

Bilinear modelling via Augmented Lagrange Multipliers (BALM)

Alessio Del Bue, *Member, IEEE*, João Xavier, *Member, IEEE*,
 Lourdes Agapito, *Member, IEEE* Marco Paladini, *Member, IEEE*

Abstract—This paper presents a unified approach to solve different bilinear factorization problems in computer vision in the presence of missing data in the measurements. The problem is formulated as a constrained optimization where one of the factors must lie on a specific manifold. To achieve this, we introduce an equivalent reformulation of the bilinear factorization problem that decouples the core bilinear aspect from the manifold specificity. We then tackle the resulting constrained optimization problem via Augmented Lagrange Multipliers. The strength and the novelty of our approach is that this framework can handle seamlessly different computer vision problems. The algorithm is such that only a projector onto the manifold constraint is needed. We present experiments and results for some popular factorization problems in computer vision such as rigid, non-rigid and articulated Structure from Motion; photometric stereo and 2D-3D non-rigid registration.

Index Terms—Bilinear Optimization, Augmented Lagrangian, SfM, Photometric Stereo, Image Registration.



1 INTRODUCTION

Many computational problems in Computer Vision fit the form of bilinear problems since often observations are influenced by two independent factors where each can be described by a linear model. These bilinear models are highly representative of a set of problems not only restricted to the Computer Vision domain. Table 1 provides a list of some common bilinear estimation problems in the literature.

To cite a few classical examples, in photometric stereo [1] the shape of the object and the light source direction interact bilinearly to influence the image intensity. In rigid structure from motion [2] the 3D shape of the object is pre-multiplied by the camera matrix to determine its image coordinates. In structure from sound the time of arrival of a sound event depends both on the direction of the sound propagation and the position of the microphones [3]. In facial tracking the problem of separating head pose and facial expression can also be defined as a bilinear problem [4]. In non-rigid structure from motion [5] the 2D coordinates of features arise from a bilinear

relation between the camera matrix and the time varying shape. The interaction between two factors has also been generalized to several problems [6] in the machine learning field, perhaps the closest computational interpretation of Schlick’s philosophical analysis [7] asserting that content can be only formed and expressed through the subjectivity of the communication. In all these problems the objective is to make inferences about both factors – the goal is their simultaneous estimation.

In this paper, we present a unified approach to solve different bilinear factorization problems in computer vision. The problem is formulated as a constrained optimization where one of the factors must lie on a specific manifold. Our key observation is that the main difference amongst factorization problems is the manifold on which the solution lies. Thus, intuitively, it should be possible to construct a unified optimization framework in which a change of the manifold constraint just implies replacing an inner module of the algorithm (as opposed to an overall redesign of the optimization method from scratch). In this paper, we propose such a modular approach. To achieve this, we start by introducing an equivalent reformulation of the bilinear factorization problem. In loose terms, the reformulation decouples the core bilinear aspect from the manifold specificity. We then tackle the resulting constrained optimization problem via an Augmented Lagrange Multipliers (ALM) iterative algorithm. The mechanics of our algorithm are such that only a projector onto the manifold constraint is needed. That is the strength and the novelty of our approach: it can handle seamlessly different computer vision factorization problems. What will differ in each case is the projector of the solution onto the correct manifold. If the manifold projector exists, the factorization problem can be formulated using our unified approach.

Our approach is designed to be able to deal with

-
- *Alessio Del Bue is with the Istituto Italiano di Tecnologia (IIT), Via Morego 30, 16163 Genova, Italy. E-mail: alessio.delbue@iit.it*
 - *João Xavier is with the Institute for Systems and Robotics (ISR), Instituto Superior Técnico (IST), Technical University of Lisbon, Portugal.*
 - *Lourdes Agapito and Marco Paladini are with the School of Electronic Engineering and Computer Science (EECS) at Queen Mary University of London. Mile End Road, London E1 4NS, UK.*

This research has received funding from the European Research Council under the European Community’s Seventh Framework Programme (FP7/2007-2013) ERC Starting Grant agreement 204871-HUMANIS. The work of J. Xavier was partially supported by FCT grant CMU-PT/SIA/0026/2009 and by ISR/IST plurianual funding (POSC program, FEDER). We thank J. Buenaposada for providing an implementation of [1]. We gratefully acknowledge R. Proietti for the hardware support in the large scale 3D reconstruction experiments. The box and head data for the articulated test was given by P. Tresadern. The data for the Casa da Musica 3D reconstruction was provided by M. Marques and J. Costeira.

TABLE 1
Different problems and their bilinear forms $Y = S M$.

Bilinear Problem	S	M	Manifold Constraints
Rigid SfM [2]	3D points position	Camera motion	Stiefel Matrix
Non-Rigid SfM [5]	3D basis shapes	Camera motion and deformations	Stiefel Matrix
Articulated SfM [8]	Two 3D bodies	Camera motion and joint property	Stiefel and sub-Stiefel Matrices
Surface unfolding problem [9]	Template	Isometry operator	sub-Stiefel Matrix
Photometric Stereo [1]	Lighting direction	Albedo and surface normals	Normal constraints
Bilinear 2D-3D registration [10]	Affine camera	Deformation weights	Normal constraints
Non-Rigid SfM with DCT bases [11]	Camera motion	Deformation weights	Stiefel Matrix
Structure from Sound [3]	Microphone location	Sound direction	None
Head pose and facial expression [4]	Expression	Pose	None
Face identity and facial expression [12]	Identity	Expressions	None
Face identity and viewpoint [13]	Identity	Pose	None
Face identity and illumination [14]	Identity	Illumination	None
Emotion and speech [15]	Viseme	Expression	None
{identity, action, viewpoint } [16]	{identity, action}	{view, identity}	Case dependent
Nonlinear style and content [17]	Content	Style	Custom (learned)
Factorization of BRDFs [18]	Lighting/Viewpoint	Surface property	Stiefel

missing data. For instance in the structure from motion scenario this is the case when not all the points are visible in all the views due to occlusions or broken tracks, and in photometric stereo shadows, specularities or saturated pixels are treated as missing data. In our experiments we show that we are able to deal with high percentages of missing data which has the practical implication that our approach can be used on data coming from real, not just controlled, scenarios. We illustrate our unified approach by applying it to solve several popular computer vision problems: rigid, articulated and non-rigid structure from motion; photometric stereo (PS) and non-rigid registration. To the best of our knowledge, this paper constitutes the first attempt to propose a unified optimization framework for large scale bilinear factorization problems with given manifold constraints on one of the factors and provide a practical algorithm that can deal with missing data in the measurements.

2 RELATED WORK

Bilinear models appear frequently in Computer Vision. However, it is in the area of Structure from Motion (SfM) where most of the efforts dedicated to solve this problem have come from. Stemming from Wiberg’s original algorithm [19], the first to deal with missing data in SfM¹, matrix factorization in the presence of missing data continues to be an open problem that attracts significant attention and many different solutions have been since put forward. The wealth of research in this area is such that we cannot give an exhaustive review. Instead, we will focus on describing the most important threads of research to solve the problem of low-rank matrix factorization in the case of missing data and cite representative examples of each group of approaches. Buchanan and

Fitzgibbon [21] provided a comprehensive review of these methods classifying them into several categories.

One group of approaches propose strategies that combine partial low-rank factorizations obtained for sub-blocks of the measurement matrix that contain full data. Pioneered by Jacobs [22], these *batch approaches*, reconstruct the measurement matrix by first building its row or column null-space or one of its range spaces and have been applied both to the rigid [22] and non-rigid [23] SfM problems. However, one concern about these approaches is their sub-optimality and that their performance degrades in the presence of noise.

A second group of missing data approaches that dominates the literature includes iterative methods that perform alternation of closed form solutions to solve for the two factors of the matrix [24], [25], [26], [27]. For instance, Powerfactorization [26] uses an alternated least-squares (ALS) scheme to solve for the motion and shape matrices. Alternation constitutes an attractive scheme since it guarantees convergence to a local minimum as the objective function is reduced after each iteration. Further advantages are quick iteration steps combined with a fast convergence rate in the initial iterations. However, after a few iterations, the convergence rate drops and these algorithms are susceptible to flatlining.

Finally non-linear optimization algorithms have also been proposed to optimize directly the cost function. Buchanan and Fitzgibbon [21] proposed a Damped Newton algorithm that provides faster and more accurate solutions than standard alternation approaches. Later, Chen revitalized the use of the Levenberg-Marquardt algorithm to solve the missing data problem by formulating the low-rank matrix approximation problem as a minimization on subspaces [28]. The idea is to consider the shape as an implicit function of the motion and measurement matrices and solve only for the motion. This results in a smaller system to be solved in every iteration, which makes this method well suited to large matrices where it outperforms [19] and [21].

1. Often misunderstood in the literature as an Alternate Least Squares (ALM) algorithm, Wiberg’s algorithm [19] has seen a recent revitalization due to Okatani and Deguchi [20]

Although these non-linear methods do exhibit a superior performance, proper initialization remains an open problem and, more importantly, integrating additional constraints in the optimization process is not an easy task. The common drawback of all the methods described above is that they only solve the low-rank matrix factorization problem without imposing manifold constraints. These are applied afterwards, once the low-rank matrix has been estimated. Crucially, the constraints are not imposed during the minimization.

On the other hand, a relatively recent set of algorithms have attempted to solve the problem by including explicitly the non-linear constraints given by the specific problem structure in the low-rank minimization. Marques and Costeira [29] introduced the concept of *motion manifold* in rigid SfM to obtain motion matrices that exactly satisfy the camera constraints. Similarly, Paladini et al. [30] propose an alternation algorithm associated with an optimal projector onto the *motion manifold* of non-rigid shapes. The practical implication of their algorithm is that it can deal with very high percentages of missing data. Shaji et al. [31] also propose to solve a non-linear optimization problem directly on the product manifold of the Special Euclidean Group claiming better results than [21] in a rigid real sequence.

