

Analyse de trajectoires sur des variétés de matrices pour la reconnaissance des expressions faciales

A. Kacem

M. Daoudi

B. Ben Amor

IMT Lille Douai,
Univ. Lille CNRS, UMR 9189 - CRISAL
Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France

Résumé

Dans ce travail nous nous intéressons au problème de reconnaissance des expressions faciales en se basant sur une analyse invariante aux variabilités temporelles des trajectoires de matrices vivant sur des variétés Riemanniennes bien définies. Nous considérons l'évolution temporelle des matrices X des marqueurs de visages pour construire une trajectoire sur une Grassmannienne $\mathcal{G}(n,2)$. Une autre trajectoire est introduite sur une variété moins connue, $\mathcal{S}^+(n,2)$ ou encore le cône des matrices symétriques semi-définie positives de rang fixe 2. Cette dernière est obtenue en considérant l'évolution temporelle des matrices de Gram XX^t obtenues à partir des matrices originales des marqueurs de visages X . Nous avons développé des outils géométriques pertinents pour aligner, comparer et calculer des moyennes statistiques d'un ensemble de trajectoires sur les deux variétés Riemanniennes considérées. L'approche proposée, testée sur la base de données CK+, a donné des résultats de reconnaissance compétitifs par rapport à ceux de la littérature tandis qu'elle ne requiert pas des techniques d'apprentissage automatique.

Mots Clef

Variété Grassmannienne, variété Riemannienne des matrices symétriques semi-définie positives de rang fixe, trajectoires sur des variétés Riemanniennes, alignement temporel, reconnaissance des expressions faciales.

Abstract

In this work, we address the problem of facial expression recognition based on a rate-invariant analysis of time-parametrized curves on special Riemannian manifolds. We consider the temporal evolution of landmark configurations X (of size n) to draw a trajectory representation $X(t)$ on $\mathcal{G}(n,2)$, the set of 2-frame of \mathbb{R}^n , called also a Grassmann manifold. A second trajectory representation is also introduced in a less familiar manifold, $\mathcal{S}^+(n,2)$ or the cone of symmetric positive semi-definite matrices of (low) fixed rank 2. The latter is obtained using the Gram matrices, i.e. XX^t , computed from the original landmark

configuration X . We develop relevant geometric tools to align, average and compare such trajectories on both underlying manifolds. Tested on the CK+ dataset, under commonly used experimental settings, the proposed approach achieved competitive accuracy whereas no Machine Learning technique required. Both representations are computationally efficient and allow also invariance to rigid motions.

Keywords

Grassmann manifold, Riemannian manifold of symmetric positive semi-definite matrices of fixed rank, trajectories on Riemannian manifold, temporal alignment, facial expression recognition.

1 Introduction

La comparaison de séquences faciales est un problème fondamental en reconnaissance des expressions faciales. En plus des propriétés d'invariance telles que l'invariance aux transformations géométriques nécessaires aux algorithmes d'analyse des expressions faciales, il faut s'assurer que ces méthodes sont robustes aux variabilités temporelles des séquences faciales. En effet, des séquences non alignées produisent des erreurs dans le calcul des métriques lors de la comparaison des séquences. Par conséquent, les quantités statistiques, telles que la moyenne des séquences temporelles, ne sont pas pertinentes pour la tâche de classification car leurs valeurs sont erronées. La méthode la plus utilisée pour résoudre le problème de l'alignement temporelle des caractéristiques temporelles représentant les séquences est la programmation dynamique (DTW). Bien que cette approche et d'autres variantes aient été utilisées pour l'alignement temporelle [21], ces méthodes utilisent des métriques Euclidiennes, et n'exploitent pas la non-linéarité et la dynamique de l'évolution des caractéristiques extraites à partir des séquences. Récemment, la géométrie Riemannienne a trouvé un large panel d'applications en vision par ordinateur telles que la reconnaissance des piétons [17] et l'analyse des séquences vidéos [2]. Elle offre un cadre unifié pour calculer des métriques dans des variétés non

linéaires, et permet aussi de faire des calculs statistiques dans ces variétés. Dans cet article, nous cherchons une représentation suffisamment discriminative pour classer une séquence donnée à l'une des classes d'expressions faciales invariantes à la variabilité temporelle et aux mouvements rigides des marqueurs du visage. Nous proposons d'embarquer les séquences temporelles des configurations de marqueurs du visage dans la variété de Grassmann et dans la variété de matrices symétriques semi-définies positives de rang fixe. Ainsi, les séquences sont considérées comme des trajectoires paramétrées par le temps sur ces deux variétés. La figure 1 résume l'approche proposée.

Les principales contributions de ce travail sont :

- Une nouvelle représentation géométrique de la dynamique des séquences faciales par des trajectoires paramétrées par le temps sur la variété Riemannienne des matrices symétriques semi-définies positives de rang fixe.
- Des outils statistiques pour la classification des trajectoires sur ces variétés.
- Une analyse invariante aux variabilités temporelles en utilisant le DTW (Dynamic Time Warping) appliqué à ces trajectoires.

Le reste du papier est organisé comme suit : Dans la section 2 nous présentons brièvement la géométrie des variétés étudiées. Dans la section 3 nous décrivons notre approche proposée. Les résultats expérimentaux et les discussions sont fournis dans la section 4. Finalement, nous concluons et présentons quelques perspectives dans la section 5.

2 Fondements théoriques

Dans cette section, nous présentons brièvement les bases théoriques des variétés Grassmannienne, la variété des matrices symétriques définies positives et la variété des matrices symétriques semi-définies positives de rang fixe.

2.1 Variété Grassmannienne

Soit $\mathcal{G}(n, k)$ l'ensemble des sous-espaces vectoriels de dimension k de \mathbb{R}^n (e.g. $\mathcal{G}(3, 2)$ est l'ensemble des plans de dimension 2 de \mathbb{R}^3) où $n > k$. Un sous-espace \mathcal{X} de $\mathcal{G}(n, k)$, est représenté par une matrice X de taille $n \times k$ dont les vecteurs colonnes forment une base orthogonale de ce sous-espace. On dit que la matrice X engendre le sous-espace \mathcal{X} .

