# WEIGHTED FACTORIZATION

*Pedro M. Q. Aguiar*, José M. F. Moura†*

Carnegie Mellon University
Pittsburgh PA, USA

## ABSTRACT

Factorization methods use linear subspace constraints to recover 3D rigid structure from 2D motion. Usually, these methods give equal weight to the contribution of each region (or feature) to the estimates of the 3D structure. In this paper, we accommodate different confidence weights for the 2D motion parameter estimates of each region, by rewriting the problem as the factorization of a *modified* matrix. This incurs no additional computational cost.

## 1. INTRODUCTION

Recovery of 3D shape and 3D motion (3D structure) from an image video sequence is useful in many areas. Among available approaches, factorization methods have become popular. These methods, originally introduced in [1], exploit the rigidity of the scene over a set of frames and use linear space constraints to solve for the larger number of unknowns. Reference [1] relies on tracking pointwise features. In prior work [2], we extend the factorization method to handle more general models that describe parametrically the 3D shape. This allows tracking larger regions where the image motion is described by a set of parameters and leads to simpler algorithms and better 3D structure recovery.

Clearly, the accuracy of the reconstruction will improve with better estimates of the 2D motion parameters. In turn, these estimates depend on the spatial variability of the brightness intensity pattern and on the size of the image region being tracked. The original factorization method [1] and the surface-based factorization method [2] give equal weight to the contribution of each feature or region to the final 3D shape and 3D motion estimates. Intuitively, however, we should expect that weighting more the trajectories corresponding to "sharp" features than the trajectories corresponding to features with smooth textures should lead to better overall estimates. In this paper, we develop such an approach, which leads to the factorization of a modified measurement matrix rather than the original matrix in [2]. Besides better performance, our method computes the weighted estimates with no additional cost.

Reference [3] considers reliability weights to address occlusion – when a feature is lost, it is given the weight zero – and uses an iterative method to recover the 3D structure. As reported in [3], the iterative method may fail to converge. We show here that when the weights are time-invariant, the problem can be reformulated as the non-weighted factorization of a *modified* matrix. Then, any method can be used to factorize this matrix. We call this extension the *weighted factorization* method.

## 2. SURFACE-BASED RANK 1 FACTORIZATION

To keep it simple, and without loss of generalization, we restrict the discussion to piecewise planar objects. The 3D motion of such objects induces on the image plane a parametric description for the 2D motion of the brightness pattern. We show in [2] that, at frame $f$, the image coordinates of the points belonging to the object surface are affine mappings of their image coordinates in frame 1. For each planar patch $n$, the parameters of the 2D affine motion model, represented by the $2 \times 1$ vector $\mathbf{d}_f^n$ and the $2 \times 2$ matrix $\mathbf{D}_f^n$, are related to the 3D motion parameters, in the $2 \times 2$ matrix $\mathbf{N}_f$ and the $2 \times 1$ vectors $\mathbf{n}_f$ and $\mathbf{t}_f$, and the 3D shape parameters corresponding to that patch, $a_{00}^n$ and the $2 \times 1$ vector $\mathbf{a}^n$, by

$$\begin{cases} \mathbf{d}_f^n = \mathbf{N}_f \mathbf{s}_0^n + \mathbf{n}_f a_{00}^n + \mathbf{t}_f \\ \mathbf{D}_f^n = \mathbf{N}_f + \mathbf{n}_f \mathbf{a}^{nT}, \end{cases} \tag{1}$$

where the $2 \times 1$ vector $\mathbf{s}_0^n$ is the centroid of the support region of the planar patch $n$. For $F$ frames and $N$ planar patches, there are $N(F-1)$ systems like (1), since we take the first frame as the reference.

The problem of inferring 3D rigid structure from 2D motion is formulated as the LS inversion of the $N(F-1)$ systems of equations (1) for the $F$ frames and $N$ planar patches. After replacing the translation estimates, $\mathbf{t}_f$, which are given by a closed-form expression, this system of equations (1) relating the image motion parameters and the 3D structure parameters is written in matrix format as

$$\mathbf{R} = \mathbf{M}\mathbf{S}^T, \tag{2}$$

*The first author is also affiliated with ISR-IST, Lisboa, Portugal. He was partially supported by INVOTAN.

