

3D Rotation Invariant Decomposition of Motion Signals

Quentin Barthélemy^{1,2}, Anthony Larue¹, and Jérôme I. Mars²

¹ CEA-LIST, LOAD, 91191 Gif-sur-Yvette, France
{quentin.barthelemy, anthony.larue}@cea.fr

² Grenoble-INP, GIPSA-Lab, DIS, 38402 Grenoble, France
jerome.mars@gipsa-lab.grenoble-inp.fr

Abstract. A new model for describing a three-dimensional (3D) trajectory is introduced in this article. The studied object is viewed as a linear combination of rotatable 3D patterns. The resulting model is now 3D rotation invariant (3DRI). Moreover, the temporal patterns are considered as shift-invariant. A novel 3DRI decomposition problem consists of estimating the active patterns, their coefficients, their rotations and their shift parameters. Sparsity allows to select few patterns among multiple ones. Based on the sparse approximation principle, a non-convex optimization called 3DRI matching pursuit (3DRI-MP) is proposed to solve this problem. This algorithm is applied to real and simulated data, and compared in order to evaluate its performances.

Keywords: 3D, motion trajectory, rotation invariant, shift-invariant, matching pursuit, Procrustes, registration.

1 Introduction

In 3D space, a time-varying 3D trajectory composed of N temporal samples is considered. This trajectory is decomposed on elementary patterns and is thus described as the sum of K vectors. Different models can be considered.

In computer vision, Akhter et al. [1] described a non-rigid 3D object of P points as P temporal trajectories of N samples. A single point will be considered in this work. Thus, the 3D trajectory $y \in \mathbb{R}^{3 \times N}$ is defined as:

$$y = \sum_{k=1}^K a_k \theta_k, \quad (1)$$

where $a_k \in \mathbb{R}^{3 \times 1}$ are the coefficients, and $\theta_k \in \mathbb{R}^{1 \times N}$ are the trajectory basis vectors. The trajectory y is the sum of K trajectory basis vectors $\{\theta_k\}_{k=1}^K$, as illustrated in Fig. 1 (*top*). Contrary to its dual model based on a *shape* basis [2], the advantage of a *trajectory* basis [1] is to be defined independently of the data. So, generic basis as fast transforms can be employed with this model. More particularly, discret cosine transform (DCT) appears to be a well-adapted generic basis to study motion signals (as also noticed in [3]).

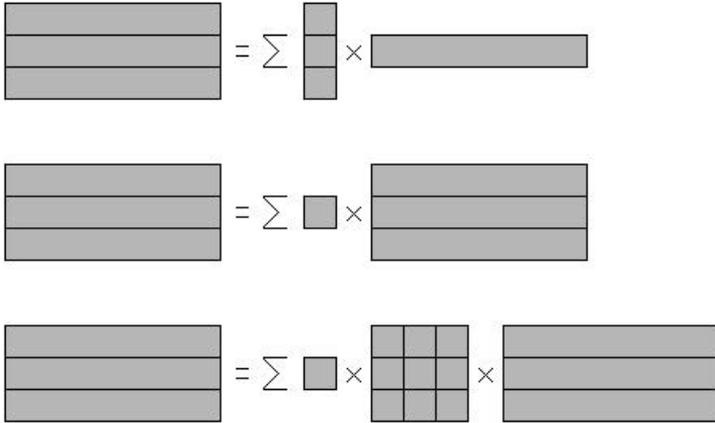


Fig. 1. Illustration of the three models to describe a 3D trajectory. Akter’s et al. model (*top*), Barthélemy’s et al. model (*middle*) and the introduced 3DRI model (*bottom*).