L'ensemble des matrices de taille $n \times k$ avec des vecteurs colonnes orthogonaux forme une variété de Stiefel $\mathcal{V}(n, k)$. Les éléments de la Grassmannienne $\mathcal{G}(n, k)$ sont des classes d'équivalence des éléments de $\mathcal{V}(n, k)$ de façon que deux éléments sont équivalents si leurs bases orthogonales engendrent le même sous-espace [15]. Plus précisément, soit une matrice X qui représente un élément de $\mathcal{V}(n, k)$, si on effectue une multiplication à droite de X par une matrice orthogonale O de taille $k \times k$ on obtient une nouvelle matrice X_1 . D'une part, X_1 représente un autre élément différent de celui de X dans $\mathcal{V}(n, k)$, d'autre part X_1 représente le même élément que X dans $\mathcal{G}(n, k)$. Ainsi, on

peut noter l'invariance de $\mathcal{G}(n, k)$ aux transformations orthogonales de \mathbb{R}^k .

Une distance géodésique de la variété Grassmannienne est donnée par la norme l_2 du vecteur formé par les angles principaux entre deux sous-espaces. Etant donnés deux sous-espaces $\mathcal{X}_1, \mathcal{X}_2 \in \mathcal{G}(n, k)$ engendrés, respectivement, par X_1 et X_2 , la distance géodésique d_g entre \mathcal{X}_1 et \mathcal{X}_2 est définie par :

$$d_g^2(\mathcal{X}_1, \mathcal{X}_2) = \sum_i \theta_i^2 \quad (1)$$

Où, $\theta_i = \cos^{-1} \left(\max_{u_i \in \mathcal{X}_1} \max_{v_i \in \mathcal{X}_2} \langle u_i, v_i \rangle \right)$, u et v sont les vecteurs des bases qui engendrent, respectivement, \mathcal{X}_1 et \mathcal{X}_2 , $\langle \cdot, \cdot \rangle$ désigne le produit scalaire dans \mathbb{R}^n .

2.2 Variété Riemannienne des matrices symétriques définies positives

Soit $\mathcal{S}^{++}(k)$ l'ensemble des matrices symétriques définies positives de taille $k \times k$. Une matrice symétrique à coefficients réels est dite définie positive, si et seulement si, pour tout x non nul $\in \mathbb{R}^n$, $x^T M x > 0$. L'espace de ces matrices n'est pas un espace vectoriel, si on multiplie une matrice appartenant à $\mathcal{S}^{++}(k)$ par un scalaire négatif, on obtient une matrice qui n'est plus définie positive et donc n'appartient plus à $\mathcal{S}^{++}(k)$. En équipant $\mathcal{S}^{++}(k)$ d'une métrique Riemannienne, plusieurs travaux l'ont utilisé pour étudier la variété Riemannienne des matrices de covariance [16, 20, 18]. Plusieurs métriques ont été proposées pour $\mathcal{S}^{++}(k)$, les plus utilisées sont : (1) "affine-invariant metric" [13], (2) "log-euclidean metric" [4]. Dans ce travail, nous considérons uniquement la métrique "affine-invariant metric" qui est une vraie métrique Riemannienne. Etant données deux matrices S_1 et $S_2 \in \mathcal{S}^{++}(k)$, la distance géodésique est donnée par ,

$$d_{\mathcal{S}^{++}}^2(S_1, S_2) = \|\log(S_1^{-1/2} S_2 S_1^{-1/2})\|_F \quad (2)$$

Où $\|\cdot\|_F$ désigne la norme *Frobenius*.

2.3 Variété Riemannienne des matrices symétriques semi-définies positives de rang fixe

Soit $\mathcal{S}^+(n, k)$ l'ensemble des matrices symétriques semi-définies positives de taille $n \times n$ et de rang $k < n$. Une matrice symétrique à coefficients réels est dite semi-définie positive, si et seulement si, pour tout x non nul $\in \mathbb{R}^n$, $x^T M x \geq 0$. De même que $\mathcal{S}^{++}(k)$, $\mathcal{S}^+(n, k)$ n'est pas un espace vectoriel. Bonnabel *et al.* [6] ont introduit une métrique Riemannienne pour $\mathcal{S}^+(n, k)$ qui est presque égale à la somme des métriques dans la Grassmannienne $\mathcal{G}(n, k)$ et la variété des matrices symétriques définies positives $\mathcal{S}^{++}(k)$. Pour aboutir à ces conclusions, les auteurs de [6] ont procédé par les factorisations matricielles suivantes :

