

# Joint Transform Correlator pour le suivi de visages : application à la détection des chutes de la personne âgée

Philippe KATZ, Michael ARON, Ayman ALFALOU

VISION\_L@BISEN, ISEN : Institut Supérieur de l'Électronique et du Numérique  
20 rue Cuirassé Bretagne, CS 42807, 29228 Brest Cedex 2, France

philippe.katz@isen.fr, michael.aron@isen.fr, ayman.al-falou@isen.fr

**Résumé** – Dans un souci de maintien (sécurisé) à domicile des personnes dépendantes (e.g. personnes âgées), nous proposons une approche de détection des chutes basée sur l'étude de la trajectoire de la tête. Notre système est composé de différents modules (suivi, identification, fusion de données, décision...). Dans cet article, nous présentons une optimisation du suivi, basé sur un JTC (Joint Transform Correlator) et une comparaison d'histogrammes. Des expérimentations préliminaires de détection des chutes ont été réalisées, validant le principe.

**Abstract** – Seeking for a solution for dependent people (e.g. elderly) to remain safely at home, we propose a fall detection approach based on a head trajectory analysis. Our system is made up of various modules (tracking identification, data fusion, decision...). In this paper, we present an optimisation of the tracking stage based on a JTC (Joint Transform Correlator) and a histogram comparison. Preliminary experiments of fall detection have been performed, leading to the principle validation.

## 1 Introduction

Du fait du vieillissement de la population dans les pays industrialisés, les infrastructures traditionnelles (e.g. maisons de retraite) sont dans l'incapacité de couvrir le besoin croissant de prise en charge. Ainsi, l'urgence d'imaginer de nouvelles solutions s'est faite ressentir et le concept d'habitat intelligent a émergé. La chute d'une personne dépendante isolée étant le premier cas de décès accidentel, les solutions actuellement commercialisées utilisent en majorité des données issues de capteurs portés (bouton poussoir actionné par la personne elle-même, accéléromètres...). Cependant, le fait de porter le système engendre des contraintes : incapacité d'actionner l'alerte, casse du matériel lors de la chute... Afin de pallier ces problèmes, et en coopération avec Malakoff-Médéric, nous avons décidé d'utiliser une caméra vidéo pour la détection des chutes. Ainsi, nous proposons d'effectuer un suivi de la tête de la personne pour en extraire sa trajectoire et en déduire la chute éventuelle. Un grand nombre de méthodes de détection vidéo existent dans la littérature [1]. Deux familles majeures se distinguent : les méthodes numériques (e.g. détection de points d'intérêt [2], soustraction de fond [3]) et les méthodes optiques (e.g. Joint Transform Correlator [4]). Malgré un essoufflement de la production scientifique sur les corrélateurs, ceux-ci ont l'avantage de réaliser simultanément la détection, la localisation et l'identification de l'objet cible dans une scène [4].

Ainsi nous présentons dans cet article un système de suivi itératif basé sur le Joint Transform Correlator (JTC) et couplé avec une méthode numérique [5] pour proposer un système fiable. Nous introduisons tout d'abord la méthode de corréla-

tion ainsi que notre algorithme de suivi par JTC. Puis nous présentons notre protocole d'expérimentation, permettant d'étudier l'impact des paramètres du JTC ainsi que les performances de suivi.

## 2 Détection des chutes par corrélation

### 2.1 Joint Transform Correlator : principe

Introduit par Weaver et Goodman [6], le JTC est une méthode de reconnaissance optique basée sur la comparaison d'une image cible (image à reconnaître) et d'une image de référence dans un unique plan d'entrée. Son implémentation numérique s'effectue de la façon suivante : (i) un plan de Fourier est construit à partir de la transformée de Fourier ( $TF$ ) du plan d'entrée ; (ii) le module au carré de ce plan est calculé ; (iii) la transformée de Fourier inverse ( $TF^{-1}$ ) donne finalement le plan de corrélation. Ce plan contient trois pics principaux : (i) un pic central, ou « ordre zéro », correspondant à la somme des autocorrélations de chacune des images présentes dans le plan d'entrée ; (ii) deux pics périphériques, positionnés symétriquement par rapport au pic central, correspondant à la corrélation croisée entre les images cible et référence. L'intensité de ces pics de corrélation est conditionnée par le degré de similarité entre les images d'entrée. Quant à leur position, elle est dépendante de la localisation relative des images sur le plan d'entrée. Ce sont ces deux propriétés fondamentales qui permettent la détection et la localisation du visage de la personne dans l'image. Malheureusement, dans sa version classique, ce corrélateur a deux principaux défauts : l'ordre zéro, extrêmement