However, all these approaches are tailored to specific problems. Therefore, for different manifold constraints an overall redesign of the optimization method would be needed. The purpose of our work is to present a generic approach that is not problem dependent. In similar spirit, Chandraker and Kriegman [10] have proposed a globally optimal bilinear fitting approach for general Computer Vision problems. The key contribution of their approach is that they can prove convergence to a global minimizer using a branch and bound approach. Their insight is that in many bilinear problems in computer vision the size of one set of variables in the relation is much smaller than the size of the other one. Exploiting this property allows them to solve the problem by restricting the branching dimensions to just one set of variables (the smaller one). However, the main drawback is that the scenarios to which their method can be applied are restricted to simple bilinear problems where the number of variables in one of the sets must be very small (for instance just 9 variables in one of their examples).

Our Bilinear factorization via Augmented Lagrange Multipliers (BALM) is designed to deal with large-scale optimization problems with the inclusion of non-linear constraints. Our approach is not the first one to adopt the Augmented Lagrangian Multipliers (ALM) framework in the Computer Vision or related contexts. In perspective 3D reconstruction [32] ALM was used to enforce constraints on the perspective depths. In [33] ALM is successfully employed as a single matrix imputation algorithm which can deal with large scale problems.

3 PROBLEM STATEMENT

We denote by $Y \in \mathbb{R}^{n \times m}$ the measurement matrix. In this paper, we consider the general case of missing data. We let the finite set $\mathcal{O} := \{(i, j) : Y_{ij} \text{ is observed}\}$ enumerate the indices of the entries of Y which are available. The bilinear factorization problem we address is the following constrained optimization problem:

$$\begin{aligned} & \text{minimize} && \sum_{(i,j) \in \mathcal{O}} (Y_{ij} - s_i^\top m_j)^2 \\ & \text{subject to} && M_i \in \mathcal{M}, \quad i = 1, \dots, f, \end{aligned} \quad (1)$$

where s_i^\top denotes the i th row of the matrix $S \in \mathbb{R}^{n \times r}$ and m_j denotes the j th column of the matrix $M = [M_1 \ \dots \ M_i \ \dots \ M_f] \in \mathbb{R}^{r \times m}$, $M_i \in \mathbb{R}^{r \times p}$. Note that we are using a dummy variable i twice in (1): in the cost function (i.e., $(i, j) \in \mathcal{O}$) and in the constraints (to enumerate the submatrices M_i of M).

The variable to optimize in (1) is (S, M) . Here, f is the number of frames and we consider throughout the paper that $n \geq r \geq p$. For instance in the structure from motion problem S would be the 3D structure and M the camera matrices, in photometric stereo S would be the lighting parameters and M the surface normals and albedo.

In words, problem (1) consists in finding the best rank r factorization SM of Y , given the available entries enumerated by \mathcal{O} and subject to the constraints on M . More precisely, each submatrix $M_i \in \mathbb{R}^{r \times p}$ of M must belong to the manifold $\mathcal{M} \subset \mathbb{R}^{r \times p}$. Our aim in this paper is to construct an algorithm to solve problem (1) which takes advantage of the fact that the projector onto \mathcal{M} is available (easily implementable). That is, we assume that, for a given $A \in \mathbb{R}^{r \times p}$, it is known how to solve the projection problem onto \mathcal{M}

$$\begin{aligned} & \text{minimize} && \|A - X\|^2, \\ & \text{subject to} && X \in \mathcal{M} \end{aligned} \quad (2)$$

where $\|X\|$ denotes the Frobenius norm of X . In the sequel, we let $p_{\mathcal{M}}(A)$ denote a solution of (2).

Problem reformulation. Let us define a new set of variables $z := \{Z_{ij} : (i, j) \notin \mathcal{O}\}$. Think of them as representing the non-observed entries of Y . We can introduce these variables in (1) and obtain the following equivalent optimization problem

$$\begin{aligned} & \text{minimize} && \|Y(z) - SM\|^2 \\ & \text{subject to} && M_i \in \mathcal{M}, \quad i = 1, \dots, f, \end{aligned} \quad (3)$$

where the (i, j) entry of the matrix $Y(z)$ is defined as

$$(Y(z))_{ij} := \begin{cases} Y_{ij} & , \text{ if } (i, j) \in \mathcal{O} \\ Z_{ij} & , \text{ if } (i, j) \notin \mathcal{O} \end{cases}.$$

In words, $Y(z)$ is just Y where we fill-in the missing entries with z . Note that the variable to optimize in (3) is (z, S, M) . Problem (3) is equivalent to (1) because once we fix (S, M) in (3) and minimize over z we fall back into (1). Finally, we clone M into a new variable $N = [N_1 \ \dots \ N_i \ \dots \ N_f] \in \mathbb{R}^{r \times m}$, $N_i \in \mathbb{R}^{r \times p}$, and transfer the manifold constraint to the latter. By

doing so, we roughly separate the bilinear issue from the manifold restriction. This is our final reformulation:

$$\begin{aligned} & \text{minimize} && \|Y(z) - SM\|^2 \\ & \text{subject to} && M_i = N_i, \quad i = 1, \dots, f \\ & && N_i \in \mathcal{M}, \quad i = 1, \dots, f. \end{aligned} \quad (4)$$

The variable to optimize in (4) is (z, S, M, N) . We note that duplicating the variable M into N and attaching the manifold constraint to the latter can be also interpreted as variable splitting; see [34].

4 THE BALM ALGORITHM

The main difficulty in the constrained optimization problem (4) are the equality constraints $M_i = N_i$. We propose to handle them through an augmented Lagrangian approach, see [35], [36] for details on this optimization technique. In our context, the augmented Lagrangian corresponding to (4) is given by

$$L_\sigma(z, S, M, N; R) = \|Y(z) - SM\|^2 - \sum_{i=1}^f \text{tr}(R_i^\top (M_i - N_i)) + \frac{\sigma}{2} \sum_{i=1}^f \|M_i - N_i\|^2. \quad (5)$$

where $\sigma > 0$ is the weight of the penalty term and R_i , $i = 1, \dots, f$, denote Lagrange multipliers. We let $R = [R_1 \dots R_f]$. The optimization problem (4) can then be tackled by our Bilinear factorization via Augmented Lagrange Multipliers (BALM) algorithm detailed in Algorithm 1.

Algorithm 1 Bilinear factorization via Augmented Lagrange Multipliers (BALM)

- 1: set $k = 0$ and $\epsilon_{\text{best}} = +\infty$
 - 2: initialize $\sigma^{(0)}$, $R^{(0)}$, $\gamma > 1$ and $0 < \eta < 1$
 - 3: initialize $z^{(0)}$, $S^{(0)}$ and $M^{(0)}$
 - 4: **repeat**
 - 5: solve

$$\begin{aligned} & (z^{(k+1)}, S^{(k+1)}, M^{(k+1)}, N^{(k+1)}) = \\ & = \text{argmin} \quad L_{\sigma^{(k)}}(z, S, M, N; R^{(k)}) \\ & \text{subject to} \quad N_i \in \mathcal{M}, \quad i = 1, \dots, f, \end{aligned} \quad (6)$$
 - 6: using the iterative Gauss-Seidel scheme described in **Algorithm 2**
 - 7: compute $\epsilon = \|M^{(k+1)} - N^{(k+1)}\|^2$
 - 8: **if** $\epsilon < \eta \epsilon_{\text{best}}$
 - 9: $R^{(k+1)} = R^{(k)} - \sigma^{(k)} (M^{(k+1)} - N^{(k+1)})$
 - 10: $\sigma^{(k+1)} = \sigma^{(k)}$
 - 11: $\epsilon_{\text{best}} = \epsilon$
 - 12: **else**
 - 13: $R^{(k+1)} = R^{(k)}$
 - 14: $\sigma^{(k+1)} = \gamma \sigma^{(k)}$
 - 15: **endif**
 - 16: update $k \leftarrow k + 1$
 - 17: **until** some stopping criterion
-

Regarding the initialization of the BALM algorithm, we used $\sigma^{(0)} = 1$, $R^{(0)} = 0$, $\gamma = 5$ and $\eta = 1/2$ in all our computer experiments. With respect to $z^{(0)}$, $S^{(0)}$ and $M^{(0)}$, we feel that there is no universally good method, that is, the structure of \mathcal{M} must be taken into account. We discuss the initialization $(z^{(0)}, S^{(0)}, M^{(0)})$ for several examples in the experimental section of this paper.

Clearly, solving the inner problem (6) at each iteration of the BALM method is the main computational step. Note that in (6) the optimization variable is (z, S, M, N) ($\sigma^{(k)}$ and $R^{(k)}$ are constants). To tackle (6) we propose an iterative Gauss-Seidel scheme which is described in Algorithm 2. For simplicity, we run a fixed number of iterations (L_{max}) in Algorithm 2. However, adopting other stopping criteria might be more efficient and boost the overall BALM algorithm speed (we leave this issue for future work). We now show that each of the subproblems (7), (8) and (9) inside the Gauss-Seidel scheme are easily solvable.