$$Y = X X^T = (UR)(UR)^T = U R^2 U^T \quad (3)$$

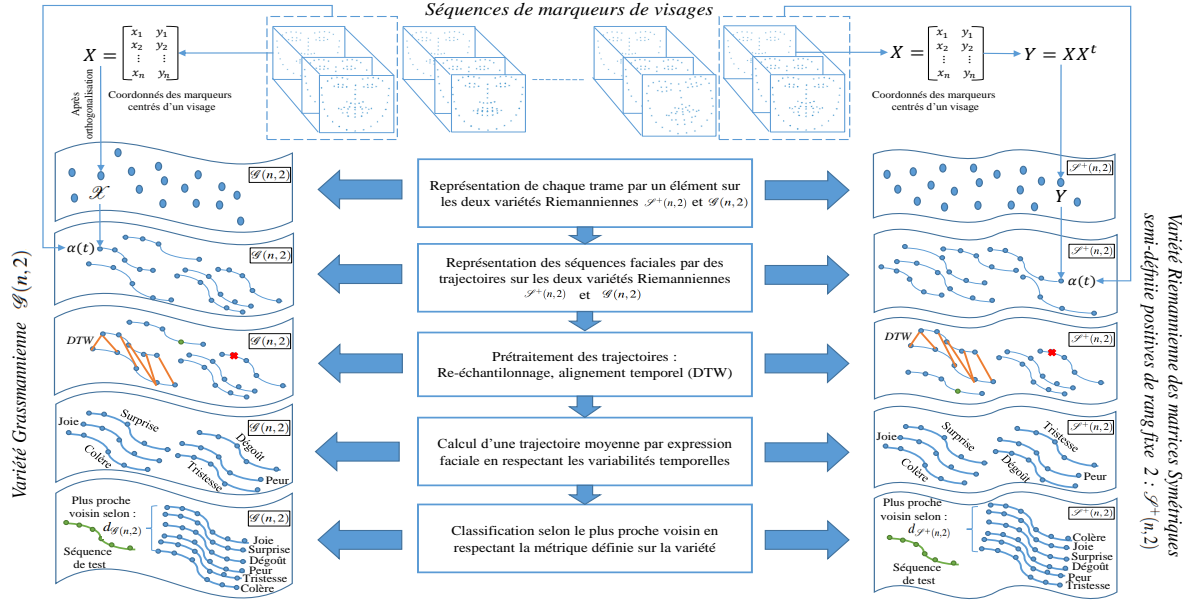


FIGURE 1 – Vue d'ensemble de l'approche proposée

Où $Y \in \mathcal{S}^+(n, k)$, X est une matrice de taille $n \times k$ et de rang k , $U \in \mathcal{V}(n, k)$ et $R \in \mathcal{S}^{++}(k)$.

La première factorisation est effectuée par une décomposition de *Cholesky* et la deuxième par une décomposition polaire [8]. Suite à la première factorisation, on peut remarquer l'invariance de ces matrices $Y \in \mathcal{S}^+(n, k)$ par rapport aux transformations orthogonales appliquées à la matrice X .

Démonstration 1. Soit $O \in \mathcal{O}(k)$ une matrice orthogonale, X_1 et X_2 deux matrices de taille $n \times k$ et de rang k et $Y_1, Y_2 \in \mathcal{S}^+(n, k)$ tel que $X_1 = X_2 O$, $Y_1 = X_1 X_1^t$ et $Y_2 = X_2 X_2^t$ alors,

$$Y_1 = X_1 X_1^t = (X_2 O)(X_2 O)^t = X_2 O O^t X_2^t = X_2 X_2^t = Y_2$$

Avec la deuxième factorisation, une matrice $Y \in \mathcal{S}^+(n, k)$ peut être représentée par un couple $(U, R^2) \in \mathcal{V}(n, k) \times \mathcal{S}^{++}(k)$. Bonnabel *et al.* [6] ont identifié une représentation de $\mathcal{S}^+(n, k)$ par un espace quotient où les éléments de $\mathcal{S}^+(n, k)$ sont des classes d'équivalence du groupe orthogonal $\mathcal{O}(k)$:

$$\mathcal{S}^+(n, k) \approx \mathcal{V}(n, k) \times \mathcal{S}^{++}(k) / \mathcal{O}(k) \quad (4)$$

En se basant sur cette représentation, les auteurs ont introduit une quasi-géodésique reliant deux éléments $Y_1, Y_2 \in \mathcal{S}^+(n, k)$,

$$\begin{aligned} \gamma_{Y_1 \rightarrow Y_2} : [0, 1] &\rightarrow \mathcal{S}^+(n, k) \\ \gamma_{Y_1 \rightarrow Y_2}(t) &= U(t) R^2(t) U^t(t) \end{aligned} \quad (5)$$

Où $U(t)$ est une géodésique sur la Grassmannienne $\mathcal{G}(n, k)$ et $R^2(t)$ est une géodésique sur $\mathcal{S}^{++}(k)$. La longueur de cette quasi-géodésique représente une mesure de similarité

entre $Y_1 = U_1 R_1^2 U_1^t$ et $Y_2 = U_2 R_2^2 U_2^t$,

$$d_{\mathcal{S}^+}^2(Y_1, Y_2) = d_{\mathcal{G}(n, k)}^2(U_1, U_2) + \lambda d_{\mathcal{S}^{++}(k)}^2(R_1^2, R_2^2) \quad (6)$$

Où $\lambda > 0$ est un paramètre qui contrôle la contribution des distances de $\mathcal{S}^{++}(k)$ et $\mathcal{G}(n, k)$. Il est à préciser que $d_{\mathcal{S}^+(n, k)}$ n'est pas une distance car elle ne vérifie pas l'inégalité triangulaire et que des valeurs faibles de λ sont recommandées [6].

2.4 Moyenne intrinsèque

Etant donné L échantillons S_1, S_2, \dots, S_L sur une variété Riemannienne \mathcal{M} , nous nous intéressons au calcul d'une moyenne intrinsèque \tilde{S} qui vit sur cette variété \mathcal{M} . En effet, une moyenne Euclidienne calculée à partir de ces matrices n'appartiendra pas à cette variété à cause de la non-linéarité de cet espace. Pour remédier à ce problème, nous utilisons la métrique $d_{\mathcal{M}}$ définie sur \mathcal{M} pour le calcul d'une moyenne intrinsèque $\tilde{S} \in \mathcal{M}$ selon :

$$\tilde{S} = \arg \min_{S \in \mathcal{M}} \sum_{i=1}^L d_{\mathcal{M}}(S, S_i)^2, \quad (7)$$

qui minimise l'erreur quadratique moyenne en utilisant une métrique appropriée $d_{\mathcal{M}}$ [9]. Des algorithmes itératifs ont été proposés pour résoudre ce problème pour $\mathcal{G}(n, k)$ [15] et $\mathcal{S}^{++}(k)$ [5].