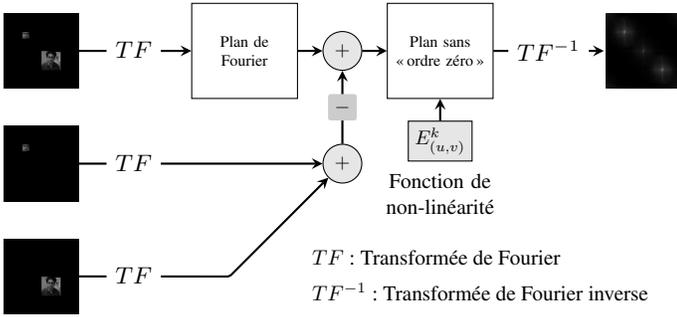


FIG. 1 – Joint Transform Correlator non-linéaire non-zéro.

intense par rapport aux pics de corrélation, et les pics de corrélation eux-mêmes, qui sont de faible intensité et très larges. La suppression de l'ordre zéro peut être effectuée en créant deux plans d'entrée supplémentaires, un pour chaque image contenue dans le plan d'entrée principal, et en déduisant leur spectre du plan de Fourier (Fig. 1).

Pour remédier au problème de faible intensité des pics de corrélation, Javidi et al. [7] proposent l'utilisation d'une fonction de non-linéarité dans le plan de Fourier, élevant ce dernier à une puissance  $k < 1$ . L'utilisation d'une valeur de  $k$  proche de 1 engendre de larges pics de corrélation de faible intensité, tandis qu'une valeur proche de 0 a l'effet inverse. Ainsi, le choix de la fonction de non-linéarité a un effet à la fois sur la robustesse du JTC et la précision de la localisation des pics. Il est donc nécessaire de déterminer expérimentalement un compromis entre ces deux propriétés. Cette étape de non-linéarité est appliquée après la suppression des spectres des images d'entrée (Fig. 1).

## 2.2 Application au suivi de visages

Notre algorithme est basé sur une méthode itérative : la région détectée dans l'image cible au temps  $t$  est utilisée comme image de référence au temps  $t + 1$ , la corrélation nous permettant d'effectuer dans un même processus la détection et le suivi. Le non-zero non-linéaire JTC (nz-nl-JTC) est initialisé par une image de référence, positionnée en haut à gauche du plan d'entrée. L'image cible est insérée sur le côté opposé. La position de l'objet dans l'image cible est obtenue à l'aide de la position relative des pics de corrélation. Cette position est finalement utilisée pour former l'image de référence pour l'itération suivante. Pour notre application de détection de chutes, nous proposons de choisir comme image de référence le visage de la personne. Notre méthode itérative est initialisée sur la première image de référence en utilisant un classifieur de Viola et Jones [8].

## 2.3 Optimisation du suivi

Malgré l'utilisation d'une fonction de non-linéarité, la faiblesse de l'intensité du pic de corrélation subsiste en cas de faible ressemblance de l'objet suivi entre les deux images (e.g. flou généré par une chute rapide). De plus, l'utilisation d'un

suivi itératif rend définitive toute perte de l'objet suivi (e.g. une corrélation avec l'arrière-plan de l'image cible). Finalement, pour être en mesure d'effectuer le suivi en temps réel (i.e. entre deux captures d'images), il est nécessaire de réduire la taille du plan d'entrée. Pour pallier à cela, nous proposons deux optimisations : l'une basée sur une information déterministe (région d'intérêt), l'autre sur une comparaison d'histogrammes.

