permettra de gérer facilement et naturellement les imprécisions provenant des limitations dues à l'acquisition (i.e. séquences vidéo) et des erreurs inherentes au problème de suivi et à son analyse. D'autre part, cette modélisation nous permettra d'intégrer du raisonnement sous forme de règles expertes qui seront traitées par du raisonnement approximatif. Ainsi on pourra par exemple intégrer des connaissances sur les contraintes physiques du mouvement pour améliorer le suivi (qui est analysé sur une projection en deux dimensions). Finalement, en sachant qu'un des aspects intrinsèques du stress est la saturation par aggregation de stimulations, le cadre théorique proposé apporte un réel avantage. En effet, il est classiquement utilisé pour faire de la fusion guidée. Ici on s'intéressera en particulier aux aspects de renforcement de signaux lors de l'aggrégation. Ces travaux prendront comme point de départ les premiers résultats presentés dans [12]. Création d'une base de données. Comme nous l'avons dit, très peu de travaux aujourd'hui proposent un étude poussée du stress lors d'échanges sociaux. Ainsi, il n'existe aucun cadre formel permettant de valider une approche. En parallèle des travaux prévus dans le cadre de cette thèse énumérés ci-dessus, nous prévoyons une réflexion sur la création et la mise en place d'une base de données expérimentale, avec une attention particulière sur la communication non verbale, qui sera mise à disposition de la communauté scientifique. Concernant l'acquisition de cette base, il faudra tenir compte du caractère ambigü du stress et son aspect multimodal demande une profonde réflexion concernant les processus d'acquisition à choisir, les capteurs à utiliser (le capteur RGB-d de la Kinect et des micro synchronisés) et les scénarii à envisager d'intéraction diadiques. Enfin, il faudra fournir une vérité terrain et des outils de validation afin que les chercheurs puissent valider leur approche et la comparer à d'autres, à l'image de HumanEva (http://vision.cs.brown.edu/humaneva/) qui est proposée depuis 2007 pour tester des techniques de suivi d'objets articulées en environnement d'acquisitions multi-vues.

Informations complémentaires (Langue 2)

Références : [1] D. Giakoumis, Anastasios Drosou, P. Cipresso, D. Tzovaras, G. Hassapis, A. Gaggioli and G. Riva. "Using Activity-Related Behavioural Features towards More Effective Automatic Stress Detection." PloS one Journal, Vol. 7.9: e43571, 2012. [2] J. Tepperman and S. Narayanan. "Automatic syllable stress detection using prosodic features for pronunciation evaluation of language learners." IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP. Vol. 1. 2005. [3] M. Lai, Y. Chen, C. Min, Y. Zhao and F. Hu . "A hierarchical approach to automatic stress detection in English sentences." IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP, Vol. 1. IEEE, 2006. [4] J.A. Healey and R.W. Picard. "Detecting stress during real-world driving tasks using physiological sensors." IEEE Transactions on Intelligent Transportation Systems, Vol. 6.2, 2005. [5] J. Zhai and B. Armando. "Stress detection in computer users through non-invasive monitoring of physiological signals." Blood journal, Vol. 5, 2008. [6] M. Zuckerman, B.M. DePaulo and R. Rosenthal. "Verbal and nonverbal communication of deception." Advances in experimental social psychology, Vol. 14.1, 1981. [7] P. Ekman and W. V. Friesen. Manual on Facial action coding system, 1977. [8] S. Mitra and A. Tinku. "Gesture recognition: A survey." IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Vol. 37.3, 2007 [9] V. Bettadapura, "Face Expression Recognition and Analysis: The State of the Art", Tech Report, arXiv:1203.6722, April 2012. [10] N.J. Gordon, D.J. Salmond and A.F.M. Smith. “Novel approach to nonlinear/non-Gaussian Bayesian state estimation". IEEE Proceedings F on Radar and Signal Processing, 1993, 140 (2): 107–113. [11] E. Erdem, S. Dubuisson and I. Bloch. “Visual Tracking by Fusing Multiple Cues with Context-Sensitive Reliabilities.” Pattern Recognition, 2012, 45(5) :1948-1959. [12] Z. Gao, M. Detyniecki, M.-Y. Chen, A.G. Hauptmann, H.D. Wactlar and A. Cai. “The Application of Spatio-temporal Feature and Multi-Sensor in Home Medical Devices.” Journal of Digital Content Technology and its Applications, vol. 4(7): 69-78, 2010 [13] Z. Gao, M.-Y. Chen, M. Detyniecki, W. Wu, A.G. Hauptmann, H.D. Wactlar and A. Cai. “Multi-camera Monitoring of Infusion Pump Use.” IEEE International Conference on Semantic Computing (ICSC 2010) 2010: 105-111 [14] N. Widynski, S. Dubuisson and I. Bloch. “Fuzzy Spatial Constraints and Ranked Partitioned Sampling Approach for Multiple Object Tracking,” Computer Vision and Image Understanding, 2012, 116(10):1076–1094.