Dans [6], les auteurs ont introduit une moyenne intrinsèque $\tilde{Y} \in \mathcal{S}^+(n, k)$ de deux éléments $Y_1 = U_1 R_1^2 U_1^t$ et $Y_2 = U_2 R_2^2 U_2^t \in \mathcal{S}^+(n, k)$:

$$\tilde{Y} = \tilde{U} \tilde{R}^2 \tilde{U}^t \quad (8)$$

Où \tilde{U} est une moyenne intrinsèque de U_1 et U_2 dans $\mathcal{G}(n, k)$ et \tilde{R}^2 est une moyenne intrinsèque de R_1^2 et R_2^2 dans

$\mathcal{S}^{++}(k)$.

Pour calculer une moyenne intrinsèque d'un ensemble d'éléments sur $\mathcal{S}^+(n, k)$, nous proposons de calculer une moyenne intrinsèque des matrices obtenues après la factorisation sur $\mathcal{G}(n, k)$ selon la méthode décrite dans [15] et une autre moyenne sur $\mathcal{S}^{++}(k)$ selon [9]. La moyenne de cet ensemble d'échantillons sur $\mathcal{S}^+(n, k)$ est obtenue en combinant les deux moyennes selon l'équation (8).

3 Représentation des séquences faciales par des trajectoires sur des variétés Riemanniennes

Pour analyser les séquences faciales, nous nous sommes intéressés uniquement aux marqueurs des visages et leur évolution temporelle. Dans ce contexte, plusieurs travaux ont représenté l'évolution temporelle des marqueurs par des trajectoires paramétrées par le temps sur des variétés Riemanniennes bien définies : variété Grassmannienne [14, 1], espace de formes de Kendall [3], etc. Dans cette section, nous introduisons une nouvelle approche pour représenter les marqueurs des visages et leur évolution temporelle par des trajectoires de matrices de Gram qui vivent sur la Variété Riemannienne des matrices symétriques semi-définies positives de rang fixe 2. Nous commençons par aligner ces trajectoires dans le temps pour avoir une distance significative entre elles. Ensuite, nous présentons quelques outils permettant de les analyser.

3.1 Trajectoires de marqueurs de visages

La représentation par des marqueurs de visages est une technique très utilisée pour modéliser approximativement la géométrie des visages. Dans un premier temps, en s'inspirant de [14], nous avons modélisé la géométrie du visage par une matrice $X = [(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]^t$, de taille $n \times 2$ où n représente le nombre de marqueurs pour un visage. Après des procédures d'orthogonalisation et de centrage, ces matrices vivent sur une Grassmannienne $\mathcal{G}(n, 2)$ que nous avons présentée dans la section 2.1.

En se basant sur cette représentation, nous proposons dans notre travail de représenter chaque trame par une matrice $Y = XX^t$ de taille $n \times n$ appelée matrice de Gram. D'une part, ces matrices sont très riches car elles décrivent la corrélation entre les coordonnées des marqueurs du visage et donc, les petits déplacements des marqueurs seront plus prononcés dans ces matrices. D'autre part, elles permettent de représenter et analyser les matrices des marqueurs X sans passer par une orthogonalisation qui engendre une perte d'information. Les matrices de Gram sont symétriques semi-définies positives et de rang égal à celui de la matrice primitive, i.e. $\text{rang}(Y) = \text{rang}(X)$. Dans notre cas, on peut remarquer que $\text{rang}(X) = 2$, et donc ces matrices de Gram ont un rang fixe égal à 2 et vivent sur la variété Riemannienne $\mathcal{S}^+(n, 2)$ des matrices symétriques semi-définies positives de rang 2 qui a été présentée dans la section 2.3. Il est à préciser que nous n'effectuons pas une

décomposition de *Cholesky* comme indiqué dans l'équation (3). En effet, les matrices primitives des marqueurs X vérifient les conditions de la première factorisation (X est une matrice de taille $n \times 2$ et de rang 2) et sont utilisées pour la deuxième factorisation.

En considérant la succession de ces matrices, l'analyse des séquences faciales revient à analyser les trajectoires de ces matrices sur les variétés sur lesquelles elles vivent. Une trajectoire sur une variété Riemannienne \mathcal{M} qui généralise $\mathcal{S}^+(n, 2)$ et $\mathcal{G}(n, 2)$ peut être considérée comme une courbe $\alpha : I \rightarrow \mathcal{M}$, où I désigne le domaine temporel.

Un autre avantage de ces représentations, est l'invariance des marqueurs aux rotations de \mathbb{R}^2 due à l'invariance aux transformations orthogonales mentionnée dans les sections 2.1 et 2.3. Ainsi, l'alignement spatial des marqueurs de visage n'est pas nécessaire avant l'analyse de ces trajectoires.

Veillez noter que nous utilisons une notation $(\mathcal{M}, d_{\mathcal{M}})$ pour généraliser $(\mathcal{G}(n, 2), d_{\mathcal{G}(n, 2)})$ et $(\mathcal{S}^+(n, 2), d_{\mathcal{S}^+(n, 2)})$.

3.2 Re-échantillonnage adaptatif des trajectoires

Un outil important dans notre approche consiste à augmenter ou réduire le nombre d'échantillons dans une trajectoire.

D'une part, l'augmentation des échantillons implique une plus haute résolution temporelle des séquences qui garantit une meilleure performance dans l'alignement temporel ou le calcul de la moyenne des séquences temporelles. Pour augmenter le nombre d'échantillons d'une trajectoire, nous cherchons les formes de visages qui ont des distances maximales par rapport à l'échantillon précédent et nous générons un nouvel échantillon entre eux, défini par la moyenne géométrique de l'échantillon et son précédent.

D'autre part, pour réduire les échantillons, nous supprimons de la séquence les échantillons ayant les distances minimales par rapport à l'échantillon précédent. En d'autres termes, nous éliminons les échantillons les moins importants représentant des formes de visages similaires à celles qui les précèdent. La réduction des échantillons nous permet de réduire les temps de calcul.