Algorithm 2 Iterative Gauss Seidel scheme to solve for (6)

- 1: set $l = 0$ and choose L_{max}
 - 2: set $z^{[0]} = z^{(k)}$, $S^{[0]} = S^{(k)}$ and $M^{[0]} = M^{(k)}$
 - 3: **repeat**
 - 4: solve

$$\begin{aligned} & N^{[l+1]} = \\ & = \text{argmin} \quad L_{\sigma^{(k)}}(z^{[l]}, S^{[l]}, M^{[l]}, N; R^{(k)}) \\ & \text{subject to} \quad N_i \in \mathcal{M}, \quad i = 1, \dots, f, \end{aligned} \quad (7)$$
 - 5: solve

$$\begin{aligned} & (S^{[l+1]}, M^{[l+1]}) = \\ & = \text{argmin} \quad L_{\sigma^{(k)}}(z^{[l]}, S, M, N^{[l+1]}; R^{(k)}) \end{aligned} \quad (8)$$
 - 6: solve

$$\begin{aligned} & z^{[l+1]} = \\ & = \text{argmin} \quad L_{\sigma^{(k)}}(z, S^{[l+1]}, M^{[l+1]}, N^{[l+1]}; R^{(k)}) \end{aligned} \quad (9)$$
 - 7: update $l \leftarrow l + 1$
 - 8: **until** $l = L_{\text{max}}$
 - 9: set $S^{(k+1)} = S^{[L_{\text{max}]}$, $M^{(k+1)} = M^{[L_{\text{max}]}$ and $N^{(k+1)} = N^{[L_{\text{max}]}$
-

Solving for (7), (8) and (9). Problem (7) requires a minimization over $N_i \in \mathcal{M}$, $i = 1, \dots, f$, the remaining variables being constant. Thus, by using (5) and dropping the constant terms, problem (7) becomes equivalent to

$$\begin{aligned} N^{[l+1]} = \text{argmin} \quad & \sum_{i=1}^f \left\| N_i - \left(M_i^{[l]} - \frac{1}{\sigma^{(k)}} R_i^{(k)} \right) \right\|^2 \\ & \text{subject to} \quad N_i \in \mathcal{M}, \quad i = 1, \dots, f \end{aligned} \quad (10)$$

That is, problem (7) decouples into f projections onto the manifold of constraints \mathcal{M} . More precisely, if we

partition

$$N^{[l+1]} = \left[N_1^{[l+1]} \quad \dots \quad N_i^{[l+1]} \quad \dots \quad N_f^{[l+1]} \right] \in \mathbb{R}^{r \times m},$$

with $N_i^{[l+1]} \in \mathbb{R}^{r \times p}$. The solution of (7) is given by

$$N_i^{[l+1]} = p_{\mathcal{M}} \left(M_i^{[l]} - \frac{1}{\sigma^{(k)}} R_i^{(k)} \right), \quad i = 1, \dots, f. \quad (11)$$

We recall that $p_{\mathcal{M}}$ stands for the projector onto \mathcal{M} , see (2), which we assume is available. This is the only part of the algorithm where the constraint manifold \mathcal{M} plays a role. Thus, replacing \mathcal{M} amounts to replace the projector $p_{\mathcal{M}}$. This is the modularity which we alluded to previously.

To simplify notation, we let $Y = Y(z^{[l]})$, $N = N^{[l+1]}$, $\sigma = \sigma^{(k)}$ and $R = R^{(k)}$. Solving (8) corresponds to solving

$$\text{minimize} \quad \|Y - SM\|^2 + \frac{\sigma}{2} \sum_{i=1}^f \|M_i - (N_i + \frac{1}{\sigma} R_i)\|^2.$$

It can be solved as described in [37, Sec. 4.2] or further decoupled in two least-squares problems, first over M (fixed S) and then over S (fixed M). In fact, in all our numerical experiments, we implemented this option, i.e., we approach (8) by solving once for M and then once for S , in each iteration of the loop between the lines 3 and 8 of Algorithm 2.

After solving for $N^{[l+1]}$ and $(S^{[l+1]}, M^{[l+1]})$, problem (9) updates the missing data. The solution of (9) is trivial: we just have to take $Z_{ij}^{[l+1]}$ as the (i, j) th entry of $S^{[l+1]}M^{[l+1]}$ for all $(i, j) \notin \mathcal{O}$.

Algorithm convergence. At best, the BALM algorithm can produce a local minimizer for (1). That is, we do not claim that BALM (algorithm 1) converges to a global minimizer. In fact, even the nonlinear Gauss-Seidel technique (algorithm 2) is not guaranteed to globally solve (6). This is the common situation when dealing with non-convex problems. A detailed theoretical study of the convergence properties of BALM is available in [38]. See [39] for some convergence results on augmented Lagrangian methods. We now apply our generic BALM algorithm to solve different bilinear computer vision problems: the rigid, non-rigid, articulated structure-from-motion problems, Photometric Stereo (PS) problem and 2D-3D image registration problem.

4.1 Example 1: Rigid and Non-Rigid SfM

The problem of recovering the non-rigid 3D shape of a deforming object from a monocular video sequence given only 2D correspondences between frames was formulated as a factorization problem by Bregler *et al.* [5] in the case of an orthographic camera. The assumption is that the 3D shape can be represented as a linear combination of a set of d basis shapes with time varying coefficients t_{id} . If the image coordinates are referred to

the centroid, the projection of the shape at frame i can be expressed as

$$Y_i = \begin{bmatrix} u_{i1} & \dots & u_{in} \\ v_{i1} & \dots & v_{in} \end{bmatrix}^\top = \left(\sum_{l=1}^d t_{il} B_l \right) Q_i = \quad (12)$$

$$= [B_1 \quad \dots \quad B_d] (t_i \otimes Q_i) = SM_i$$

where Y_i is the $n \times 2$ measurement matrix that contains the 2D coordinates of n image points in frame i , B_l are the basis shapes of size $n \times 3$ and t_{il} are the time varying shape coefficients and Q_i is the projection matrix for frame i , which in the case of an orthographic camera is a 3×2 matrix that encodes the first two columns of a rotation matrix (therefore it is a Stiefel matrix). Note that we are defining $M_i := t_i \otimes Q_i$, and \otimes denotes the Kronecker product. Rigid SfM can be instantiated with this framework by imposing $d = 1$, a single basis shape. By concatenating all the measurements for all the frames into a single matrix we have

$$Y = [B_1 \quad \dots \quad B_d] [t_1 \otimes Q_1 \quad \dots \quad t_f \otimes Q_f] = \quad (13)$$

$$= S [M_1 \quad \dots \quad M_f] = SM.$$

Now, we have expressed the measurement matrix as a bilinear interaction between the shape matrix S of size $n \times 3d$ and the motion matrix M of size $3d \times 2f$. This form fits exactly the optimization problem as presented in Eq. (1). Therefore, in the NRSfM case, the manifold constraint corresponds to

$$\mathcal{M} = \{ t \otimes Q : t \in \mathbb{R}^d, Q \in \mathbb{R}^{3 \times 2}, Q^\top Q = I_2 \}, \quad (14)$$

or in other words, the two rows of the rotation matrix Q^\top must be orthonormal (i.e. it is a Stiefel matrix). To apply our BALM algorithm, we need first to derive the projector onto \mathcal{M} . We now turn to this problem.

4.1.1 NRSfM manifold projector

In [30] Paladini *et al.* derived an exact globally optimal algorithm to project the motion matrices onto the non-rigid motion manifold (14). However, here, we discuss an alternative which provides an *approximate* projector onto \mathcal{M} . The advantage is that our proposed algorithm stills provides accurate estimates while being considerably faster (approximately 100 times in experimental tests).

Let $A \in \mathbb{R}^{3d \times 2}$ be given, and consider the partition $A = [A_1^\top A_2^\top \dots A_d^\top]^\top$, where $A_i \in \mathbb{R}^{3 \times 2}$. We want to compute $p_{\mathcal{M}}(A)$, that is, we want to solve the optimization problem

$$\begin{aligned} & \text{minimize} && \|A - t \otimes Q\|^2. \\ & \text{subject to} && t \in \mathbb{R}^d \\ & && Q^\top Q = I_2 \end{aligned} \quad (15)$$

If $A \in \mathcal{M}$, that is, if $A = t \otimes Q$ for some $t = (t_1, \dots, t_d)^\top \in \mathbb{R}^d$ and Stiefel matrix Q , then we would have the identity $\sum_{i=1}^d A_i A_i^\top = \|t\|^2 Q Q^\top$. That is, the left-hand side of the equation would reveal the underlying Q (up to a right multiplication by a 2×2 rotation). This motivates

the following approach, for a generic A : compute $\mathcal{A} = \sum_{i=1}^d A_i A_i^\top$ and estimate Q as its dominant Stiefel, that is, corresponding to the 2 top eigenvectors of \mathcal{A} . Let \widehat{Q} denotes this Stiefel matrix. Even when $A \in \mathcal{M}$ we do not have, in general, $\widehat{Q} = Q$. Rather, $\widehat{Q} = QR$ for a 2×2 orthogonal matrix R . Thus, we return to problem (15) and we solve for $t \in \mathbb{R}^d$ and R :

$$\begin{aligned} & \text{minimize} \quad \left\| A - t \otimes (\widehat{Q}R) \right\|^2 \\ & \text{subject to} \quad t \in \mathbb{R}^d \\ & \quad \quad \quad R \in \mathbb{R}^{2 \times 2}, R^\top R = I_2 \end{aligned} \quad (16)$$

For a fixed R , the optimal t is $t_i = \frac{1}{2} \text{tr} \left(R^\top \widehat{Q}^\top A_i \right)$ with $i = 1, \dots, d$. Plugging t_i into (16) gives the reduced problem over R

$$\begin{aligned} & \text{maximize} \quad \sum_{i=1}^d \left(\text{tr} \left(R^\top T_i \right) \right)^2 \\ & \text{subject to} \quad R^\top R = I_2 \end{aligned} \quad (17)$$

where $T_i := \widehat{Q}^\top A_i$. Now, a 2×2 rotation matrix R must fall into one of the two following cases $\det(R) = 1$ or $\det(R) = -1$.

Case $\det(R) = 1$. In this case

$$R = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}$$

for some $(c, s) \in \mathbb{R}^2$, $\|(c, s)\| = 1$. Using this representation in (17) yields

$$\begin{aligned} & \text{maximize} \quad \sum_{i=1}^d [c \quad s] \sum_{i=1}^d \begin{bmatrix} T_i(1,1) + T_i(2,2) \\ T_i(2,1) - T_i(1,2) \end{bmatrix}^\top \begin{bmatrix} c \\ s \end{bmatrix} \\ & \text{subject to} \quad \|(c, s)\| = 1 \end{aligned}$$

which can be solved by an eigenvalue decomposition.

Case $\det(R) = -1$. In this case

$$R = \begin{bmatrix} c & s \\ s & -c \end{bmatrix}$$

for some $(c, s) \in \mathbb{R}^2$, $\|(c, s)\| = 1$. Using this representation in (17) yields

$$\begin{aligned} & \text{maximize} \quad \sum_{i=1}^d [c \quad s] \sum_{i=1}^d \begin{bmatrix} T_i(1,1) + T_i(2,2) \\ T_i(2,1) - T_i(1,2) \end{bmatrix}^\top \begin{bmatrix} c \\ s \end{bmatrix} \\ & \text{subject to} \quad \|(c, s)\| = 1 \end{aligned}$$

which can be solved by an eigenvalue decomposition. After examining separately the two cases, we pick the best.