In signal processing, Barthélemy et al. [3] described a multicomponent temporal signal as the sum of multicomponent patterns. They used a redundant basis composed of $M > N$ elements which is called dictionary. In this case, elements of the dictionary are no more called vectors but atoms. Considering here the particular case of tricomponent data, a 3D trajectory of N samples is viewed as the sum of K 3D trajectories. The trajectory $y \in \mathbb{R}^{3 \times N}$ is defined as:

$$y = \sum_{k=1}^K x_k \phi_k, \tag{2}$$

where $x_k \in \mathbb{R}$ are the coefficients, and $\phi_k \in \mathbb{R}^{3 \times N}$ are the 3D atoms. Thereby, as seen in Fig. 1 (*middle*), the trajectory y is viewed as a weighted sum of K 3D atoms. This model is different from the Akhter model. Indeed, in model (1), each unicomponent trajectory θ_k (1D pattern) is multiplied by three coefficients, one by dimension. In model (2), each tricomponent trajectory ϕ_k (3D pattern) is multiplied by a scale factor. Here, the advantage of using this model is to deal with 3D trajectory patterns $\phi_k \in \mathbb{R}^{3 \times N}$ whose the three components can be different contrary to model (1) which has the same pattern on the three components. The differences between model (1), called *multichannel* framework, and model (2), called *multivariate*, are well detailed in [3].

The purpose of this article is to provide a 3D rotation invariant (3DRI) model. Thus, a rotation matrix $R_k \in \mathbb{R}^{3 \times 3}$ is added to each 3D atoms ϕ_k and model (2) becomes:

$$y = \sum_{k=1}^K x_k R_k \phi_k. \tag{3}$$

Each rotation matrix R_k has to be orthogonal, so it has to verify the condition: $R_k R_k^T = Id$. Finally, the trajectory y is represented as a weighted sum of rotatable 3D atoms as seen in Fig. 1 (*bottom*).

In 3DRI decomposition, we are interested in the estimation of coefficients $\mathbf{x} = \{x_k\}_{k=1}^K$ and rotation matrices $\mathbf{R} = \{R_k\}_{k=1}^K$ of the model (3). The problem to solve here is:

$$\min_{\mathbf{x}, \mathbf{R}} \left\| y - \sum_{k=1}^K x_k R_k \phi_k \right\|^2 \quad s.t. \quad \forall k \in \mathbb{N}_K, R_k R_k^T = Id, \quad (4)$$

where $\|\cdot\|$ is the Frobenius norm and $\langle A, B \rangle = trace(AB^T)$ is the associated matrix inner product, with $(\cdot)^T$ the transpose operator. To the best of our knowledge, this problem has not been addressed and we ignore if an analytic solution exists to solve it. It can be viewed as a generalization of the orthogonal Procrustes problem [4,5] which usually deals with the registration of a single pattern. Moreover, the shift-invariant case will be considered hereafter. Using a sparsity constraint, we propose a nonconvex optimization to solve this more complex problem, based on the matching pursuit (MP) principle. This introduced algorithm is called 3D rotation invariant matching pursuit (3DRI-MP).

We first present existing methods to solve 3D registration problems in Section 2 and then the shift and 3D rotation invariance problem is defined. 3D rotation invariant MP is introduced in Section 3 and is illustrated on real data in Section 4. As validation, experiments on simulation data are shown in Section 5.

2 State of the Art and Problem

In this section, the state of the art in rigid 3D registration is first presented and the shift and 3D rotation invariance problem is then detailed.

2.1 State of the Art in Rigid 3D Registration

In this paragraph, 3D decomposition problems related to problem (4) are mentioned. A rigid transformation composed of a 3D rotation R and a spatial translation T is considered here between the trivariate pattern ϕ and the original signal y . The rigid 3D registration, also called orthogonal Procrustes problem, consists of finding parameters R and T such that:

$$\min_{R, T} \| y - R \phi - T \|^2 \quad s.t. \quad R R^T = Id. \quad (5)$$

Eggert et al. [4] reviewed the several methods that give an analytical solution to this rigid 3D registration problem: singular value decomposition (SVD), unit quaternions, and orthonormal matrix.