### 2.3.1 Définition d'une région d'intérêt

Nous introduisons une première optimisation consistant à limiter la recherche de l'objet suivi à une région donnée. Pour ce faire, nous définissons dans l'image cible une région d'intérêt autour de la position du visage au temps  $t - 1$ . Notre algorithme étant destiné à une application de détection de chutes, les mouvements ascendants sont par conséquent limités. Nous choisissons de conserver une plus vaste région sous le visage. La taille optimale de cette région, dépendante de la fréquence de rafraîchissement de la caméra, a été déterminée expérimentalement (trois fois la région du visage, décalée d'un tiers vers le bas de l'image).

### 2.3.2 Mesure de la similarité des histogrammes

Comme explicité précédemment, leur incapacité à détecter une perte du suivi constitue le principal défaut des algorithmes itératifs. Pour remédier à cela, il est nécessaire d'introduire un critère supplémentaire. Ainsi, nous proposons l'utilisation de l'information contenue dans les histogrammes de l'image référence et de la région détectée. Un histogramme mesure la répartition de l'intensité des pixels : deux régions similaires visuellement auront donc des histogrammes semblables. Par conséquent, nous quantifions la similarité entre les images à l'aide du Chi Square de Pearson [9], défini par l'équation 1 :

$$X^2(H_{t-1}, H_t) = \sum_I \frac{(H_{t-1}(I) - H_t(I))^2}{H_{t-1}(I)} \quad (1)$$

- $H_{t-1}$  and  $H_t$  correspondent aux histogrammes des régions détectées aux temps  $t$  et  $t - 1$ , respectivement,
- $I$  correspond aux classes de l'histogramme.

Concrètement, cette mesure de similarité permet de détecter une perte du suivi et de réinitialiser notre algorithme.

## 2.4 Application à la détection des chutes

Il est nécessaire de déterminer un critère caractérisant une chute. Une chute brutale peut-être définie par le passage dans un faible laps de temps d'une position debout à une position allongée. Ainsi, il est possible à l'aide de la vitesse verticale du visage de définir un seuil à partir duquel une chute est considérée comme ayant eu lieu. Malheureusement, ce type de critère ne prend pas en compte les chutes suivant un mouvement elliptique (de vitesse verticale plus faible). La formule décrite par l'équation 2 permet de prendre en compte à la fois un mouvement vertical et elliptique en pondérant le mouvement horizontal par un facteur  $1/4$  (déterminé expérimentalement).

$$(y_t - y_{t-1}) + \frac{1}{4} \times |x_t - x_{t-1}| > Seuil \quad (2)$$

–  $x_t$  et  $y_t$  sont les coordonnées du visage dans l'image au temps  $t$ .

Ce critère simple a été utilisé pour valider notre approche, mais nous prévoyons d'expérimenter d'autres critères dans de futurs travaux.

### 3 Expérimentations et résultats

Dans cette partie, nous décrivons tout d'abord les performances de notre système de suivi par rapport au JTC classique, ainsi que les apports de la correction par histogrammes (partie 3.2). L'application de cet algorithme à la détection des chutes est évalué en partie 3.3.

#### 3.1 Protocole expérimental

Afin d'expérimenter notre algorithme, nous avons défini un protocole de tests et aménagé une salle de simulation (Fig. 2a) reproduisant une chambre d'hôpital ou de maison de retraite. Pour valider notre approche de détection des chutes, nous avons réalisé deux bases de données explorant les performances du suivi pour la première et de la détection de chutes pour la seconde.

Nous avons imaginé une large variété de scénarios composés de 14 différents événements (i.e. aucun mouvement, position debout, rotations du visage, translations, occlusions ou hors champ) de différentes vitesses ou amplitudes (Fig. 2b). La base de données, composée de 21087 images, a été enregistrée sur un total de 5 personnes de différentes origines ethniques. Finalement, l'emplacement du visage a été indiqué manuellement sur chaque image afin d'obtenir une « vérité terrain ».



FIG. 2 – Protocole expérimental : (a) chambre de simulation ; (b) événements d'expérimentation (chutes et occlusions).