4.2 Example 2: Articulated SfM

A bilinear formulation of motion and 3D structure components is also possible for articulated image shapes. In such case two, or more, moving object have their motion constrained by a certain extent given the type of joint among the shapes. These constraints lead to a factorization formulation of the problem and this results in a

drop in the dimensionality of the measurement matrix. For the simplified case of two shapes, we have that the overall measurement matrix is given by $Y = \begin{bmatrix} Y^{(1)} \\ Y^{(2)} \end{bmatrix}$ and it contains the 2D image point trajectories for each object with $Y^{(1)}$ and $Y^{(2)}$ being a $n_1 \times m$ and a $n_2 \times m$ matrix respectively (i.e. $n_1 + n_2 = n$). In the case of a *universal joint* the two shapes share a common translation (i.e. the distance between the centers of mass of the shapes is constant) while in the case of a *hinge joint* the shapes also share a common rotation axis [8], [40]. As an initial assumption, the approach requires that an initial segmentation stage has taken place to assign the trajectories in Y to the respective shapes.

A *universal joint* [8] constrains the distance between the centres of the two shapes to be constant (for instance, the head and the torso of a human body) but with independent rotation components. At each frame i the shapes connected by a joint satisfy:

$$t^{(1)} + d^{(1)\top} Q^{(1)} = t^{(2)} + d^{(2)\top} Q^{(2)} \quad (18)$$

where $t^{(1)}$ and $t^{(2)}$ are the 2D image centroid of the two objects, $Q_i^{(1)}$ and $Q_i^{(2)}$ are the 3×2 orthographic camera matrices and $d^{(1)}$ and $d^{(2)}$ the 3D displacement vectors of each shape from the joint. The relation in equation (18) gives the reduced dimensionality in the motion and shape subspaces. Thus, the shape matrix S can be written as:

$$S = \begin{bmatrix} S^{(1)} & 0 & 1_{n_1} \\ D^{(1)} & S^{(2)} - D^{(2)} & 1_{n_2} \end{bmatrix} \quad (19)$$

where S of size $n \times 7$ is a full rank matrix. The matrices $D^{(1)}$ and $D^{(2)}$ are given by $D^{(1)} = d^{(1)} 1_{n_2}^\top$ and $D^{(2)} = d^{(2)} 1_{n_1}^\top$ where 1_{n_1} and 1_{n_2} are vectors of n_1 and n_2 ones respectively. The motion for a frame i has to be accordingly arranged to satisfy equation (18) as:

$$M_i = \begin{bmatrix} Q_i^{(1)} \\ Q_i^{(2)} \\ t_i^{(1)\top} \end{bmatrix}. \quad (20)$$

In the case of a *hinge joint*, if we assume the image coordinates to be registered to the centroid of each segment, then the motion matrices M_i that lie on the articulated *motion manifold* can be written as:

$$M_i = \begin{bmatrix} u_i^\top \\ A_i \\ B_i \end{bmatrix} \quad (21)$$

where u is a 2-vector with the common rotation axis for both objects, A_i and B_i are 2×2 matrices such that $Q_i^{(1)} = [u_i \quad A_i]^\top$ and $Q_i^{(2)} = [u_i \quad B_i]^\top$ are the 3×2 orthographic camera matrices associated with the first and second shape respectively. The metric constraints in the case of a hinge can therefore be expressed as:

$$[u_i \quad A_i] \begin{bmatrix} u_i^\top \\ A_i^\top \end{bmatrix} = I_2 \quad \text{and} \quad [u_i \quad B_i] \begin{bmatrix} u_i^\top \\ B_i^\top \end{bmatrix} = I_2 \quad (22)$$

where, without loss of generality, we have implicitly assumed that the axis of rotation is aligned with the x -axis of the first object. According to the hinge model [8], we can write S as:

$$S = \begin{bmatrix} x_1^{(1)} & y_1^{(1)} & z_1^{(1)} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n_1}^{(1)} & y_{n_1}^{(1)} & z_{n_1}^{(1)} & 0 & 0 \\ x_1^{(2)} & 0 & 0 & y_1^{(2)} & z_1^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n_2}^{(2)} & 0 & 0 & y_{n_2}^{(2)} & z_{n_2}^{(2)} \end{bmatrix} \quad (23)$$

where now S is a $5 \times p$ matrix and $n = n_1 + n_2$ (we assume the shapes have been registered to the respective object centroids). Therefore, in the case of a hinge joint the rank of the measurement matrix is at most 5.

Regarding the projector for the articulated case, we point the reader to the solution presented in [30] that provides an optimal projection on the articulated *motion manifold*.

4.3 Example 3: Photometric Stereo

Basri et al. [1] derived a bilinear approximation of the image brightness given by luminance variations. This derivation is based on a spherical harmonics representation of lighting variations and it allows to frame PS as a factorization problem with manifold constraints on one of the bilinear factors. Given a set of images of a Lambertian object with varying illumination, it is possible to extract the dense normal to the surface of the object z , the albedo ρ and the lighting directions l . For a 1st order spherical harmonics approximation, the brightness at image pixel j at frame i can be modelled as:

$$Y_{ij} = l_i^\top \rho_j [1 \ z_j^\top]^\top = S_i M_j$$

where $l_i \in \mathbb{R}^4$, $\rho_j \in \mathbb{R}$, $z_j \in \mathbb{R}^3$ with $z_j^\top z_j = 1$. A compact matrix form can be obtained for each pixel Y_{ij} as:

$$Y = \begin{bmatrix} Y_{11} & \dots & Y_{1n} \\ \vdots & \ddots & \vdots \\ Y_{f1} & \dots & Y_{fn} \end{bmatrix} = \begin{bmatrix} l_1^\top \\ \vdots \\ l_f^\top \end{bmatrix} \left[\rho_1 \begin{bmatrix} 1 \\ z_1 \end{bmatrix} \quad \dots \quad \rho_n \begin{bmatrix} 1 \\ z_n \end{bmatrix} \right] = SM \quad (24)$$

where a single image i is represented by the vector $Y_i = [y_{i1} \ \dots \ y_{in}]$. Manifold constraints are given by the surface normal constraints which implies $M^{4 \times f}$ lying on the manifold defined by:

$$\mathcal{M} = \{ \rho [1 \ z^\top]^\top : \rho \in \mathbb{R}, z \in \mathbb{R}^3, z^\top z = 1 \}, \quad (25)$$

The matrix $S^{n \times 4}$ now contains the collection of lighting directions which which combines the first-order spherical harmonics at each frame i .

4.3.1 Photometric Stereo manifold projector

We now derive the projector onto the manifold \mathcal{M} defined in (25). That is, for a given $a \in \mathbb{R}^4$, we show how to solve the associated optimization problem

$$\begin{aligned} & \text{minimize} \quad \|a - \rho [1 \ z^\top]^\top\|^2 \\ & \text{subject to} \quad z^\top z = 1 \end{aligned} \quad (26)$$

The variable to optimize is $(\rho, z) \in \mathbb{R} \times \mathbb{R}^3$. Consider the partition $a = [\alpha \beta^\top]^\top$ with $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^3$. We can rewrite (26) as

$$\begin{aligned} & \text{minimize} \quad \rho^2 - \alpha\rho - \rho\beta^\top z \\ & \text{subject to} \quad z^\top z = 1 \end{aligned} \quad (27)$$

We denote by (ρ^*, z^*) the solution of (27). If $\beta = 0$ then $\rho^* = \alpha/2$ and z^* can be any unit-norm vector. If $\beta \neq 0$ then optimizing (27) over z (for a fixed ρ) gives

$$z^* = \frac{\rho}{|\rho|} \frac{\beta}{\|\beta\|} \quad (28)$$

and leaves the optimization problem

$$\begin{aligned} & \text{minimize} \quad \rho^2 - \alpha\rho - |\rho| \|\beta\| \\ & \rho \end{aligned} \quad (29)$$

Now, if $\alpha \geq 0$ (respectively $\alpha < 0$) then it is clear that $\rho^* \geq 0$ ($\rho^* < 0$). Inserting this constraint into (29) leaves a quadratic problem with an easy solution: $\rho^* = (\alpha + \|\beta\|)/2$ (respectively $\rho^* = (\alpha - \|\beta\|)/2$).

4.4 Example 4: Image Registration

A common problem in Computer Vision is the registration of a 3D model defined as a set of basis shapes to a set of 2D coordinates. This 2D-3D registration problem was one of the application domains of the Branch and Bound method proposed by Chandraker and Kriegman [10]. Our BALM approach can be reformulated to deal with this case as well with a simple modification to the optimization and by introducing a new manifold projector. In order to be able to relate the solution proposed in [10] with ours we will also assume that the image coordinates (i.e. x or y in Y) will be optimized independently. We want to solve:

$$\begin{aligned} & \text{minimize} \quad \sum_{j=1}^N (y_j - a^\top \sum_{i=1}^p \alpha^i X_j^i)^2 \\ & \text{subject to} \quad \|a\| = 1 \end{aligned} \quad (30)$$

The variable to optimize is $(a, \alpha) \in \mathbb{R}^q \times \mathbb{R}^p$. The vectors $y = (y_1, \dots, y_N) \in \mathbb{R}^N$, $\{X_j^i \in \mathbb{R}^q : j = 1, \dots, N, i = 1, \dots, p\}$ are given.

Problem (30) can be expressed in our canonical form (1). Indeed, problem (30) is equivalent to

$$\begin{aligned} & \text{minimize} \quad \|y - Sm\|^2 \\ & \text{subject to} \quad m \in \mathcal{M} \end{aligned} \quad (31)$$

where

$$S := \begin{bmatrix} (X_1^1)^\top & (X_1^2)^\top & \dots & (X_1^p)^\top \\ (X_2^1)^\top & (X_2^2)^\top & \dots & (X_2^p)^\top \\ \vdots & \vdots & \dots & \vdots \\ (X_N^1)^\top & (X_N^2)^\top & \dots & (X_N^p)^\top \end{bmatrix} \in \mathbb{R}^{N \times pq}$$

and

$$\mathcal{M} = \{m : m = \alpha \otimes a, \text{ for some } \alpha \in \mathbb{R}^p, a \in \mathbb{R}^q, \|a\| = 1\}. \quad (32)$$

In the following, we assume that $N > pq$ (thus, S is a tall matrix).

