

SYNTHESE DE L'ARTICLE

L'objectif de cet article est de décrire le processus de détection et de reconnaissance des visages en temps réel à l'aide d'un réseau neuronal convolutif (CNN), en utilisant une caméra sur un Raspberry Pi module B.

Outils : Raspberry Pi, module caméra Pi NoIR ,GPU NVIDIA GeForce GTX 780

Bibliothèques : NumPy, OpenCV, Theano

Ils ont utilisé 40 images par personne (6 personnes) et les ont prises dans différents environnements (intérieur et extérieur).

Les images prises étant de taille différente, ils ont dû les remodeler et utiliser LBP pour réduire la dépendance à l'éclairage. Ils ont réduit la dimension des données en les regroupant afin d'éviter les surajustements. Pour manipuler toutes les images, ils ont utilisé la taille de 46*46 pixels en faisant CNN sur faire la classification et la régression pour détecter et identifier la personne. Pour entraîner le modèle, ils ont utilisé la méthode de descente de gradient pour réduire l'échelle de la fonction de perte afin que notre algorithme puisse avoir de bonnes performances. Ils ont utilisé des images extérieures pour le test et ont obtenu des performances qui varient de 88% à 99%. Le modèle a ensuite été transféré sur un Raspberry Pi.

Ils ont utilisé différentes méthodes :

Method	Accuracy	Classification Time [msec]	Enrollment Rate N_e/N	Total Time per Face [msec]
CNN ($p_0=0.85$)	99.59%	105 ± 8	250 / 278	529 ± 64
CNN ($p_0 = 0.0$)	97.48%	105 ± 8	278 / 278	529 ± 64
Fisherfaces (no al.)	88.50%	54 ± 11	278 / 278	511 ± 89
Fisherfaces (al.)	96.87%	535 ± 89	192 / 278	1006 ± 118

Précision et performance pour différentes approches. Le temps de classification inclut également le temps de traitement propre.

Enfin, ces auteurs ont tiré les conclusions suivantes :

1. Comme référence, ils ont choisi différents classificateurs OpenCV, et la meilleure performance a été obtenue en utilisant le classificateur Fisherface.
2. Sans aucun prétraitement, le CNN a obtenu un taux de reconnaissance initial de 24%. L'application de l'opérateur LBP a réduit la dépendance à l'illumination, augmentant le taux de reconnaissance à 82%.

3. Cette précision a pu être augmentée encore plus en rejetant les images pour lesquelles la meilleure estimation, retournée par la régression, n'était pas suffisante.

4. Une analyse temporelle sur le Raspberry Pi a révélé que l'algorithme de Viola-Jones nécessitait 423 ± 64 ms pour détecter un visage dans une image donnée. Le CNN nécessite 105 ± 8 ms supplémentaires pour la reconnaissance des visages, tandis que l'approche Fisherface nécessite 535 ± 89 ms pour la même tâche.

En conclusion, cette approche fournit des améliorations significatives en termes de vitesse et de précision par rapport à OpenCV, et est adaptée aux matériels à ressources limitées.

Schema bloc du fonctionnement global :