Une seconde base de données a été réalisée pour déterminer la capacité de détection des chutes du système. Elle est composée de 60 chutes – 20 chutes verticales (descente verticale du visage dans l'image) et 40 chutes elliptiques (dans les 2 directions). Quatre personnes ont participé à son enregistrement.

#### 3.2 Performances du suivi JTC

Afin d'optimiser les performances du suivi par JTC, nous avons observé l'effet des différents paramètres disponibles dans notre environnement de test. Les meilleurs résultats sont obtenus avec : une décimation  $d = 2$  (une image sur deux est prise en compte), un coefficient de non-linéarité  $k = 0,4$ , une taille de la région d'intérêt  $s_{roi} = 3 \times s$  ( $s$  la région dans laquelle se situe le visage) et une taille de plan de corrélation  $s_{corr} = (512 \times 512)$ . Une étude détaillée est disponible dans [5].

Le tableau 1 présente une comparaison des résultats du suivi avec la « vérité terrain », obtenus en appliquant notre algorithme sur la base de 21 087 images (partie 3.1). Il présente les pourcentages d'images avec et sans suivi pour le suivi JTC classique ainsi que le nz-nl-JTC avec et sans correction par histogrammes (le critère de similarité 2.3.2 utilisé est  $X^2 > 100$ ).

TAB. 1 – Comparaison entre le JTC classique et le nz-nl-JTC avec région d'intérêt (ROI) et avec et sans correction par histogrammes (Hist) –  $s_{corr} = (512, 512)px$ ,  $s_{roi} = (3 \times s)px$ ,  $s = (78, 78)px$  and  $k = 0.4$ .

Suivi	JTC	nz-nl-JTC	
		ROI	ROI+Hist
Images avec suivi (%)	15,35	58,11	81,46
Images sans suivi (%)	84,65	41,89	18,54

On observe une amélioration significative du pourcentage d'images suivies avec l'utilisation du nl-nz-JTC par rapport au JTC classique, à savoir une augmentation de 42,76pts du nombre d'images suivies. Malgré tout, le taux de 58,11% d'images suivies reste insuffisant pour notre application. Ce taux est amélioré de 23,35pts par la correction par histogrammes, permettant d'obtenir un taux de 81,46% d'images suivies. Cependant, des pertes de suivi persistent, principalement lors de mouvements rapides ou d'occlusions [5], engendrant des limitations lors de l'étape suivante de détection des chutes.

#### 3.3 Détection des chutes

Après avoir optimisé le suivi, nous observons les performances de notre critère de décision (décrit en 2.4). Le seuil est fixé expérimentalement à 450px sur un intervalle de 1,34s. Les résultats présentés sur le tableau 2 ont été réalisés sur l'ensemble de 60 chutes verticales et latérales (partie 3.1). Le nombre et le pourcentage de chutes détectées sont représentés sur les lignes 2 et 3. On observe un taux de 57,5% et de 60% de chutes elliptique et verticales reconnues, respectivement. Nous obtenons un total de 58,34% de chutes détectées.

Ce faible taux de détection est dû à une baisse des performances du suivi lors de mouvements rapides. De plus, l'absence d'information de profondeur permet une mesure de la vitesse uniquement à une distance fixe de la caméra.

TAB. 2 – Nombre et pourcentage de chutes (direction verticale, gauche, et droite) détectées par notre système (20 chutes simulées pour chaque situation).

Direction	Verticale	Gauche	Droite	TOTAL
Chutes détectées #	13	10	12	35
Chutes détectées %	65	50	60	58,34
		57,5		

## 4 Discussion

L'utilisation d'une mesure de similarité des histogrammes améliore significativement les performances du suivi par nzl-JTC, permettant d'atteindre un taux de 86,41% d'images suivies.

L'étape de détection de chutes correspond à une première expérimentation et nécessite des travaux plus étendus. Ses déficiences sont dues à la fois à l'étape de suivi, malheureusement insuffisante, et au critère de détection de la chute utilisé. En effet, différentes lacunes de notre algorithme persistent pour une telle application.