3.3 Alignement temporel des trajectoires

Pour pouvoir comparer et analyser les séquences faciales, nous devons tenir compte des variabilités temporelles qui peuvent survenir. L'alignement temporel des séquences faciales, qui sont représentées par des trajectoires sur des variétés Riemanniennes consiste à aligner dans le temps ces trajectoires.

Etant donnée une variété Riemannienne \mathcal{M} équipée d'une métrique $d_{\mathcal{M}}$ et deux trajectoires sur $\mathcal{M} : \alpha_1(t), \alpha_2(t) : I \rightarrow \mathcal{M}$ où I désigne le domaine temporel, le problème de l'alignement des deux trajectoires α_1 et α_2 revient à trouver la fonction de re-paramétrisation optimale γ^* appliquée sur l'une des trajectoires et minimisant la distance entre elles

selon,

$$\gamma^* = \arg \min_{\gamma \in \Gamma} \int_I d_{\mathcal{M}}(\alpha_1(t), \alpha_2(\gamma(t))) dt \quad (9)$$

Où Γ désigne l'ensemble des fonctions croissantes $\gamma: I \rightarrow I$. La méthode la plus utilisée pour résoudre ce problème d'optimisation est l'algorithme de *Dynamic Time Warping* (DTW). L'adaptation du DTW pour des séquences de matrices vivant sur des variétés Riemanniennes peut être fait en considérant la métrique appropriée au lieu de la distance Euclidienne.

3.4 Calcul d'une trajectoire moyenne

Pour générer des modèles pour les différentes expressions faciales, nous proposons de calculer une séquence faciale moyenne par expression faciale tout en respectant les variabilités temporelles qui peuvent survenir. Les séquences faciales étant représentées par des trajectoires sur une variété Riemannienne \mathcal{M} , le problème du calcul de cette moyenne revient à calculer une trajectoire moyenne intrinsèque à la variété \mathcal{M} tout en respectant les variabilités temporelles. Pour cela, nous procédons comme suit : La trajectoire moyenne est initialisée aléatoirement à l'une des trajectoires en entrée. Ensuite, nous alignons dans le temps ces trajectoires à la trajectoire moyenne en utilisant le DTW introduit dans la section 3.3. Une nouvelle trajectoire moyenne est obtenue en calculant une moyenne intrinsèque (*élément par élément*) de toutes les trajectoires comme présenté dans la section 2.4. Finalement, nous itérons ces étapes jusqu'à convergence atteinte lorsque la distance $d_{\mathcal{M}}$ entre la nouvelle trajectoire moyenne et la trajectoire moyenne courante est inférieure à un certain petit seuil ε . Nous pouvons noter qu'une étape de rééchantillonnage des trajectoires, comme mentionné dans la section section 3.2, est nécessaire pour avoir un nombre fixe d'échantillons pour toutes les trajectoires.

4 Résultats expérimentaux

Pour illustrer l'efficacité de l'approche proposée et comparer les deux métriques étudiées dans ce travail (trajectoires sur $\mathcal{G}(n, 2)$ et $\mathcal{S}^+(n, 2)$), nous avons utilisé la base de données CK+.

4.1 Corpus de test

La base de données CK+ (The Cohn-Kanade Extended Facial Expression) [12] a été développée pour l'analyse et la synthèse des expressions faciales. Elle contient 123 sujets et 593 séquences d'images frontales. Parmi ces sujets, 118 ont été annotés par les sept expressions universelles (la colère, le mépris, le dégoût, la peur, la joie, la tristesse et la surprise). Les visages sont annotés avec 68 marqueurs en deux dimensions que nous utilisons dans notre approche.

4.2 Calcul des distances à la première et précédente trame

Afin d'illustrer les phases temporelles d'une expression faciale, nous calculons les distances par rapport au premier

élément d'une trajectoire, qui correspond à l'état initial d'une personne. Le calcul des distances des autres éléments de la trajectoire par rapport au premier élément nous permet de quantifier l'intensité des déformations du visage comparée à l'état initial (voir figure 2 (gauche)).

Les distances par rapport à la trame précédente sont très utiles pour quantifier l'intensité des déformations du visage entre deux trames successives (voir figure 2 (droite)).

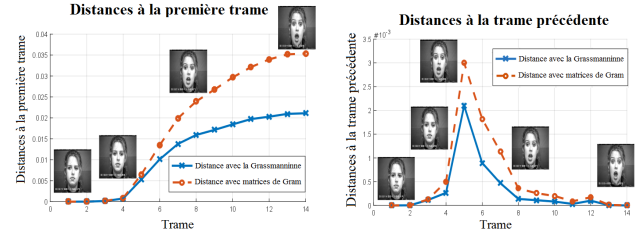


FIGURE 2 – Deux exemples de l'évolution des distances à la première et précédente trame en utilisant les métriques définies sur $\mathcal{G}(n, 2)$ et $\mathcal{S}^+(n, 2)$.

4.3 Alignement des trajectoires et calcul d'une trajectoire moyenne

Pour montrer l'importance de l'alignement temporel avant la comparaison des séquences comme indiqué dans la section section 3.3, nous avons sélectionné deux séquences de la base CK+ de deux sujets effectuant la même expression faciale (la joie) avec des variabilités temporelles (voir la partie gauche de la figure 4). Les résultats de l'alignement temporel de ces séquences sur $\mathcal{G}(n, 2)$ et $\mathcal{S}^+(n, 2)$ sont donnés par la partie droite de la figure 4. Nous rappelons que les distances à la première trame sont utilisées comme indiqué dans la section 4.2 pour illustrer les phases temporelles avant et après l'alignement temporel. La figure 3 montre les fonctions de re-paramétrisation après l'alignement sur $\mathcal{G}(n, 2)$ et $\mathcal{S}^+(n, 2)$. On remarque une différence peu significative entre les deux courbes entre la trame 5 et la trame 15 qui peut être vue dans la partie gauche de la figure 4 où, l'alignement dans $\mathcal{S}^+(n, 2)$ est visuellement peu performant que dans $\mathcal{G}(n, 2)$.