In [5], Gower and Dijksterhuis reviewed multiple different Procrustes problems and many generalizations, notably the generalized Procrustes analysis (GPA) [5]. However, problem (4) is not addressed. Note also that neither the multiview reconstruction problem [6] nor the iterative closest point (ICP) algorithm [7] solve problem (4).

2.2 Shift and 3D Rotation Invariance Problem

In this paragraph, the shift-invariance and the sparse approximation are first detailed, and the shift and 3D rotation invariance problem is then explained.

In the shift-invariant case, we want to sparsely code the signal y as a sum of a few short structures, known as kernels, that are characterized independently of their positions. This model is usually applied to time-series data, and it avoids block effects in the analysis of largely periodic signals and provides a compact kernel dictionary [3]. The L shiftable kernels of the compact dictionary Ψ are replicated at all of the positions, to provide the M atoms of the dictionary Φ . The N samples of the signal y , the residual error ϵ , and the atoms ϕ_m are indexed¹ by t . The kernels $\{\psi_l\}_{l=1}^L$ can have different lengths. The kernel $\psi_l(t)$ is shifted in the τ samples to generate the atom $\psi_l(t - \tau)$, and the subset σ_l collects the translations τ of the kernel $\psi_l(t)$. For the few kernels that generate all of the atoms, we have:

$$y(t) = \sum_{m=1}^M x_m \phi_m(t) + \epsilon(t) = \sum_{l=1}^L \sum_{\tau \in \sigma_l} x_{l,\tau} \psi_l(t - \tau) + \epsilon(t). \quad (6)$$

The signal y is thus approximated as a weighted sum of shiftable kernels ψ_l .

Due to shift-invariance, the dictionary Φ is the concatenation of L Toeplitz matrices [3] and is overcomplete. Since $M > N$, the dictionary is redundant and the linear system is thus under-determined and has multiple solutions. The introduction of constraints such as *sparsity* allows the solution to be regularized. The sparse approximation selects only K active atoms among the M possible and computes the associated coefficients vector \mathbf{x} to have the better approximation of the signal y . One way to formalize the sparse approximation is:

$$\min_{\mathbf{x}} \left\| y(t) - \sum_{l=1}^L \sum_{\tau \in \sigma_l} x_{l,\tau} \psi_l(t - \tau) \right\|^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_0 \leq K, \quad (7)$$

where $K \ll M$ is a constant and $\|\mathbf{x}\|_0$ is the number of nonzero elements of vector \mathbf{x} . But this problem is NP-hard [8]. So, non-convex pursuits tackle sequentially it such as matching pursuit (MP) [9]. The orthogonal matching pursuit (OMP) [10] assures that coefficients \mathbf{x} are the orthogonal projection of the signal over the selected atoms. Using only K active atoms among the M possible, sparsity provides the compactness so much requested by [1]. From the beginning of this paragraph, explanations have been given for univariate signals. But they are extended to trivariate signals by the multivariate OMP (M-OMP) [3] which makes the atoms choice and the coefficients estimation. The introduced model (3) allows atoms to rotate but needs an appropriate approximation method to estimate the associated rotation matrices besides.

¹ Note that $a(t)$ and $a(t - t_0)$ do not represent samples, but the signal a and its translation of t_0 samples.

Now combining shift and 3D rotation invariances problems, we obtain the following problem to solve:

$$\min_{\mathbf{x}, \mathbf{R}} \left\| y(t) - \sum_{l=1}^L \sum_{\tau \in \sigma_l} x_{l,\tau} R_{l,\tau} \psi_l(t-\tau) \right\|^2$$

$$s.t. \quad \|\mathbf{x}\|_0 \leq K \text{ and } \forall l \in \mathbb{N}_L, \forall \tau \in \sigma_l, R_{l,\tau} R_{l,\tau}^T = Id. \quad (8)$$

More than Eq.(4), Eq.(8) is the real issue that is addressed in this article. Eq.(8) combines Eq.(4) which we ignore if an analytic solution exists and Eq.(7) which is NP-hard. We propose a non-convex optimization to solve this particularly hard problem.