Note that the only variable to optimize in (31) is $m \in \mathbb{R}^{pq}$. We recall that S is given (thus, this problem is simpler than (1) where S is also a variable; also, in (31) both the matrices Y and M correspond to vectors y and m , respectively).

As usual, we reformulate (31) as

$$\begin{aligned} & \text{minimize} && \|y - Sm\|^2. && (33) \\ & \text{subject to} && m = n \\ & && n \in \mathcal{M} \end{aligned}$$

In Algorithm 3 we present the BALM algorithm for (33):

Algorithm 3 BALM for registration problems

- 1: set $k = 0$ and $\epsilon_{\text{best}} = +\infty$
 - 2: initialize $\sigma^{(0)}, r^{(0)}, \gamma > 1$ and $\eta < 1$
 - 3: initialize $m^{(0)}$
 - 4: **repeat**
 - 5: solve

$$\left(m^{(k+1)}, n^{(k+1)}\right) = \underset{n \in \mathcal{M}}{\operatorname{argmin}} L_{\sigma^{(k)}}(m, n; r^{(k)}) \quad (34)$$
 - 6: compute $\epsilon = \|m^{(k+1)} - n^{(k+1)}\|^2$
 - 7: **if** $\epsilon < \eta \epsilon_{\text{best}}$
 - 8: $r^{(k+1)} = r^{(k)} - \sigma^{(k)} (m^{(k+1)} - n^{(k+1)})$
 - 9: $\sigma^{(k+1)} = \sigma^{(k)}$
 - 10: $\epsilon_{\text{best}} = \epsilon$
 - 10: **else**
 - 10: $r^{(k+1)} = r^{(k)}$
 - 11: $\sigma^{(k+1)} = \gamma \sigma^{(k)}$
 - 12: **endif**
 - 13: update $k \leftarrow k + 1$
 - 14: **until** some stopping criterion
-

Note that

$$L_{\sigma}(m, n; r) = \|y - Sm\|^2 - r^{\top} (m - n) + \frac{\sigma}{2} \|m - n\|^2.$$

To address (34) we can use a Gauss-Seidel approach as before. A summary of the optimization is given in Algorithm 4.

We now address the subproblems (35) and (36). Subproblem (36) is just an unconstrained quadratic minimization with respect to m . The solution is

$$m^{[l+1]} = \left(S^{\top} S + \frac{\sigma^{(k)}}{2} I_{pq}\right)^{-1} \left(S^{\top} y + \frac{1}{2} r^{(k)} + \frac{\sigma^{(k)}}{2} n^{[l+1]}\right).$$

Problem (35) corresponds, in the current setting, to problem (7). Following similar manipulations, we conclude

Algorithm 4 Iterative Gauss Seidel scheme to solve for (34)

- 1: set $l = 0$ and choose L_{max}
 - 2: set $m^{[0]} = m^{(k)}$
 - 3: **repeat**
 - 4: solve

$$n^{[l+1]} = \underset{n \in \mathcal{M}}{\operatorname{argmin}} L_{\sigma^{(k)}}(m^{[l]}, n; r^{(k)}) \quad (35)$$
 - 5: solve

$$m^{[l+1]} = \underset{m \in \mathcal{M}}{\operatorname{argmin}} L_{\sigma^{(k)}}(m, n^{[l+1]}; r^{(k)}) \quad (36)$$
 - 6: update $l \leftarrow l + 1$
 - 7: **until** $l = L_{\text{max}}$
 - 8: **set** $m^{(k+1)} = m^{[L_{\text{max}}]}$ and $n^{(k+1)} = n^{[L_{\text{max}}]}$.
-

that the solution of (35) is given by

$$n^{[l+1]} = p_{\mathcal{M}} \left(m^{[l]} - \frac{1}{\sigma^{(k)}} r^{(k)} \right),$$

which is the counterpart of (11). We discuss the projector manifold $p_{\mathcal{M}}$ in the next subsection. Finally, we propose the following initialization $m^{(0)}$ for (31) (hence, for (33)):

$$m^{(0)} = p_{\mathcal{M}}(S^+ y)$$

where S^+ denotes the pseudo-inverse of S .

4.4.1 Image registration manifold projector

We now consider how to compute $p_{\mathcal{M}}(x)$ for a generic $x \in \mathbb{R}^{pq}$ where \mathcal{M} is given as in (32). That is, we address the optimization problem

$$\begin{aligned} & \text{minimize} && \|x - \alpha \otimes a\|^2 && (37) \\ & \text{subject to} && \|a\| = 1 \end{aligned}$$

where the variable to optimize is $(\alpha, a) \in \mathbb{R}^p \times \mathbb{R}^q$. Consider the partition of the given $x \in \mathbb{R}^{pq}$ as

$$x = [x_1 \quad x_2 \quad \cdots \quad x_p]^{\top}$$

and define the matrix

$$X = [x_1 \quad x_2 \quad \cdots \quad x_p] \in \mathbb{R}^{q \times p}.$$

With this notation, we can rewrite (37) as

$$\begin{aligned} & \text{minimize} && \|X - a\alpha^{\top}\|^2. && (38) \\ & \text{subject to} && \|a\| = 1 \end{aligned}$$

It is now clear that (38) asks for the optimal rank 1 approximation of the matrix X . In conclusion, problem (37) can be solved as follows:

- 1) Write the given $x \in \mathbb{R}^{pq}$ as

$$x = [x_1 \quad x_2 \quad \cdots \quad x_p]^{\top}$$

where each $x_i \in \mathbb{R}^q$ for $i = 1, 2, \dots, p$

- 2) Build the matrix

$$X := [x_1 \quad x_2 \quad \cdots \quad x_p] \in \mathbb{R}^{q \times p}.$$

In Matlab notation, this corresponds to $x = \text{reshape}(x, q, p)$.

3) Perform the SVD

$$X = U\Sigma V^T$$

4) Set $p_{\mathcal{M}}(x) = (\sigma_1 v_1) \otimes u_1$, where σ_1 is the top singular value of X (that is, the (1,1) entry of Σ), $u_1 \in \mathbb{R}^q$ is the first column of U and $v_1 \in \mathbb{R}^p$ is the first column of V .

5 EXPERIMENTS

To evaluate our unified algorithm we carry out experiments on the example problems proposed above with both synthetic and real data². The aim of our tests is twofold: to show that the performance of BALM is comparable to the best specialized algorithms and to assess its convergence. In the NRSfM and PS problems we will also assess the resilience of our approach to very high levels of missing data. We will use the 2D-3D non-rigid registration problem to validate the convergence of our algorithm comparing against a Branch & Bound based algorithm [10].

5.1 Synthetic experiments: NRSfM

First we evaluate the performance of our bilinear algorithm when applied to the NRSfM problem. We consider two different sets of synthetic experiments. The first set of tests is designed to verify the resilience of the algorithm to increasing ratios of missing data. We used a 3D motion capture sequence of a face. The sequence was captured using a VICON system tracking a subject wearing 37 markers on the face to provide 3D ground truth for the evaluation. The 3D points were then projected synthetically onto an image sequence 74 frames long using an orthographic camera model. To test the performance of our algorithm we computed the 3D reconstruction error, defined as the Frobenius norm of the difference between the recovered 3D shape S and the ground truth 3D shape S_{GT} . The relative 3D error is then computed as: $\|S - S_{GT}\| / \|S_{GT}\|$. We subtract the centroid of each shape and align them with Procrustes analysis. We evaluated the performance of the algorithm with respect to noise in the image measurements of up to 3% and up to 80% missing data in a combined test. Zero mean additive Gaussian noise was applied with standard deviation $\sigma = n \times s / 100$ where n is the noise percentage and s is defined as $\max(Y)$ in pixels. In all experiments the number of basis shapes was fixed to $k = 5$. The results for each level of noise were averaged over 100 trials.

In Figure 1 we compare the results of our proposed algorithm (BALM) with Torresani *et al.*'s algorithm [41] (EM-PPCA), Bundle Adjustment (BA) [42], the MP method [30] and the trilinear approach of Wang *et al.* [43]

². The code for the BALM method and the manifold projectors is available from: <http://www.isr.ist.utl.pt/~adb/the-balm/>.

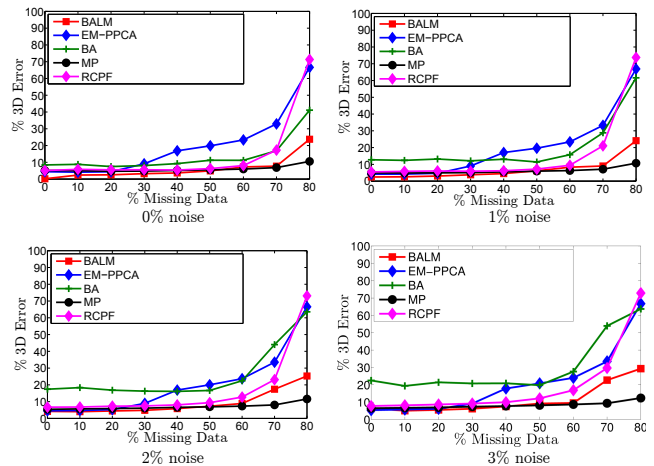


Fig. 1. Synthetic experiment results showing comparison with several NRSfM methods with different ratios of missing data and noise.

(RCPF) for different levels of noise. In practical terms, note that a reconstruction error above 20% is too high to be of any use in most applications. Regarding the overall results, while in the case of full data the performance of all algorithms is comparable, BALM and MP outperform the rest of the algorithms in the case of missing data. Notice that RCPF is closer to both BALM and MP since it also includes a projection step of the rotation matrices onto the correct manifold. However, since the projection step is only approximate, this algorithm breaks down for lower levels of missing data. On the other hand, BA and EMPPCA deteriorate for levels of missing data above 30%. Also notice that the algorithms that perform metric projections (BALM, MP and RCPF) are less affected by increasing levels of noise than others. BALM closely follows the performance of the best performing algorithm (MP) which is specific for NRSfM. A noticeable decrease in performance for BALM occurs at 80% missing data (70% for the higher percentages of noise). Regarding runtime, a single manifold projection takes approximately 1.8 msec for each frame with $d = 5$ basis shapes.