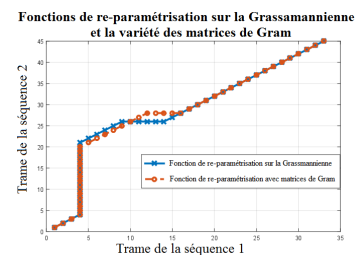


FIGURE 3 – Fonctions de re-paramétrisation des deux trajectoires (introduites dans la Figure 4) sur $\mathcal{G}(n, 2)$ (en bleu) and $\mathcal{S}^+(n, 2)$ (en rouge).

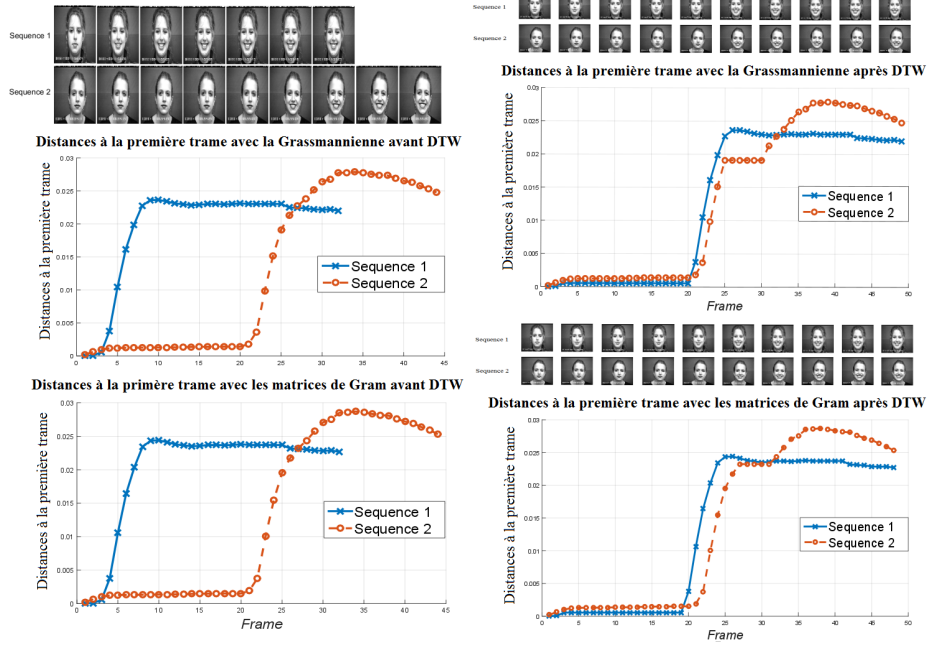


FIGURE 4 – Gauche : Distances à la première trame avant l’alignement temporel ; Droite : Après l’alignement temporel.

Pour montrer l’efficacité de la méthode proposée du calcul de la moyenne spatio-temporelle des séquences, nous avons calculé une trajectoire moyenne dans $\mathcal{G}(n,2)$ et $\mathcal{S}^+(n,2)$ de toutes les séquences de la classe "joie" (69 séquences) comme décrit dans la section 3.4. La figure 5 montre 5 trames de chaque trajectoire moyenne obtenue. On remarque visuellement que la trajectoire moyenne dans $\mathcal{S}^+(n,2)$ ¹ montre des déformations plus accentuées que celle dans $\mathcal{G}(n,2)$. Veuillez noter que nous avons choisi une rotation arbitraire pour visualiser les marqueurs puisque notre approche est invariante aux rotations de \mathbb{R}^2 .

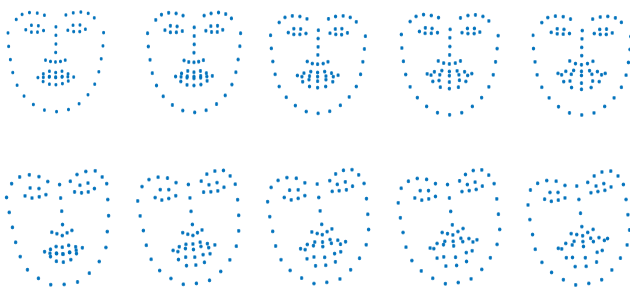


FIGURE 5 – Trajectoires moyennes (haut) sur $\mathcal{G}(n,2)$ et (bas) sur $\mathcal{S}^+(n,2)$, calculées à partir de 69 sujets effectuant l’expression joie de la base CK+.

1. Nous avons appliqué une *décomposition de Cholesky* sur les matrices de Gram pour visualiser la configuration des marqueurs correspondante.

4.4 Reconnaissance des expressions faciales

Afin de valider notre représentation et évaluer la robustesse des outils statistiques introduits, nous nous sommes intéressés à la reconnaissance des expressions faciales dans la base CK+. A cet effet, nous avons généré une trajectoire moyenne pour chacune des 6 classes (colère, dégoût, peur, joie, tristesse et surprise). Ensuite, nous avons aligné toutes les séquences (trajectoires) de test à ces trajectoires moyennes comme indiqué dans la section 3.3. Finalement, nous avons utilisé la méthode du plus proche voisin pour prédire la classe de la séquence de test. En suivant le protocole expérimental le plus utilisé pour cette base de données [19, 11, 14, 7], nous avons effectué une validation croisée (*leave-one-subject-out*). En utilisant des trajectoires sur $\mathcal{S}^+(n,2)$ nous avons obtenu un taux de reconnaissance moyen de 87.7% dépassant celui des trajectoires dans $\mathcal{G}(n,2)$ de 1.8% (et donc 5 séquences sont reconnues en plus). Cette amélioration est illustrée plus en détails dans les matrices de confusion (figure 6) où *la joie*, *la colère*, *la surprise* et *le dégoût* sont un peu plus reconnus dans $\mathcal{S}^+(n,2)$.