Note that 2DRI-OMP [3] simply tackles Eq.(8) in the 2D case. The presented article can be viewed as a non-trivial 3D extension (without orthogonal projection), that explains the name of the method presented.

3 3D Rotation Invariant Matching Pursuit

In this section, our proposed sparse 3DRI decomposition algorithm is introduced. We first detail the chosen method for the 3D registration, which will be the core of the introduced algorithm. Then, a non-convex optimization based on MP principle is introduced to solve Eq.(8) and is called 3DRI-MP.

3.1 3D Registration by SVD

Registration problem (5) is considered here with a normed trivariate pattern $\phi \in \mathbb{R}^{3 \times N}$, but without spatial translation. The sought parameters are the rotation R and the scale factor x :

$$\min_{x, R} \|y - x R \phi\|^2 \text{ s.t. } R R^T = Id. \quad (9)$$

For solving this 3D registration problem, the SVD method is chosen among the other possible methods because it is the cheapest and it simply deals with the particular cases of noise and planar patterns [4].

The method chosen is described in Algorithm 1. After having computed the correlation matrix $M_c = y \phi^T \in \mathbb{R}^{3 \times 3}$ (step 1), its SVD is carried out: $(U, \Lambda_1, V) = SVD(M_c)$ (step 2). Defining matrix Λ_2 such that $\Lambda_2 = \text{diag}(1, 1, \det(UV^T))$, the optimal rotation is: $R = U \Lambda_2 V^T$ (step 3). The correlation value which provides the scale factor is computed such that: $x = \text{trace}(R \phi y^T) \geq 0$ (step 4).

Algorithm 1 : $(\mathbf{x}, \mathbf{R}) := \text{Reg_SVD}(\phi, y)$

```

begin
1:   Mc := y*phi^T ;
2:   (U,Lambda_1,V) := SVD(Mc) ;
3:   R := U*Lambda_2*V^T ;
4:   x := trace(R*phi*y^T) ;
end.
```

3.2 3DRI-MP Description

In this section, the 3DRI-MP is going to be explained step by step. A trivariate signal $y \in \mathbb{R}^{3 \times N}$ and a dictionary $\Psi = \{\psi_l\}_{l=1}^L$ of shiftable trivariate normed kernels are considered. Given this redundant trivariate dictionary, 3DRI-MP produces a sparse approximation of the signal y (Algorithm 2).

The initialization (step 1) allocates the studied signal y to the residue ϵ^0 . At the current iteration k , the algorithm selects the atom that produces the absolute strongest decrease in the mean square error (MSE) $\|\epsilon^{k-1}\|^2$. This is equivalent to finding the registered atom that is the most correlated to the residue ϵ^{k-1} (see Appendix). The correlation value $x_{l^k, \tau^k}^k = \text{trace}(R_{l^k, \tau^k}^k \psi_{l^k}(t - \tau^k) \epsilon^{k-1}(t)^T)$ is computed for each shift τ , with R_{l^k, τ^k}^k the optimal rotation matrix to register $\psi_l(t - \tau)$ on $\epsilon^{k-1}(t)$. To carry out this step, algorithm Reg_SVD is applied for each τ and each $l=1..L$ (step 5), and then, the maximum of the values $x_{l^k, \tau^k}^k (\geq 0)$ is searched for to select the optimal atom (step 7), characterized by its kernel index l^k and its position τ^k . Selected atoms form an active dictionary. The vector x accumulates the active (*i.e.* nonzero) coefficients that are the maximum correlation values (step 8). Associated rotation matrices are grouped in R (step 9) and the current residue is computed (step 10).

