

L'AFEC (*Automatic facial expression recognition*) a suscité un grand intérêt au cours des deux dernières décennies. Cependant, l'analyse des expressions faciales est une tâche très difficile, car les expressions faciales causées par les mouvements des muscles faciaux sont subtils et transitoires. Une méthode basée sur la géométrie géométrique saisit les configurations faciales dans lesquelles un ensemble de points de repère faciaux est utilisé pour caractériser la forme du visage : Histogrammes de Gradients Orientés (HOG) à la modélisation temporelle et trois plans orthogonaux (TOP). Le but ici est de développer un cadre qui peut s'attaquer efficacement à la reconnaissance des expressions faciales dans les vidéos. Nous proposons un nouveau descripteur de caractéristiques HOG-TOP pour caractériser les changements d'apparence du visage et montrer que plusieurs caractéristiques peuvent obtenir de meilleures performances que les caractéristiques appliquées seules.

Le HOG-TOP est basé sur le LBP-TOP, mais plus orienté changement d'expression facial avec de meilleures fonctionnalités géométrique et plus compact. Le HOG TOP est un HOG possédant 3 plans orthogonaux X, Y et T (le temps). Le plan XY fournit l'aspect physique alors que XT et YT enregistre le temps et le mouvement dans le temps. Les calculs d'orientation de gradients se font donc sur les 3 plans, et sont ensuite regroupés pour n'en former plus qu'un qui est une description globale avec les caractéristiques spatiales et temporelles combinées. De plus, HOG-TOP calcule les gradients orientés de chaque pixel, ce qui le rend plus efficace aux déformations, mais il s'opère également sur des blocs de pixels plus petits, ce qui le rend plus précis. L'image est *croquée* et redimensionnée en 128x128 pixels puis par 8x8 blocks de 16x16 pixels contenant des angles allant de 0° à 180° pour enfin sortir 9 images.

L'expression faciale peut être considérée comme un processus dynamique comprenant l'apparition, le pic et le décalage. Pour trouver le changement Apparition-Pic de l'émotion, on va utiliser un maillage en triangle (dérivé du *warp transform*) sur la face du sujet. Lors d'une face neutre, chaque pixel de l'image appartient à un maillage (*triangle*) précis. En ayant cette information, nous savons que lors d'une expression, il y aura des déformations et les maillages vont bouger, il suffit donc de détecter si un pixel change de maillage pour détecter un mouvement et son type d'émotion associé.

On utilise également l'audio qui peut nous aider à valider ou non une hypothèse (ex : un son de rire sur une détection de tristesse) dans une dimension temporelle.

Pour l'apprentissage, On décide de créer plusieurs noyaux de SVM (*Support Vecteur Machine*) car plus performant, mais également car un SVM à noyaux multiples est conçu pour apprendre à la fois les frontières de décision entre les données de différentes classes mais aussi les poids de combinaison des noyaux par le biais d'un seul problème d'optimisation.

En utilisant 3 Datasets de différents, le modèle a pu s'entraîner sur 1060 séquences et être testé et validé sur 924 séquences. Pour les résultats expérimentaux, le HOG-TOP est similaire au LBP-TOP mais est meilleur de 5% sur le dataset plus réaliste (*wild*) en plus d'être deux fois plus rapide à l'exécution et de prendre 39% moins de place. En revanche, lorsque l'on a la combinaison optimale de la fonction HOG-TOP et de la fonction de déformation géométrique (appelé *Hybrid Feature ID*), le résultat est légèrement meilleur que les meilleures techniques utilisées jusqu'ici pour tous les datasets ce qui promet de nouvelles avancées dans ce domaine.