Impact du paramètre λ – Nous rappelons que la métrique définie sur $\mathcal{S}^+(n,2)$ dans l’équation (6) fait intervenir un paramètre $\lambda > 0$ qui contrôle la contribution des métriques sur la Grassmannienne $\mathcal{G}(n,2)$ et $\mathcal{S}^{++}(2)$. Bonnabel *et al.* [6] recommandent d’utiliser des petites valeurs pour ce paramètre. Des résultats de reconnaissance pour différentes valeurs de λ sont illustrés dans la figure 6. Une meilleure performance est obtenue pour une valeur de $\lambda = 0.01$ qui est maintenue pour le reste des expérimentations.

Différences entre $\mathcal{G}(n,2)$ and $\mathcal{S}^+(n,2)$ – Comme indiqué dans l’équation (6), la mesure de similarité (qui n’est

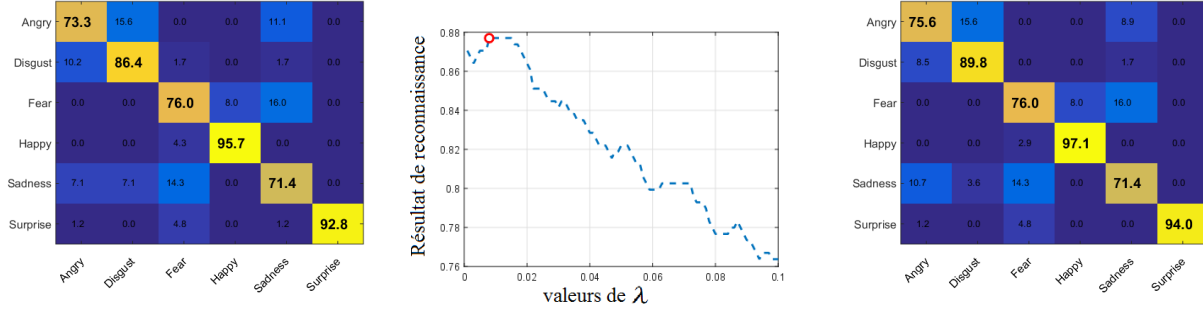


FIGURE 6 – De gauche à droite - Matrice de confusion de la reconnaissance des trajectoires sur $\mathcal{G}(n, 2)$; Taux de reconnaissance pour différentes valeurs de λ ; Matrice de confusion de la reconnaissance des trajectoires sur $\mathcal{S}^+(n, 2)$.

pas une distance) contient deux termes. Le premier terme se résume à la métrique de la Grassmannienne qui décrit la forme construite par les marqueurs, et le deuxième encode la corrélation entre ces marqueurs et se résume à la métrique sur $\mathcal{S}^{++}(2)$. Le paramètre λ discuté dans la section précédente permet de contrôler la contribution de ces deux termes. En considérant uniquement le premier terme ($\lambda = 0$), l'étude sur $\mathcal{S}^+(n, 2)$ se résume à celle sur $\mathcal{G}(n, 2)$. En ajoutant le deuxième terme, le taux de reconnaissance passe de 86.08% à 87.87% pour une valeur de $\lambda = 0.01$.

Impact de la résolution des trajectoires – Un autre paramètre important dans notre approche est le nombre d'échantillons à considérer dans une trajectoire. En effet, les trajectoires qui représentent les séquences doivent contenir le même nombre d'échantillons (trames) pour pouvoir être comparées. A cet effet, nous avons utilisé la méthode de re-échantillonnage adaptatif décrite dans la section 3.2. Le meilleur taux de reconnaissance est obtenu pour une valeur de 50 échantillons.

Impact de l'alignement temporel (DTW) – Pour quantifier la contribution de l'alignement temporel des trajectoires, nous avons réalisé les mêmes expériences sans appliquer le DTW. La performance de notre approche avec des trajectoires sur $\mathcal{G}(n, 2)$ diminue de 86.08% à 78.96% et de 87.87% à 79.05% avec des trajectoires sur $\mathcal{S}^+(n, 2)$.

Complexité de calcul – Un avantage de notre méthode est son efficacité en termes de temps d'exécution. Contrairement à [7], notre approche ne nécessite aucune normalisation pour une analyse invariante aux rotations de \mathbb{R}^2 . Dans le tableau 1, nous présentons le temps d'exécution² nécessaire (en millisecondes) pour comparer deux trajectoires arbitraires avec et sans alignement temporel (DTW). Nous rappelons que nous utilisons 68 marqueurs pour chaque visage et que le nombre d'échantillons considérés pour chaque trajectoire est égal à 50.

Étude comparative avec l'état de l'art – Dans la littérature plusieurs travaux ont choisi la base de données CK+ pour évaluer leurs méthodes. Dans notre étude, nous nous comparons uniquement aux travaux basés sur une représentation par les marqueurs du visage sans considérer l'infor-

TABLE 1 – Temps d'exécution de la comparaison de deux trajectoires sur $\mathcal{G}(n, 2)$ et $\mathcal{S}^+(n, 2)$.

Trajectoires sur	Sans DTW	Avec DTW
$\mathcal{G}(n, 2)$	1.055	78.985
$\mathcal{S}^+(n, 2)$	10.118	148.645

mation donnée par l'image couleur. Notre approche donne des résultats comparables à ceux de l'état de l'art tandis qu'elle ne requiert aucune technique d'apprentissage automatique. Comme le montre le tableau 2, le taux de reconnaissance que nous avons obtenu dépasse celui de toutes les méthodes existantes à l'exception de [7]. En particulier, tandis que Taheri *et al.* [14] représentent les séquences par des trajectoires sur $\mathcal{G}(n, 2)$ et calculent les vecteurs vélocités pour les utiliser dans un classifieur SVM, notre approche se base sur une métrique appropriée. De plus, le calcul de ces vecteurs vélocités est coûteux en termes de temps d'exécution.