A threshold on k the number of iterations or a threshold on the relative root MSE (rRMSE) $\|\epsilon^k\|/\|y\|$ can be used as stopping criteria (step 12). In the end, the 3DRI-OMP provides a K -sparse approximation of y using the K selected active elements:

$$\hat{y}^K = \sum_{k=1}^K x_{l^k, \tau^k}^k R_{l^k, \tau^k}^k \psi_{l^k}(t - \tau^k). \quad (10)$$

Without considering the nonconvexity of the algorithm, if there is no overlap between the selected atoms, 3DRI-MP gives the orthogonal projection of the signal on the active dictionary in Eq. (8). Otherwise, it is suboptimal since atom overlaps generate cross terms that are not treated by 3DRI-MP.

Algorithm 2 : (x,R) := 3DRI_MP (y,Psi)

```

begin
1:   initialization: k:=1, epsilon^0:=y, x:=[], R:=[] ;
2:   repeat
3:     for l = 1 .. L
4:       3D Registration for each tau:
5:       (x^k_{l,tau}, R^k_{l,tau}) :=
           Reg_SVD (psi_l(t-tau), epsilon^{k-1}(t)) ;
6:     end for
7:     (l^k,tau^k) := arg max_{l,tau} x^k_{l,tau} ;
8:     x := x U x^k_{l^k,tau^k} ;
9:     R := R U R^k_{l^k,tau^k} ;
10:    epsilon^k := epsilon^{k-1}
           - x^k_{l^k,tau^k} * R^k_{l^k,tau^k} * psi_{l^k}(t-tau^k) ;
11:    k := k + 1 ;

```

```
12: until stopping criterion
end.
```

4 Illustration on Real Data

An application of 3DRI-MP on real data is shown in this section. Studied data are motion signals of French cued speech, which is a gestural language to complement speech reading [11]. This language associates speech articulation to cues formed by the hand. To make the acquisition, retroreflective markers are put on the hand of a skilled cuer. Data are acquired by an optical system which records the 3D coordinates of the markers. At the end of the acquisition, tricomponent coordinates are obtained for each marker, and we focus on the one located on the top on the thumb. Velocity signals $v = [v_x ; v_y ; v_z]^T$ are the inputs of the 3DRI-MP, and a dictionary of $L = 6$ kernels is designed to be adapted to such data. These few kernels represent the main motion primitives.

The original velocity signal v is plotted in Fig. 2 (*top*), and is composed of the three velocity components v_x (*solid blue line*), v_y (*dashed red*) and v_z (*dotted green*). This signal is processed by 3DRI-MP with $K = 10$ iterations and gives the approximated signal \hat{v} plotted in Fig. 2 (*bottom*). The rRMSE of this approximation is 28.8 %. This error is quite high, but the goal is to decompose the signal on its main motion primitives, and not to code all the variabilities with numerous small coefficients.

Velocity signals are integrated to have a more visual representation of the data. The original trajectory associated to the studied signal v is plotted in Fig. 3 (*top left*). Now, we are interested by the contributions of the largest atoms. The trajectory is reconstructed using the contributions of the $K = 5$

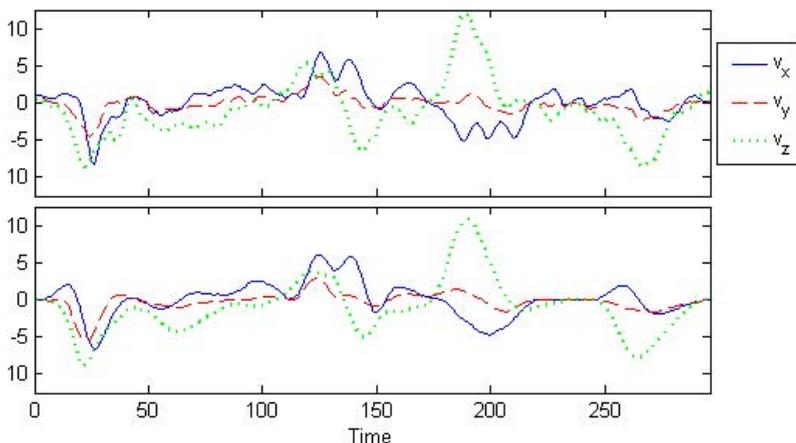


Fig. 2. The original signal (*top*) and its approximation with $K = 10$ atoms (*bottom*). Each signal is composed of the three velocity components v_x (*solid blue line*), v_y (*dashed red*) and v_z (*dotted green*)