Premièrement, lors d'une perte de suivi au cours de la chute, l'étape d'initialisation du suivi ne permet pas d'obtenir un taux de détection du visage acceptable. De plus, des mouvements rapides engendrent un flou sur l'image capturée, perturbant la corrélation. Ces deux effets combinés augmentent à la fois la probabilité de décrochage du système tout en réduisant sa capacité à se réinitialiser.

Deuxièmement, le critère de détection n'est adapté qu'aux chutes brutales. Les chutes molles ou syncopales (le sujet se retient à un meuble ou perd connaissance) ne peuvent être détectées que par une méthode décision beaucoup plus développée, basée sur une analyse approfondie de la chute.

Finalement, des déplacements dans la profondeur influent à la fois sur le critère de détection et sur le JTC. La corrélation étant très sensible au changement d'échelle, la probabilité de perte de suivi en est d'autant plus augmentée. En outre, notre critère de détection ne peut prendre en compte des chutes dans cette direction.

Des améliorations peuvent être apportées. Nous développons un système multicaméra afin d'obtenir les informations de profondeur et de détecter la personne dans l'ensemble de la pièce. Également, notre méthode nécessite que le visage soit continuellement visible par la caméra [5]. Une détection du squelette améliorerait notre système, afin de connaître la posture de la personne dans l'environnement.

## 5 Conclusion et perspectives

Cet article présente une étude de l'architecture JTC dans une application de détection de chutes de la personne âgée. Pour optimiser le suivi, deux améliorations de l'algorithme ont été apportées, à savoir la définition d'une région d'intérêt ainsi qu'un critère de similarité permettant une réinitialisation en cas

d'échec du suivi.

En ce qui concerne notre méthode de suivi, bien que présentant des résultats prometteurs, elle doit être améliorée afin de prendre en compte la profondeur – une implantation multicaméra (stéréovision) est actuellement en cours. De plus, elle doit être en mesure d'offrir des informations de position et de posture de la personne lorsque le visage n'est plus dans le champ de vision de la caméra. Pour cela, nous proposons de fusionner notre algorithme avec des méthodes de détection de silhouette ou de squelette.

Notre critère de détection des chutes fait actuellement l'objet d'une étude plus approfondie, afin d'être à même de caractériser l'ensemble des différents types de chutes (chutes brutales, molles et syncopales).

Pour terminer, la réalisation d'une base de données étendue doit être réalisée pour expérimenter plus précisément notre algorithme ainsi que pour le comparer avec des méthodes existantes [10].

**Remerciements** Ces travaux sont supportés par le groupe Malakoff-Médéric et la société Open (Projet 00R251).

## Références

- [1] A. Yilmaz, O. Javed and M. Shah, Object tracking : A survey, *ACM Comput. Surv.*, 38, 2006
- [2] D. Lowe, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vision*, 60, 91–110, 2004
- [3] A. Monnet, A. Mittal, N. Paragios and V. Ramesh, Background modeling and subtraction of dynamic, *IEEE ICCV*, 1305–1312, 2003
- [4] A. Alfalou and C. Brosseau, Understanding correlation techniques for face recognition : from basis to application, *Face Recognition*, M. Oravec (Ed.), 353–380, 2010
- [5] P. Katz, M. Aron and A. Alfalou, Joint Transform Correlation for face tracking : elderly fall detection application, *SPIE 8748, Optical Pattern Recognition XXIV*, 87480I–14, 2013
- [6] C. Weaver and J. W. Goodman, A technique for optically convolving two functions, *Appl. Opt.*, 5(7), 1248–1249, 1966
- [7] B. Javidi, J. Wang and Q. Tang. Nonlinear joint transform correlators. *Pattern recognition*, 27(4), 523–542, 1994
- [8] P. Viola and M. Jones, Rapid object detection using a boosted cascade of simple features, *IEEE CVPR*, 1, 511–518, 2001
- [9] O. Pele and M. Werman, The quadratic-chi histogram distance family, *Computer Vision ECCV*, 6312, 749–762, 2010
- [10] C. Rougier, A. St-Arnaud, J. Rousseau and J. Meunier, Video surveillance for fall detection, *Video Surveillance*, Prof. Weiyao Lin (Ed.), InTech, 2011