†The second author is with the Dep. of Electrical Eng. and Computer Science, Massachusetts Institute of Technology, on leave from CMU.



0-7803-6297-7/00/$10.00 

where $\mathbf{R}$ is a $(2(F-1)\times 3N)$ matrix that collects the image motion parameters $\{\mathbf{d}_f^n, \mathbf{D}_f^N\}$,

$$\mathbf{R} = \begin{bmatrix} \mathbf{d}_2^1 & \mathbf{D}_2^1 & \cdots & \mathbf{d}_2^N & \mathbf{D}_2^N \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{d}_F^1 & \mathbf{D}_F^1 & \cdots & \mathbf{d}_F^N & \mathbf{D}_F^N \end{bmatrix}, \quad (3)$$

and $\mathbf{M}$ and $\mathbf{S}$ are $(2(F-1)\times 3)$ and $(3N\times 3)$ matrices that collect the 3D motion and 3D shape parameters, respectively, $\{\mathbf{N}_f, \mathbf{n}_f\}$ and $\{a_{00}^n, \mathbf{a}^n\}$, see [2]. A factorization approach solves for $\mathbf{M}$ and $\mathbf{S}$ from $\mathbf{R}$ by factorizing a rank 1 matrix $\widetilde{\mathbf{R}}$, see [2]. The matrix $\widetilde{\mathbf{R}}$ is the matrix $\mathbf{R}$ multiplied by the orthogonal projector onto the orthogonal complement of the space spanned by the first two rows of $\mathbf{S}^T$.

## 3. WEIGHTED FACTORIZATION

We consider the model in (1). Each 2D motion parameter of the surface patch $n$, in the vector $\mathbf{d}_f^n$ and the matrix $\mathbf{D}_f^n$, is observed with additive Gaussian white noise. For reasons that will become clear below, we assume the following: the two components of the vector $\mathbf{d}_f^n$ have equal variance, denoted by $(\sigma_d^n)^2$; the noise variances for the two entries of the first column of the matrix $\mathbf{D}_f^n$ are the same and denoted by $(\sigma_{D1}^n)^2$; and the noise variance for the two entries of the second column of the matrix $\mathbf{D}_f^n$ is the same and denoted by $(\sigma_{D2}^n)^2$. The variances $(\sigma_d^n)^2$ and $(\sigma_{Di}^n)^2$ are estimated from the spatial gradient of the image brightness pattern as described elsewhere.

We first estimate the 3D translation parameters. By choosing the origin of the object coordinate system in such a way that $\sum_n \mathbf{s}_0^n/(\sigma_d^n)^2 = [0,0]^T$ and $\sum_n a_{00}^n/(\sigma_d^n)^2 = 0$, we get the estimate

$$\widehat{\mathbf{t}}_f = \frac{\sum_{n=1}^N \frac{\mathbf{d}_f^n}{(\sigma_d^n)^2}}{\sum_{n=1}^N \frac{1}{(\sigma_d^n)^2}} \quad (4)$$

for the translation along the camera plane. Replacing these translation estimates in (1), and defining the matrices $\mathbf{R}$, $\mathbf{M}$, and $\mathbf{S}$ as in [2], we obtain, similarly to the non-weighted factorization, that

$$\mathbf{R} = \mathbf{M}\mathbf{S}^T. \quad (5)$$

To take into account the different variances of the errors of the entries of the measurement matrix $\mathbf{R}$, we define the $(2(F-1)\times 3N)$ weight matrix $\mathbf{W}$ as