largest coefficients of the 3DRI-MP decomposition. The reconstruction is plotted in Fig. 3 (*top right*), showing each of the 5 atoms of the approximation. The stars represent the beginning of the kernels trajectories. In Fig. 3 (*top right*), we remark that kernel $l = 6$ (*dashed red line*) is employed three times with different orientations (and with different coefficients and shifts). This shows the 3D rotation invariance of the 3DRI-MP which provides a good matching of the studied trajectory allowing the kernels rotation. The original trajectory is now randomly rotated in Fig. 3 (*bottom left*), and it is reconstructed with its $K = 5$

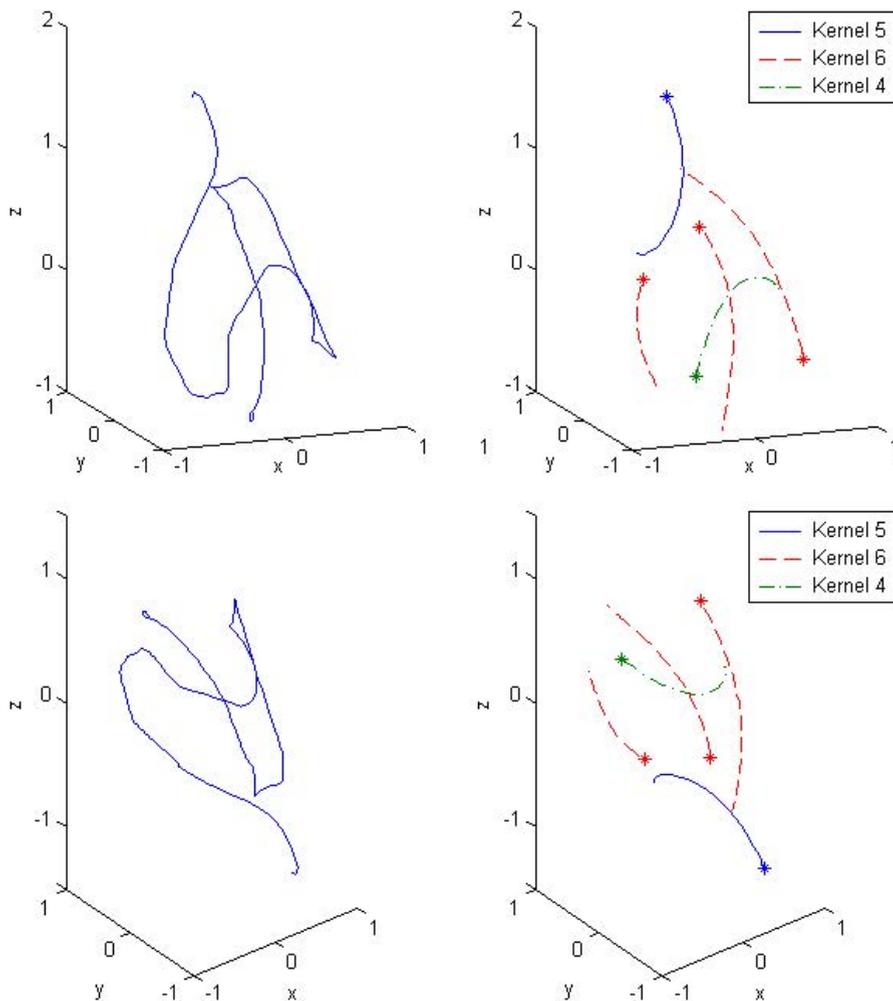


Fig. 3. The original (*resp.* rotated) trajectory (*top left*) (*resp.* (*bottom left*)) and its reconstruction with the contributions of the $K = 5$ largest atoms (*top right*) (*resp.* (*bottom right*))

largest coefficients in Fig. 3 (*bottom right*). Reconstructions of Fig. 3 (*top right*) and (*bottom right*) are similar, taking into account the rotation: this highlights the rotation robustness of the 3DRI-MP.