Regarding the initialization of the ALM algorithm in the case of NRSfM, the missing data tracks are first filled in using [29] which enforces metric constraints on the motion matrices. The camera matrices are initialized assuming rigid motion. Torresani *et al.*'s initialization [41] is used to estimate configuration weights and basis shapes given the residual of the first rigid solution.

5.2 Synthetic experiments: image registration

To evaluate the convergence of our BALM algorithm we compare with the global minimizer proposed by Chandraker and Kriegman [10], based on Branch & Bound. Since the use of this algorithm is restricted to problems with a small number of variables, we choose to test the convergence of our algorithm on the same small scale problem of 2D-3D non-rigid registration. To

ease the comparison Figure 2 shows the results using an identical experimental setup to the one used for the validation of their method [10]. We show the L2 norm in order to maintain consistency with the cost functions optimized by both approaches.

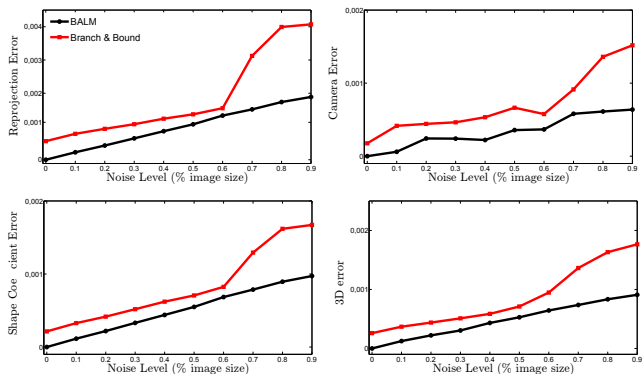


Fig. 2. Comparison of the errors obtained with our BALM algorithm (black line) and the Branch & Bound approach of [10] (red line) applied to the 2D-3D non-rigid registration problem. We show errors in mean 2D image reprojection, camera pose, shape coefficients and 3D shape. A set of 1000 experiments was run for each level of image noise and the errors averaged. The BALM plots, shown in black, converge to the global minimum in the noiseless case. Notice that these plots replicate the convergence behavior shown in [10].

Interestingly, BALM compares favorably with respect to [10]. The reason why our algorithm consistently achieves a lower error than Branch & Bound is related to the MATLAB implementation of [10]³ since the bound could not be reduced to less than 10^{-5} . This lack of further branching is a common and known issue of this type of optimization algorithms. Clearly, BALM can reach a closer solution because of its continuous optimization nature (in contrast with the *discrete* nature of Branch & Bound approaches).

Notice that the solution obtained with the Branch & Bound converges to the global optimum (as we explained above, the small error is due to the branching stopping criteria used in the tested implementation). BALM also converged to the global optimum without the use of any bounds on the non-linear parameters of the problem (i.e. the Stiefel matrix) thus obtaining a closer solution. Regarding execution times, Branch & Bound approaches are known to be computationally expensive while our ALM based approach is four orders of magnitude faster as Table 2 confirms.

The initialization for the registration problem is made by first centering the 2D image points to the image centroid t . Then we obtain an estimate of the motion matrix not satisfying the manifold constraints by simply

3. Note that we have used the same code and sequences provided directly by the authors of [10]

TABLE 2
Mean computational time for BALM and B&B.

Noise	0.1	0.3	0.5	0.7	0.9
B&B	21s	28s	31s	32s	20s
BALM	0.0022s	0.0028s	0.0032s	0.0155s	0.0245s

pseudo inverting the known S such that $S^+Y = M$. This initial M is then fed to the BALM algorithm as presented in Algorithm 3.

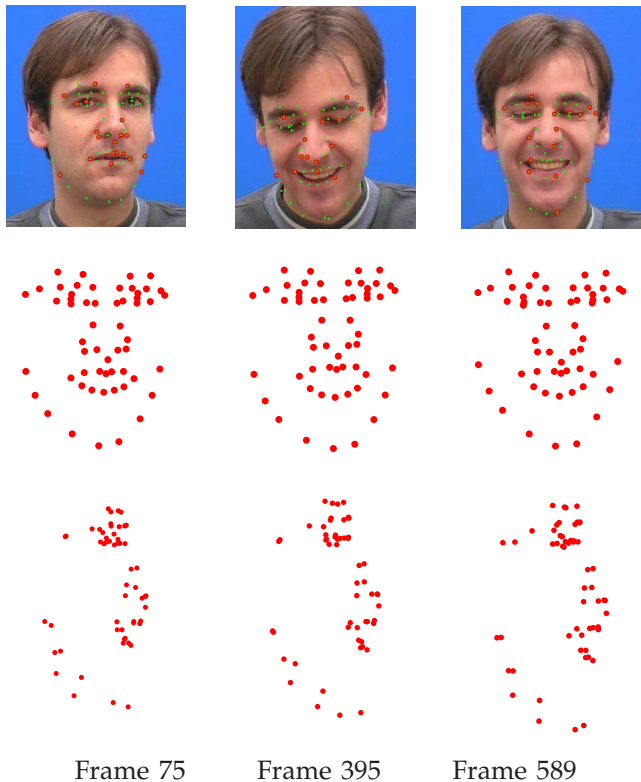


Fig. 3. The *Franck* sequence (first row) with 30% missing data. Tracked points are in green while red circles show missing entries. Second row: frontal view of 3D reconstruction. Third row: side view of the 3D shape.

5.3 Real data: NRSfM

We have tested BALM for NRSfM in the face modelling domain on the Franck sequence⁴. The face points were tracked with Active Appearance Models giving 56 points in 700 frames and ratio of 30% missing data was simulated synthetically. The first row of Figure 3 shows a sample of the sequence and the bottom row shows the corresponding reconstructions. The resulting 3D shape and deformations describe the shape well, even in occluded areas (e.g. lips).

4. The image sequence is freely available at: www-prima.inrialpes.fr/FGnet/data/01-TalkingFace/talking_face.html

5.4 Real data: Rigid SfM

We tested our BALM algorithm in the case of rigid scenes. Figure 4 shows results for the classic dinosaur sequence with 72% missing data. The results confirm that BALM performs closely to the best performing methods for rigid structure. The overall 2D rms error obtained with BALM⁵ was 1.3039 which is a slight improvement over the error reported by Marques and Costeira (1.3705) but higher than the error of the Damped-Newton approach by Buchanan and Fitzgibbon (1.0847). However notice that their best result [21] does not enforce exact orthonormality constraints and only obtains an affine 3D reconstruction while ours guarantees a metric one. We then initialized a non-linear bundle adjustment optimization step [44] obtaining a marginal improvement in the 2D error which went down to 1.2302. This points to the fact that BALM provided a solution that was already close to the local minimum.

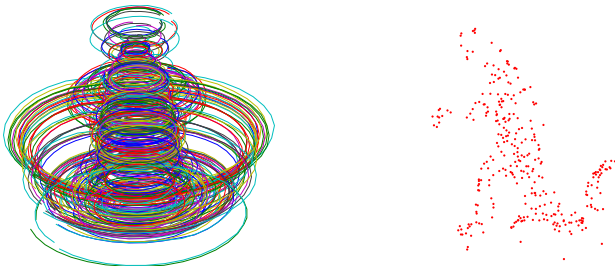


Fig. 4. The BALM algorithm applied to the *Dinosaur* sequence. Left: complete 2D image trajectories resulting from our algorithm. Right: reconstructed 3D shape.

We have also attempted the reconstruction of the *casa da musica* sequence [29] with 60% missing data where the images were obtained from *Google images* and thus generally shot far apart and with unknown cameras (no temporal consistency of camera views). The sequence also contains degenerate configurations since only planar surfaces are seen from each camera view. The 3D reconstruction in Figure 5 shows that most of the planar surfaces are correctly reconstructed and they provide a credible 3D reconstruction.

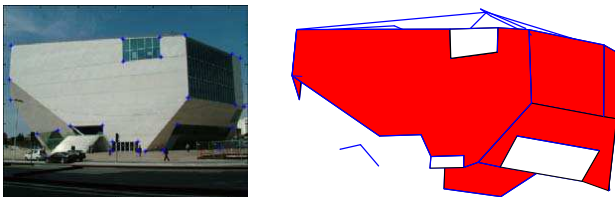


Fig. 5. BALM applied to the *casa da musica* dataset. Left: sample image of the dataset. Note that the images have a large baseline. Right: reconstructed 3D shape.

Finally, in order to test the performance of BALM in large scale reconstruction problems we have run

⁵. Notice that this error is calculated only from the known entries to compare correctly with [29] and [21]

experiments on a modification of the Venice dataset released by Agarwal et al. [45]. The original dataset could not be used in our experiments since the perspective effects were found to be too strong for our orthographic camera model assumptions. Our modified dataset was constructed by projecting the large scale 3D model of *Piazza San Marco* in Venice with almost 10^6 points using 100 random orthographic camera matrices. This allows us to perform an evaluation with ground truth data on a challenging dataset. Overall, the final measurement matrix Y was of size 939551×200 leading to a factorisation problem with nearly 200 million entries. We generated a random missing data pattern with 90% missing entries. Our BALM algorithm gave a final 3D reconstruction error of 6.67% on this dataset. Notice in Figure 6 that most of the error is localised where the 3D model points are sparser (i.e. top of the bell tower and surrounding areas of the square).

