



HAL
open science

Les Descripteurs de covariance profonds pour la reconnaissance des expressions faciales

Naima Otberdout, Mohamed Daoudi, Lahoucine Ballihi

► **To cite this version:**

Naima Otberdout, Mohamed Daoudi, Lahoucine Ballihi. Les Descripteurs de covariance profonds pour la reconnaissance des expressions faciales. COMpression et REprésentation des Signaux Audiovisuels (CORESA), Nov 2021, Sophia Antipolis, France. hal-03425705

HAL Id: hal-03425705

<https://hal.science/hal-03425705>

Submitted on 11 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Les Descripteurs de covariance profonds pour la reconnaissance des expressions faciales

Naima Otberdout , Mohamed Daoudi, Lahoucine Ballihi

Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France

IMT Lille Douai, Univ. Lille, CNRS, UMR 9189 CRIStAL, F-59000 Lille, France

LRIT - CNRST URAC 29, Univ. Mohammed V de Rabat, Faculté des Sciences, Rabat, Maroc.

Résumé : *Dans cet article, nous proposons une nouvelle approche exploitant la puissance des réseaux de neurones profonds pour encoder les caractéristiques non-linéaires des expressions faciales d'un visage et la puissance des matrices de covariance pour encoder les relations entre ces caractéristiques. Notre approche est basée sur l'idée d'encoder les cartes de caractéristiques locales et globales de la couche la plus profonde d'un réseau neuronal convolutif (DCNN) extraites des images fixes dans des matrices de covariance. Ces matrices de covariance sont symétriques définies positives (SPD). En effectuant la classification des expressions faciales en utilisant le noyau Gaussien sur la variété des matrices SPD, nous montrons que notre approche donne des résultats de classification meilleure que celle qui utilise les couches entièrement connectées et softmax. En effectuant des expérimentations sur trois bases de données des expressions faciales (Oulu-CASIA, CK+, SFEW), nous montrons que l'approche proposée atteint des performances compétitives par rapport à l'état de l'art.*

Mots-clés : Apprentissage profond, matrices de covariance, matrices symétriques définies positives, reconnaissance des expressions faciales.

1 Introduction

L'analyse automatique des expressions faciales est un sujet de recherche très important en vision par ordinateur depuis longtemps en raison de ses nombreuses applications potentielles qui vont de l'interaction homme-machine aux investigations médicales et psychologiques. Comme plusieurs autres applications de la vision par ordinateur les méthodes d'apprentissage profond ont beaucoup fait avancer ce problème. Suite à cette motivation, plusieurs recherches ont proposé d'utiliser les réseaux de neurones profonds (Deep Convolutional Neural Networks DCNN) pour la reconnaissance des expressions faciales [2, 1, 4, ?]. Toutes ces méthodes utilisent une stratégie similaire dans l'architecture de réseau : plusieurs couches convolutionnelles pour l'extraction des caractéristiques ; des couches entièrement connectées avec une couche softmax sont utilisées pour la classification. Cependant, le pipeline de classification des DCNN est basé sur des informations globales et ne prend pas en compte les relations spatiales à l'intérieur du visage. Dans une autre direction, d'autres recherches ont utilisé les descripteurs de covariance pour modéliser les corrélations entre les parties du visage. Ces descripteurs locaux ont montré leur efficacité pour la recon-

naissance des expressions faciales. La question est donc, comment est-il possible d'encoder des caractéristiques profondes locales dans une représentation compacte et discriminative pour une classification plus efficace que celle obtenue globalement par la couche softmax classique ? Dans de ce papier nous proposons une nouvelle représentation des expressions faciales qui encode les caractéristiques DCNN globales et locales par des matrices de covariance et les classer à l'aide d'un noyau Gaussien sur la variété des matrices SPD.