5 Comparisons on Simulation Data

In this section, in order to evaluate its performances, 3DRI-MP is applied to simulation data and compared to M-OMP used in a trivariate case. Remark that it is not possible to compare 3DRI-MP to algorithm dealing with model (1) mentioned in Introduction since it is too much different.

In the first experiment, a dictionary Ψ of $L = 50$ trivariate kernels is randomly created: kernels are drawn from white Gaussian noise and they are then normed. The kernels length is $T = 65$ samples. One hundred signals of $N = 1600$ samples are composed of the sum of $K = 15$ atoms, for which the coefficients (strictly positive), the rotation matrices and the kernels indices are randomly drawn on a uniform distribution. Shift parameters are drawn in a way that atoms do not overlap. Each signal is approximated by 3DRI-MP and M-OMP with $K = 15$ iterations in order to recover the 15 atoms. The rRMSE $\|\epsilon^k\|/\|y\|$ is averaged (mean and standard deviation) over the 100 signals and is plotted in Fig. 4 (*left*) as a function of the inner iterations $k = 1..K$ of the two algorithms. We observe that 3DRI-MP gives better approximation performances than M-OMP. At the end of the K iterations, 3DRI-MP has a rRMSE of 0% whereas M-OMP has a rRMSE of 92.0%. The rRMSE of M-OMP is huge since it is not able to recover the good atoms. It shows that this algorithm is not appropriate for rotated data. This experiment highlights the relevance of 3DRI algorithm for the decomposition of revolved data.

The second experiment is close to the first one. The difference is that the signals length is now $N = 250$, and the shift parameters are randomly drawn

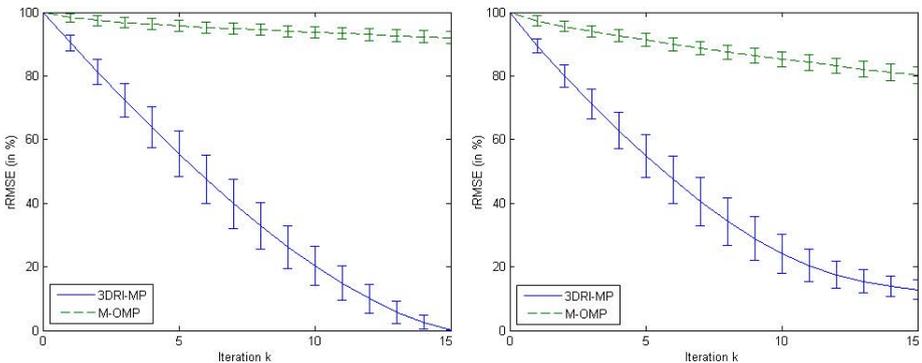


Fig. 4. Comparisons between the performances of 3DRI-MP and M-OMP without (*left*) and with (*right*) atoms overlaps. The rRMSE, averaged over 100 signals, is plotted as a function of the inner iteration k .

on a uniform distribution. As a consequence, the different atoms overlap. In the same way, the averaged rRMSE is plotted in Fig. 4 (*right*) as a function of the inner iterations k . The final rRMSE is 12.8% for 3DRI-MP, that is worse than the first experiment. In fact, when atoms overlap, 3DRI-MP which does not provide the orthogonal projection of Eq. (8) since it does not take into account the cross terms due to overlaps. This experiment shows the need to improve the 3DRI decomposition method. Concerning M-OMP, the final rRMSE is 80.2%, so performances seem to be improved. In fact, it does not recover the atoms better than in the first experiment. Signal energy is more compact since atoms overlap, so M-OMP approximates more energy at each iteration.