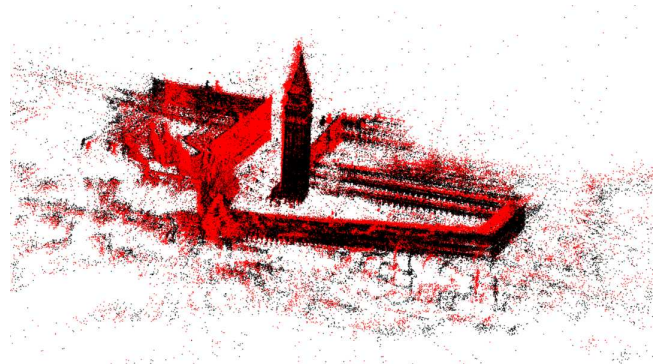


Fig. 6. Results of BALM applied to reconstruct a modification of the *Venice* dataset [45] with 90% missing data. Ground truth 3D points are in black while 3D reconstructed points are in red

5.5 Real data: Articulated SfM

We present the reconstruction of the *hinge2* sequence [8] with two boxes linked by an hinge joint as shown in Figure 7. The bigger box has 72 points and the smaller one on top has 25 points. The missing data amounted to 60% of the available data in Y of size 1630×97 . After an initialization obtained by filling the missing entries for the two shapes independently with rigid SfM, we apply our BALM algorithm using the projector described in [30]. The results show that even in this case of high levels of missing the data, the position of the axis is estimated correctly and it reflects the real motion of the objects.

Additionally, we show results using a motion capture sequence of a person kicking a football. A motion capture system tracked 333 markers on the body. We selected 15 and 18 points respectively on the upper and lower leg to reconstruct the knee joint. The 3D points are projected with a synthetic camera using orthographic projection. From the 2D tracks we recover the rotation axis and the 3D structure of the leg, as shown in Figure 8. The reconstructed 3D points and correctly recovered axis

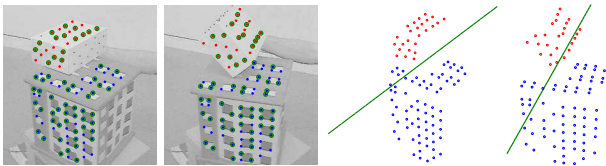


Fig. 7. BALM applied to the *hinge2* sequence. First 2 images: two sample images and tracked points. Dark green circles represent missing data for that frame (60% for the whole sequence). The remaining images show the respective 3D reconstructions and estimated 3D hinge joint (green axis).

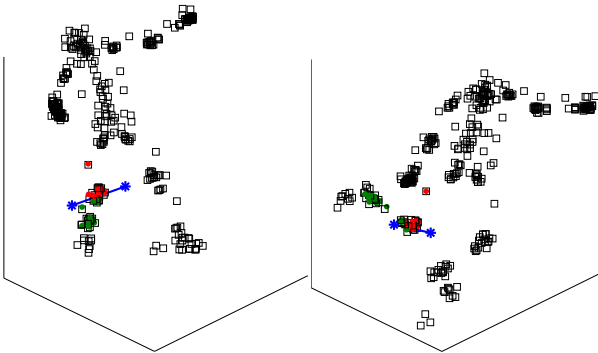


Fig. 8. Motion capture *football* sequence. Reconstruction obtained with BALM for frames 15 and 35. The 3D error in the reconstruction achieved with BALM was 4.97% against motion capture ground truth data. The recovered axis is the knee joint.

have been aligned to the MOCAP data to show the full body pose. The overall 3D error was 4.97% over the 51 frame long sequence.

5.6 Real data: Photometric Stereo

We present results for our BALM factorization algorithm applied to the photometric stereo problem using the projector presented in Section 4.3.1. The aim is to extract the 3D surface, albedo and luminance parameters in the presence of significant occlusions in the input image data. We consider missing data to be the dark pixels and the saturated parts of the images in the sequence. This is a reasonable assumption since in these areas the Lambertian model considered will not be satisfied. Similarly to the NRSfM case, we need to provide an initialization for the missing entries in Y . A heuristic for initialization which gives good results is to use an inpainting technique [46] which fills the image holes given the known part of the image. The initialization for $(S^{(0)}, M^{(0)})$ is then given by a simple rank-4 SVD on the “inpainting” Y . These affine low-rank components are then normalized as in [1] to comply with the equal norm constraints of the spherical harmonics model.

The first row of Figure 9 (Sculpture sequence) shows that even with large occluded areas the initialization

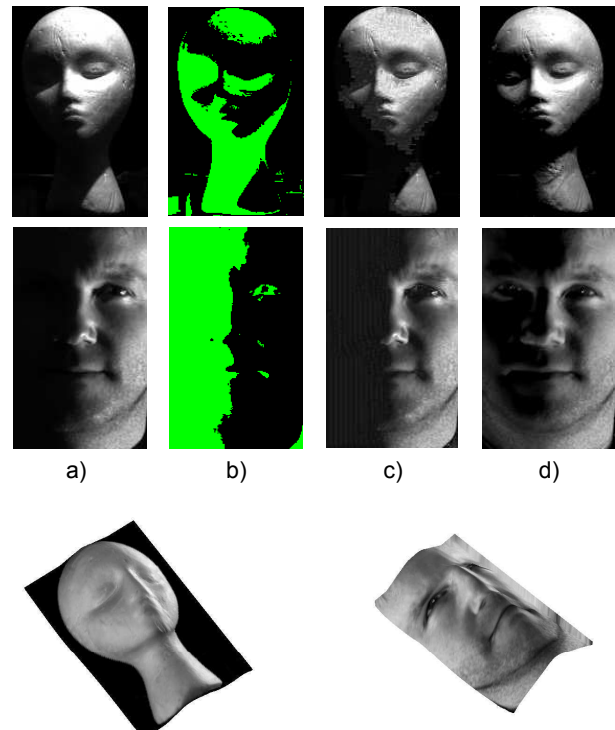


Fig. 9. Photometric stereo results for the 46-frame long Sculpture sequence with 48% missing data (1st row) and 49-frame long YaleB10 sequence with 31% missing data (2nd row). Images show: a) selected frame from the original sequence; b) image mask where green pixels represent missing entries; c) image used as initialization for BALM; d) resulting BALM optimized image. Last row: computed surface after needle map integration.

with inpainting copes well if the overall texture of the object is fairly homogenous (although errors can still be noticed on the top of the head and the darkest areas). Note that the final optimized image (Figure 9(d)) conforming to a strict Lambertian model reveals some further details in the neck area which were hidden in the original frame. The sequence on the second row, taken from a YaleB database sequence, shows an extreme occlusion in which half of the face of the subject is not visible. In this case inpainting clearly fails to provide a good initialization but the reconstructed shape still resembles the subject, estimating the closest image that satisfies the Lambertian model assumptions.

In a further test, we applied photometric stereo to a large-scale experiment *Potatoes* sequence of the *robot data set* [47]⁶ comprising 19 high resolution images of size 1600×1200 (see Figure 10). Since the objects of interest occupied a restricted area of the image those regions were pre-selected. The data set provides 3D ground truth (given by a structured light acquisition system) in addition to the calibration parameters required to align the 3D shape at each image. The sequence offers several challenges: the setup uses led lights for illumination and

6. <http://roboimagedata.imm.dtu.dk/>

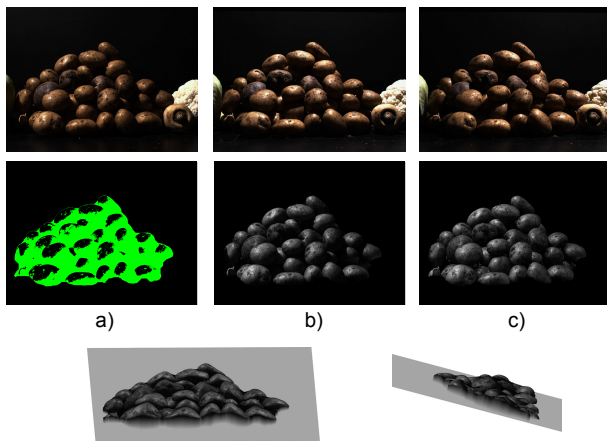


Fig. 10. Results on the *potatoes* sequence. First row: two sample images. Second row: a) image mask: green pixels represent missing entries; b) image used as initialization; c) resulting optimized image with a Lambertian model. Third row: two views of the reconstructed 3D surface after normal integration. The reference gray plane refers to the image plane of size 1600×1200 pixels. Left – top view showing the curvature of the objects. Right – side view showing the relief of the reconstruction.

is completely uncalibrated; strong shadows affect the images; finally the surface properties of the objects depart from the assumed Lambertian model. By thresholding highly saturated pixels and dark regions the overall ratio of missing data is 62.46% with a measurement matrix Y of nearly 32 million entries.

Figure 10 also shows an example of the missing data mask for some of the frames and it presents two views of the 3D reconstruction after normal integration. After aligning with the ground truth, we obtained a mean 3D error of 17.34% while the median 3D error was 8.4% denoting some localized regions with higher errors: mostly the areas in shadow and the frontal surfaces which are most affected by specularities.

5.7 Real data: 2D-3D Image registration

We also present an experiment with a real image sequence using a set of points tracked with an AAM algorithm. The first row of Figure 11 shows two frames of the sequence showing a face changing pose and expression. The bottom row shows the registered 2D points given a generic deformable 3D model $S_{12 \times 68}$ trained with a different data set of deforming faces. The lower row of Figure 11 shows the estimated 2D points along with the 3D shape fitted to the image measurements.

6 CONCLUSIONS

We have provided a novel optimization framework for a broad range of bilinear problems in Computer Vision with manifold constraints on the space where the data lies. Our results match state of the art methods in NRSfM and show a considerable improvement in performance

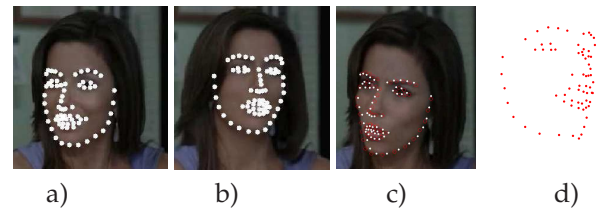


Fig. 11. a-b) The image sequence with tracking (white dots) using an AAM. c) registration results given the deformable model using BALM (red circles) and d) registered 3D shape.

on the PS problem, where our method can deal with 34% more missing data than previous solutions [48]. Our approach can deal with data matrices Y with 10^8 entries as demonstrated in the 3D reconstruction of the *Piazza San Marco* dataset. These features, together with its robustness to missing data, make BALM a good choice for bilinear modelling in large-scale inference scenarios.