2 Travail proposé

Dans ce travail, nous proposons de construire des matrices de covariance des caractéristiques obtenues au niveau des couches profondes pour la reconnaissance des expressions faciales à partir des images. L'idée consiste à encoder les cartes de caractéristiques globales et locales dans une représentation compacte et discriminative. L'utilisation des matrices de covariance avec les cartes de caractéristiques des couches profondes permet de construire une représentation qui combine les caractéristiques de premier et deuxième ordre. En outre, cette méthode permet de combiner l'information locale et globale du visage en construisant des descripteurs locaux sur des régions spécifiques sur le visage (e.x, les yeux, les joues,...). Un modèle DCNN entraîné pour la reconnaissance des expressions faciales est capable de déterminer automatiquement les caractéristiques pertinentes spécifiques à chaque expression. Dans l'approche générale, la classification de ces caractéristiques est effectuée en utilisant des couches totalement connectées pour vectoriser ces caractéristiques, puis une couche softmax est ajoutée pour obtenir une probabilité pour chaque expression faciale. En revanche, dans ce travail, nous proposons d'exploiter les corrélations entre les cartes de caractéristiques des expressions faciales des visages extraites de la dernière couche de convolution en les agrégeant dans des matrices de covariance locales et globales. De ce fait, nous obtenons des représentations locales ainsi que globales compactes, robustes et efficaces. Ces matrices de covariances vivent dans la variété des matrices semi-définies positives, une variété non-linéaire. Nous avons classé ces descripteurs en utilisant les machines à vecteurs de support (SVM) avec un noyau Gaussien défini sur la variété SPD.

3 Résultats

Nous avons effectué les expérimentations sur trois bases de données des expressions faciales ((Oulu-CASIA, CK+ et SFEW) en utilisant deux modèles DCNN (VGG-face et ExpNet). Les résultats ont montré l'efficacité de l'approche proposée. En fait, nos résultats ont surpassé la classification avec la couche softmax classique. En plus, nous avons obtenu des résultats compétitifs en comparaison avec l'état de l'art.

Dataset	FC-Softmax	G-FMs	G+R-FMs
Oulu-CASIA	82.29	83.55	84.80
CK+	90.38	97.07	98.40
SFEW	48.26	49.18	49.18

TABLE 1 – Comparaison des méthodes proposées ; Global(G-FMs) et Global-Local (G+R-FMs)) avec la classification standard en utilisant les couches totalement connectées et softmax.

4 Conclusion et perspectives

Dans ce travail, nous avons proposé une nouvelle représentation compacte et discriminative pour la reconnaissance des expressions faciales. Cette représentation consiste à encoder les caractéristiques profondes du visage dans des matrices de covariance, matrices symétriques définies positives. Nous avons proposé une méthode de classification qui permet de les classer en respectant la géométrie de la variété des matrices SPD.

Références

- [1] Hui Ding, Shaohua Kevin Zhou, and Rama Chellappa. Facenet2expnet : Regularizing a deep face recognition net for expression recognition. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pages 118–126. IEEE, 2017.
- [2] Anis Kacem, Mohamed Daoudi, Boulbaba Ben Amor, and Juan Carlos Alvarez-Paiva. A novel space-time representation on the positive semidefinite cone for facial expression recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3180–3189, 2017.
- [3] Ikechukwu Ofodile, Kaustubh Kulkarni, Ciprian Adrian Corneanu, Sergio Escalera, Xavier Baro, Sylwia Hyniewska, Juri Allik, and Gholamreza Anbarjafari. Automatic recognition of deceptive facial expressions of emotion. *arXiv preprint arXiv :1707.04061*, 2017.
- [4] Zhiding Yu and Cha Zhang. Image based static facial expression recognition with multiple deep network learning. In *Proceedings of the 2015 ACM on international conference on multimodal interaction*, pages 435–442, 2015.
- [5] Xiangyun Zhao, Xiaodan Liang, Luoqi Liu, Teng Li, Yugang Han, Nuno Vasconcelos, and Shuicheng Yan. Peak-piloted deep network for facial expression recognition. In *European conference on computer vision*, pages 425–442. Springer, 2016.

Méthod	Kacem et al.[2]	Ding et al[1]	Yu et al.[4]	Zhao et al.[5]	Ofodil et al.[3]	G-FMs	G+ R-FMs
Oulu-CASIA	83.13	82.2	-	84.59	89.60	83.55	84.80
CK+	96.87	98.60	-	97.30	98.70	97.25	98.40
SFEW	-	48.29	52.29	-	-	49.18	49.18

TABLE 2 – Comparaison avec l'état de l'art sur les trois bases de données.