6 Conclusion and Prospects

This article has proposed a new model for describing a time-varying 3D object as the sum of rotatable 3D patterns. The considered model combines the 3D rotation invariance (3DRI) and the shift-invariance of the patterns. The introduced 3DRI-MP solves the 3DRI decomposition problem estimating the active atoms, their coefficients, their matrices and their shifts. It has been illustrated on real data and evaluated on simulation data. There are multiple applications in various domains: non-rigid structure-from-motion, 3D curve matching, 3D tracking, gesture representation and analysis and all other processings based on 3DRI decomposition.

The considered prospects are to improve the 3DRI decomposition method for coping with overlaps, to present a dictionary learning algorithm able to learn a kernels dictionary with model (3) and to add a classification step to make gesture recognition.

Acknowledgements. The authors thank F. Elisei from GIPSA-DPC for his explanations about cued speech data (acquired during the RNRT 01/37 ARTUS project).

References

1. Akhter, I., Sheikh, Y., Khan, S., Kanade, T.: Trajectory space: A dual representation for nonrigid structure from motion. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 33, 1442–1456 (2011)
2. Bregler, C., Hertzmann, A., Biermann, H.: Recovering non-rigid 3D shape from image streams. In: *Proc. Computer Vision and Pattern Recognition, CVPR* (2000)
3. Barthélemy, Q., Larue, A., Mayoue, A., Mercier, D., Mars, J.: Shift & 2D rotation invariant sparse coding for multivariate signals. *IEEE Trans. on Signal Processing* 60, 1597–1611 (2012)
4. Eggert, D., Lorusso, A., Fisher, R.: Estimating 3-D rigid body transformations: a comparison of four major algorithms. *Machine Vision and Applications* 9, 272–290 (1997)
5. Gower, J., Dijkstra, G.: *Procrustes Problems*. Oxford Statistical Science Series (2004)

6. Bergevin, R., Soucy, M., Gagnon, H., Laurendeau, D.: Towards a general multi-view registration technique. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 18, 540–547 (1996)
7. Besl, P., McKay, H.: A method for registration of 3-D shapes. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 14, 239–256 (1992)
8. Davis, G.: Adaptive Nonlinear Approximations. PhD thesis, New York University (1994)
9. Mallat, S., Zhang, Z.: Matching pursuits with time-frequency dictionaries. *IEEE Trans. on Signal Processing* 41, 3397–3415 (1993)
10. Pati, Y., Rezaifar, R., Krishnaprasad, P.: Orthogonal Matching Pursuit: recursive function approximation with applications to wavelet decomposition. In: *Proc. Asilomar Conf. on Signals, Systems and Comput.* (1993)
11. Gibert, G., Bailly, G., Beautemps, D., Elisei, F., Brun, R.: Analysis and synthesis of the three-dimensional movements of the head, face, and hand of a speaker using cued speech. *Journal of Acoustical Society of America* 118, 1144–1153 (2005)

Appendix: Selection Step of 3DRI-MP

Defining the error as: $\epsilon = y - \sum_{m=1}^M x_m R_m \phi_m$, the derivation of the criterion $J = \|\epsilon\|^2 = \text{trace}(\epsilon\epsilon^T)$ with respect to x_m gives:

$$\frac{\partial J}{\partial x_m} = -2 \text{trace}(R_m \phi_m \epsilon^T) = -2 \langle R_m \phi_m, \epsilon \rangle , \quad (11)$$

and in the shift-invariant formalism, it provides the selection step of Section 3.2:

$$-\frac{1}{2} \frac{\partial \|\epsilon\|^2}{\partial x_{l,\tau}} = \text{trace}(R_{l,\tau} \psi_l(t - \tau) \epsilon^T) = \langle R_{l,\tau} \psi_l(t - \tau), \epsilon \rangle . \quad (12)$$