$$\mathbf{W} = \begin{bmatrix} \frac{1}{\sigma_d^1} & \frac{1}{\sigma_{D1}^1} & \frac{1}{\sigma_{D2}^1} & \cdots & \frac{1}{\sigma_d^N} & \frac{1}{\sigma_{D1}^N} & \frac{1}{\sigma_{D2}^N} \\ \frac{1}{\sigma_d^1} & \frac{1}{\sigma_{D1}^1} & \frac{1}{\sigma_{D2}^1} & \cdots & \frac{1}{\sigma_d^N} & \frac{1}{\sigma_{D1}^N} & \frac{1}{\sigma_{D2}^N} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sigma_d^1} & \frac{1}{\sigma_{D1}^1} & \frac{1}{\sigma_{D2}^1} & \cdots & \frac{1}{\sigma_d^N} & \frac{1}{\sigma_{D1}^N} & \frac{1}{\sigma_{D2}^N} \end{bmatrix}, \quad (6)$$

where the entry $(i,j)$ of the matrix $\mathbf{W}$ represents the weight to be given to the entry $(i,j)$ of the matrix $\mathbf{R}$, see (3).

We formulate the weighted factorization by writing the *Maximum Likelihood* estimate,

$$\min_{\mathbf{M},\mathbf{S}} \left\| (\mathbf{R} - \mathbf{M}\mathbf{S}^T) \odot \mathbf{W} \right\|_F, \quad (7)$$

where the matrices $\mathbf{M}$ and $\mathbf{S}$ are constrained to have the special structure of the motion and shape matrices in [2]. The symbol $\odot$ denotes the elementwise product of two matrices, known as the *Hadamard product*.

Due to the structure of $\mathbf{W}$, we rewrite (7) as the factorization of a modified matrix $\mathbf{R}_w$,

$$\min_{\mathbf{M},\mathbf{S}_w} \left\| \mathbf{R}_w - \mathbf{M}\mathbf{S}_w^T \right\|_F, \quad (8)$$

where the matrices $\mathbf{R}_w$ and $\mathbf{S}_w$ are related to $\mathbf{R}$, $\mathbf{S}$, and the entries of the weight matrix $\mathbf{W}$ by

$$\mathbf{R}_w = \mathbf{R}\,\mathrm{diag}\left(\{w_{1i}, 1 \le i \le 3N\}\right), \quad (9)$$

$$\mathbf{S}_w = \mathrm{diag}\left(\{w_{1i}, 1 \le i \le 3N\}\right)\mathbf{S}, \quad (10)$$

where $\mathrm{diag}\left(\{w_{1i}, 1 \le i \le 3N\}\right)$ is a $3N \times 3N$ diagonal matrix whose entry $(i,i)$ is equal to the entry $(1,i)$ of $\mathbf{W}$.

The factorization in (8) is similar to the one studied in [2]. Note that the modified measurement matrix $\mathbf{R}_w$ and the first two rows of the matrix $\mathbf{S}_w^T$ are known from (10) and (9) and the motion matrix $\mathbf{M}$ is the same matrix involved in the rank 1 factorization method of [2]. We minimize (8) by using the rank 1 factorization procedure described in [2] to compute the factor matrices $\widehat{\mathbf{M}}$ and $\widehat{\mathbf{S}_w}$. To compute the estimate $\widehat{\mathbf{S}}$ of the shape matrix from the matrix $\widehat{\mathbf{S}_w}$, we invert (10),

$$\widehat{\mathbf{S}} = \mathrm{diag}\left(\left\{w_{1i}^{-1}, 1 \le i \le 3N\right\}\right)\widehat{\mathbf{S}_w}. \quad (11)$$

While [3] also considers reliability weights in estimating the matrices $\mathbf{M}$ and $\mathbf{S}$, in this case within the original feature-based factorization method, the solution in [3] is found by an iterative process that, as reported in [3], may fail to converge. In our formulation, this is not the case because we restrict the weight matrix $\mathbf{W}$ to have the structure of (6). For a general matrix $\mathbf{W}$, it is not possible to write the minimization (7) in the form of a factorization such as in (8). In fact, the unconstrained bilinear problem $\min_{\mathbf{M},\mathbf{S}} \left\| (\mathbf{R} - \mathbf{M}\mathbf{S}^T) \odot \mathbf{W} \right\|_F$ has a single global minimum, up to a scale factor, when $\mathbf{W}$ has rows that are all equal or columns that are all equal. It can be shown that this is not true for a generic matrix $\mathbf{W}$. In this case, the existence of local minima makes nontrivial the use of iterative numerical techniques.