TABLE 2 – Étude comparative de la méthode proposée avec les méthodes existantes (basées sur les marqueurs de visages) sur la base CK+

Méthode	RR (%)
Taheri <i>et al.</i> [14]	85.8
Wang <i>et al.</i> [19]	86.3
Li <i>et al.</i> [10]	87.43
Ghimire <i>et al.</i> [7]	97.35
Traj. sur $\mathcal{G}(n, 2)$	86.08
Traj. sur $\mathcal{S}^+(n, 2)$	87.87

Dans [7], les auteurs obtiennent un taux de reconnaissance de 97.35% en utilisant un classifieur SVM sur des descripteurs géométriques (distances et angles entre les marqueurs) boostés par AdaBoost. Cette représentation rend l'approche sensible aux systèmes de détection et suivi de marqueurs. De plus, cette approche requiert, (1) une normalisation géométrique pour chaque trame de la séquence, (2) un nombre de trames fixe pour toutes les séquences qui est obtenu par une interpolation linéaire tandis qu'une

2. Notre programme Matlab est exécuté sur un ordinateur avec un processeur 2.8 GHZ CPU.

interpolation non-linéaire par les géodésiques est utilisée dans notre approche.

5 Conclusions

Dans ce travail, nous avons proposé une nouvelle approche géométrique pour modéliser la dynamique des séquences faciales. Les matrices de Gram relatives aux matrices des coordonnées des marqueurs des visages ont été utilisées pour représenter chaque trame d'une séquence. L'évolution temporelle des marqueurs est étudiée en considérant des trajectoires paramétrées par le temps sur la variété Riemannienne des matrices symétriques semi-définies positives de rang fixe. L'adaptation du DTW pour des trajectoires sur ces variétés a résolu le problème de variabilités temporelles des séquences. Nous avons aussi introduit une trajectoire moyenne qui tient en compte des variabilités temporelles. En utilisant un simple classifieur basé sur l'algorithme du "plus proche voisin" par rapport à ces modèles, les résultats de reconnaissance obtenus sont comparables par rapport à ceux de la littérature.

Combiner notre approche de représentation avec des techniques avancées d'apprentissage automatique sur ces trajectoires pourrait augmenter nos résultats de reconnaissance. Une extension de ce travail pourrait aussi inclure l'utilisation d'autres informations sur les séquences faciales telles que la texture ou l'image 3D.

Références

- [1] T. Alashkar, B. B. Amor, S. Berretti, and M. Daoudi. Analyzing trajectories on grassmann manifold for early emotion detection from depth videos. In *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, volume 1, pages 1–6. IEEE, 2015.
- [2] T. Alashkar, B. B. Amor, M. Daoudi, and S. Berretti. A grassmann framework for 4d facial shape analysis. *Pattern Recognition*, 3(3) :349–365, 2016.
- [3] B. B. Amor, J. Su, and A. Srivastava. Action recognition using rate-invariant analysis of skeletal shape trajectories. *IEEE transactions on pattern analysis and machine intelligence*, 38(1) :1–13, 2016.
- [4] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic resonance in medicine*, 56(2) :411–421, 2006.
- [5] D. A. Bini and B. Iannazzo. Computing the karcher mean of symmetric positive definite matrices. *Linear Algebra and its Applications*, 438(4) :1700–1710, 2013.
- [6] S. Bonnabel and R. Sepulchre. Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank. *SIAM Journal on Matrix Analysis and Applications*, 31(3) :1055–1070, 2009.
- [7] D. Ghimire and J. Lee. Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines. *Sensors*, 13(6) :7714–7734, 2013.
- [8] N. J. Higham. Computing the polar decomposition-with applications. *SIAM Journal on Scientific and Statistical Computing*, 7(4) :1160–1174, 1986.
- [9] H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on pure and applied mathematics*, 30(5) :509–541, 1977.
- [10] Y. Li, S. Wang, Y. Zhao, and Q. Ji. Simultaneous facial feature tracking and facial expression recognition. *IEEE Transactions on Image Processing*, 22(7) :2559–2573, 2013.
- [11] M. Liu, R. Wang, S. Shan, and X. Chen. Learning prototypes and similes on grassmann manifold for spontaneous expression recognition. *Computer Vision and Image Understanding*, 147 :95–101, 2016.
- [12] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended cohn-kanade dataset (ck+) : A complete dataset for action unit and emotion-specified expression. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pages 94–101. IEEE, 2010.
- [13] X. Pennec, P. Fillard, and N. Ayache. A riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1) :41–66, 2006.
- [14] S. Taheri, P. Turaga, and R. Chellappa. Towards view-invariant expression analysis using analytic shape manifolds. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 306–313. IEEE, 2011.
- [15] P. Turaga, A. Veeraraghavan, A. Srivastava, and R. Chellappa. Statistical computations on grassmann and stiefel manifolds for image and video-based recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(11) :2273–2286, 2011.
- [16] O. Tuzel, F. Porikli, and P. Meer. Region covariance : A fast descriptor for detection and classification. In *European conference on computer vision*, pages 589–600. Springer, 2006.
- [17] O. Tuzel, F. Porikli, and P. Meer. Pedestrian detection via classification on riemannian manifolds. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(10) :1713–1727, Oct. 2008.
- [18] R. Wang, H. Guo, L. S. Davis, and Q. Dai. Covariance discriminative learning : A natural and efficient approach to image set classification. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2496–2503. IEEE, 2012.
- [19] Z. Wang, S. Wang, and Q. Ji. Capturing complex spatio-temporal relations among facial muscles for facial expression recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3422–3429, 2013.
- [20] C. Yuan, W. Hu, X. Li, S. Maybank, and G. Luo. Human action recognition under log-euclidean riemannian metric. In *Asian Conference on Computer Vision*, pages 343–353. Springer, 2009.
- [21] X. Zhang, Y. Wang, M. Gou, M. Sznajder, and O. Camps. Efficient temporal sequence comparison and classification using gram matrix embeddings on a riemannian manifold. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.