REFERENCES

- [1] R. Basri, D. Jacobs, and I. Kemelmacher, "Photometric stereo with general, unknown lighting," *International Journal of Computer Vision*, vol. 72, no. 3, pp. 239–257, 2007.
- [2] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization approach," *International Journal of Computer Vision*, vol. 9, no. 2, 1992.
- [3] S. Thrun, "Affine structure from sound," *Advances in Neural Information Processing Systems*, vol. 18, p. 1353, 2006.
- [4] B. Bascle and A. Blake, "Separability of pose and expression in facial tracking and animation," in *Proc. 6th International Conference on Computer Vision, Bombay, India, 1998*, pp. 323–328.
- [5] C. Bregler, A. Hertzmann, and H. Biermann, "Recovering non-rigid 3D shape from image streams," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head, South Carolina, June 2000*, pp. 690–696.
- [6] J. Tenenbaum and W. Freeman, "Separating style and content with bilinear models," *Neural Computation*, vol. 12, no. 6, pp. 1247–1283, 2000.
- [7] M. Schlick, "Form and content: An introduction to philosophical thinking," *Philosophical Papers*, 1938.
- [8] P. Tresadern and I. Reid, "Articulated structure from motion by factorization," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition, San Diego, California, vol. 2, June 2005*, pp. 1110–1115.
- [9] R. Ferreira, J. Xavier, and J. Costeira, "Shape From Motion of Nonrigid Objects: The Case of Isometrically Deformable Flat Surfaces," *Proc. 20th British Machine Vision Conference, London, 2009*.
- [10] M. Chandraker and D. Kriegman, "Globally optimal bilinear programming for computer vision applications," in *IEEE Conference on Computer Vision and Pattern Recognition, 2008*, pp. 1–8.
- [11] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade, "Nonrigid structure from motion in trajectory space," in *Neural Information Processing Systems, 2008*.
- [12] B. Abboud and F. Davoine, "Bilinear factorisation for facial expression analysis and synthesis," *IEE Proceedings Vision, Image and Signal Processing*, vol. 152, no. 3, pp. 327–333, 2005.
- [13] J. Gonzalez-Mora, F. De la Torre, R. Murthi, N. Guil, and E. Zapata, "Bilinear active appearance models," in *IEEE 11th International Conference on Computer Vision, 2007*, pp. 1–8.
- [14] D. Shin, H. Lee, and D. Kim, "Illumination-robust face recognition using ridge regressive bilinear models," *Pattern Recognition Letters*, vol. 29, no. 1, pp. 49–58, 2008.
- [15] E. Chuang and C. Bregler, "Mood swings: expressive speech animation," *ACM Transactions on Graphics*, vol. 24, no. 2, pp. 331–347, 2005.

- [16] F. Cuzzolin, "Using bilinear models for view-invariant action and identity recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, New York, NY, vol. 2, 2006, pp. 1701–1708.
- [17] A. Elgammal and C. Lee, "Separating style and content on a nonlinear manifold," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Washington D.C., vol. 1, 2004.
- [18] F. Cole, "Automatic BRDF Factorization," Ph.D. dissertation, Harvard College, Cambridge, Massachusetts, 2002.
- [19] T. Wiberg, "Computation of principal components when data are missing," in *COMPSTAT 1976, Proc. Comput. Stat., 2nd Symp. Berlin*, 229–236., 1976.
- [20] T. Okatani and K. Deguchi, "On the wiberg algorithm for matrix factorization in the presence of missing components," *International Journal of Computer Vision*, vol. 72, pp. 329–337, May 2007.
- [21] A. M. Buchanan and A. Fitzgibbon, "Damped newton algorithms for matrix factorization with missing data," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, California, vol. 2, 2005, pp. 316–322.
- [22] D. Jacobs, "Linear fitting with missing data for structure-from-motion," *Computer Vision and Image Understanding*, vol. 82, no. 1, pp. 57 – 81, 2001.
- [23] S. Olsen and A. Bartoli, "Using priors for improving generalization in non-rigid structure-from-motion," *Proc. British Machine Vision Conference*, 2007.
- [24] P. Chen and D. Suter, "Recovering the missing components in a large noisy low-rank matrix: application to sfm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1051–1063, August 2004.
- [25] R. Guerreiro and P. Aguiar, "Factorization with missing data for 3d structure recovery," in *Multimedia Signal Processing, 2002 IEEE Workshop on*, December 2002, pp. 105 – 108.
- [26] R. I. Hartley and F. Schaffalitzky, "Powerfactorization: an approach to affine reconstruction with missing and uncertain data," in *Australia-Japan Advanced Workshop on Computer Vision*, Adelaide, Australia, September 2003.
- [27] L. Torresani, D. Yang, E. Alexander, and C. Bregler, "Tracking and modeling non-rigid objects with rank constraints," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Kauai, Hawaii, 2001.
- [28] P. Chen, "Optimization algorithms on subspaces: Revisiting missing data problem in low-rank matrix," *International Journal of Computer Vision*, vol. 80, no. 1, pp. 125–142, 2008.
- [29] M. Marques and J. Costeira, "Estimating 3d shape from degenerate sequences with missing data."
- [30] M. Paladini, A. Del Bue, J. Xavier, L. Agapito, M. Stosic, and M. Dodig, "Optimal metric projections for deformable and articulated structure-from-motion," *International Journal of Computer Vision*, pp. 1–25, 2011.
- [31] A. Shaji, S. Chandran, and D. Suter, "Manifold Optimisation for Motion Factorisation," in *19th International Conference on Pattern Recognition (ICPR 2008)*, 2008, pp. 1–4.
- [32] F. Mai and Y. S. Hung, "Augmented lagrangian-based algorithm for projective reconstruction from multiple views with minimization of 2d reprojection error," *Journal of Signal Processing Systems*, 2009.
- [33] Z. Lin, M. Chen, L. Wu, and Y. Ma, "The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices," *UIUC Technical Report UILU-ENG-09-2215*, 2009.
- [34] M. Guignard, "Lagrangean relaxation," *TOP – Sociedad de Estadística e Investigación Operativa*, vol. 11, no. 2, pp. 151–200, 2003.
- [35] M. Hestenes, "Multiplier and gradient methods," *Journal of Optimization Theory and Applications*, vol. 4, no. 5, pp. 303–320, 1969.
- [36] D. Bertsekas, *Constrained optimization and Lagrange multiplier methods*. Academic Press New York, 1982.
- [37] A. Del Bue, J. Xavier, L. Agapito, and M. Paladini, "Bilinear factorization via augmented lagrange multipliers," in *11th European Conference on Computer Vision (ECCV 2010)*, Crete, Greece, vol. 6314, 2010, pp. 283–296.
- [38] J. Xavier, A. Del Bue, L. Agapito, and M. Paladini, "Convergence Analysis of BALM," Tech. Rep., 2011, available from <http://users.isr.ist.utl.pt/~jxavier/technical-report-balm.pdf>.
- [39] A. Conn, N. Gould, A. Sartenaer, and P. Toint, "Convergence properties of an augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints," *SIAM Journal on Optimization*, vol. 6, no. 3, pp. 674–703, 1996.
- [40] J. Yan and M. Pollefeys, "A factorization-based approach for articulated non-rigid shape, motion and kinematic chain recovery from video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 5, May 2008.
- [41] L. Torresani, A. Hertzmann, and C. Bregler, "Non-rigid structure-from-motion: Estimating shape and motion with hierarchical priors." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 878–892, 2008.
- [42] A. Del Bue, F. Smeraldi, and L. Agapito, "Non-rigid structure from motion using ranklet-based tracking and non-linear optimization," *Image and Vision Computing*, vol. 25, no. 3, pp. 297–310, March 2007.
- [43] G. Wang, H. Tsui, and Q. Wu, "Rotation constrained power factorization for structure from motion of nonrigid objects," *Pattern Recognition Letters*, vol. 29, no. 1, pp. 72–80, 2008.
- [44] M. A. Lourakis and A. Argyros, "SBA: A Software Package for Generic Sparse Bundle Adjustment," *ACM Trans. Math. Software*, vol. 36, no. 1, pp. 1–30, 2009.
- [45] S. Agarwal, N. Snavely, S. Seitz, and R. Szeliski, "Bundle adjustment in the large," vol. 6314, pp. 29–42, 2010.
- [46] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, 2004.
- [47] H. Aanæs, A. Dahl, and K. Pedersen, "On recall rate of interest point detectors," in *3DPVT 2010: Fifth International Symposium on 3D Data Processing, Visualization and Transmission*, 2010.
- [48] C. Julia, A. Sappa, F. Lumbreras, J. Serrat, and A. Lopez, "Recovery of Surface Normals and Reflectance from Different Lighting Conditions," *Lecture Notes in Computer Science*, vol. 5112, pp. 315–325, 2008.



Alessio Del Bue received the Laurea degree in Telecommunication engineering in 2002 from University of Genova and his Ph.D. degree in Computer Science from Queen Mary University of London in 2006. He was a researcher in the Institute for Systems and Robotics (ISR) at the Instituto Superior Tecnico (IST) in Lisbon, Portugal. Currently, he is a Senior PostDoc researcher at the PAVIS department of the Istituto Italiano di Tecnologia (IIT) in Genova.



João Xavier (S97–M03) received the Ph.D. degree in electrical and computer engineering from the Instituto Superior Tecnico (IST), Lisbon, Portugal, in 2002. Currently, he is an Assistant Professor in the Department of Electrical and Computer Engineering, IST. He is also a Researcher at the Institute of Systems and Robotics (ISR), Lisbon, Portugal. His current research interests are in the area of optimization, sensor networks, and signal processing on manifolds.



Lourdes Agapito received the BSc degree in Physics in 1991 and the PhD in 1996 from the Universidad Computense in Madrid (Spain). She was then a Marie Curie fellow at Oxford's Robotics Research Group. She is currently a Reader in Computer Vision at the School of Electronic Engineering and Computer Science of Queen Mary, University of London. In 2008 she was awarded an ERC Starting Grant. Her research focuses in the area of 3D reconstruction of non-rigid structure from image sequences.



Marco Paladini received his MSc in Physics from Naples University Federico II in 2006. He is currently a PhD candidate at Queen Mary, University of London. His research interests are in the areas of non-rigid structure from motion, machine vision and robotics.