## 4. EXPERIMENT

To illustrate the effect of taking into account the 2D motion estimation errors when recovering 3D structure from

2D motion, we synthesized two subsets of trajectories of feature points, represented in Figure 1, each with a different level of observation noise. We applied to the two sets of feature point trajectories both the non-weighted factorization and the weighted factorization methods.

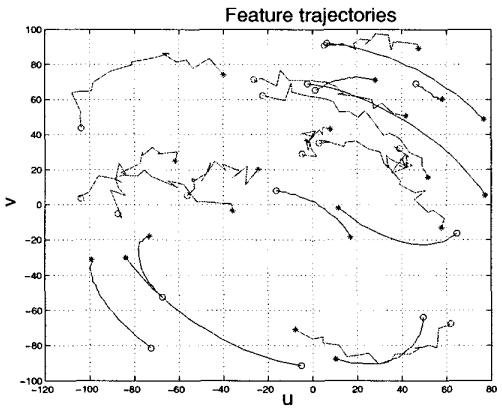


**Fig. 1**. Feature trajectories with two levels of noise.

The two subsets of features have 10 and 11 points with coordinates $x$ and $y$ randomly located inside a square. To facilitate the visualization of the experimental results, the depth $z$ was generated with a sinusoidal shape applied to the ordered set of 21 features. The coordinate $z$ is represented in both plots of Figure 2 with small circles. The dots on the plots of Figure 2 represent the estimates of the relative depth $z$. The 3D rotational motion was simulated by synthesizing a smooth time evolution for the Euler angles that specify the orientation of the object coordinate system relative to the camera coordinate system. The time evolution of the 6 entries of the 3D rotation matrix that are involved in the orthogonal projection, in matrix $\mathbf{N}_f$ and vector $\mathbf{n}_f$, is represented in both plots of Figure 4 with thick lines. The 3D translation is also smooth. The two components of the translation along the camera plane, in vector $\mathbf{t}_f$, are represented in the plots of Figure 3 with thick lines. The thin lines in the plots of Figures 3 and 4 represent estimates of the 3D motion parameters.

Figure 1 shows the feature trajectories on the image plane, after adding noise. For each trajectory, the initial position is marked with "o" and the final position is marked with "*". The noise variance is $\sigma_1^2 = 1$ for the first subset of 10 features and $\sigma_2^2 = 5$ for the second subset of 11 features. As expected, the trajectories corresponding to the features of the first subset have a smooth evolution (see trajectories clustered at the bottom of the figure), while the trajectories corresponding to the features of the second subset (top of figure) for which the noise has a larger variance have a more noisy shape, see Figure 1.

We applied both the non-weighted feature-based factorization and the weighted factorization method we describe

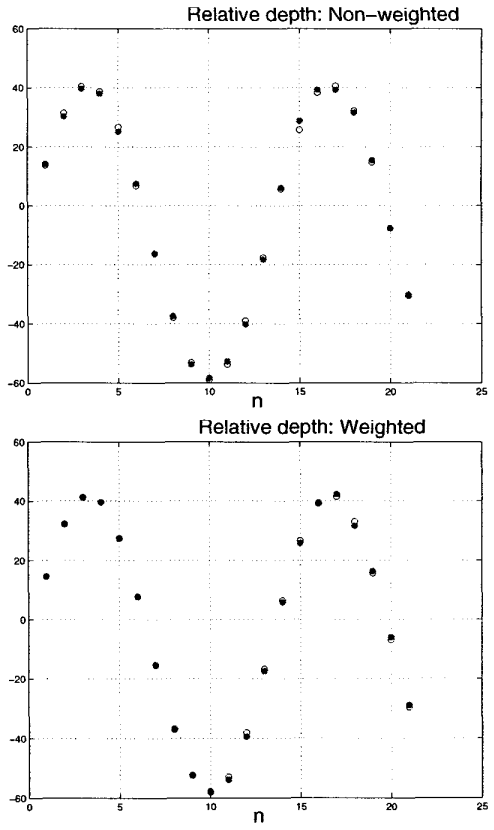


**Fig. 2**. Estimates of the relative depth, points, superimposed with the true value, circles.

in this paper to the feature trajectories of Figure 1. The estimates of the 3D shape and 3D motion are shown in the plots of Figures 2, 3, and 4, superimposed to the true values. In Figures 2, 3, and 4, the top plots represent the non-weighted estimates, and the bottom plots represent the weighted factorization results. We see that the 3D motion estimates obtained through the weighted factorization method are more accurate than the ones obtained without taking into account the different noise levels. This is particularly evident for the translation estimates – compare the top and bottom plots of Figure 3. The difference is smaller for the estimates of the entries of the 3D rotation matrix, see Figure 4, but these small differences originate much larger differences in the feature projections because the projections are obtained by multiplying the 3D rotation matrix by the 3D position of the features. The 3D shape estimates represented by the relative depths in the top and bottom plots of Figure 2 show a very good agreement with the real 3D shape, both for the non-weighted factorization and the weighted factorization methods. Thus, we conclude that while giving different credit for different trajectories improves the 3D motion estimates, the 3D shape estimate is almost independent of the weights

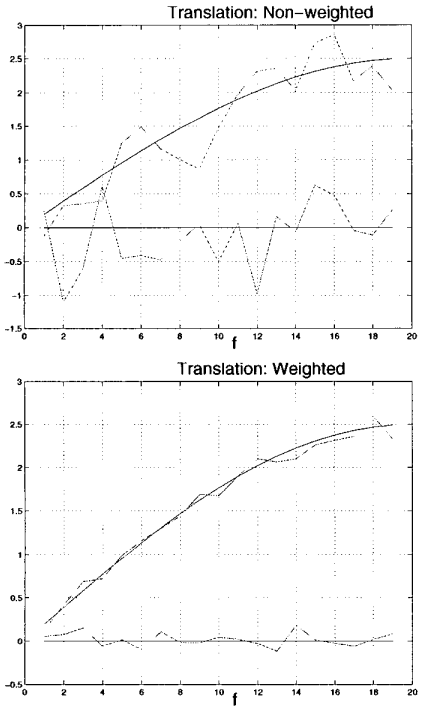


**Fig. 3**. Estimates of the translation along the camera plane, thin lines, superimposed to the true translation, thick lines.

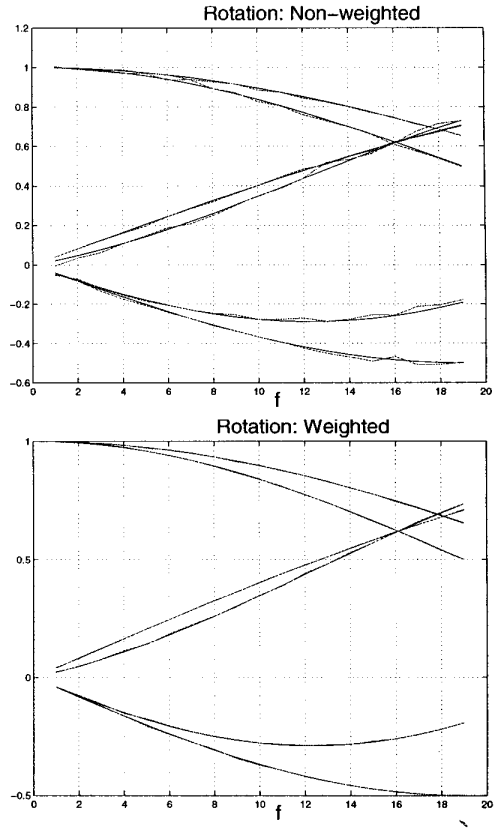


**Fig. 4**. Estimates of the 6 entries of the 3D rotation matrix that are involved in the orthogonal projection, thin lines, superimposed to the true values, thick lines.

given to the trajectories.

To interpret this behavior, we think of the estimation of the 3D shape as a filtering of the observations across time, and of the estimation of the 3D motion as a filtering of the observations across space. Since the weights given to the observations in the weighted factorization method are different from feature to feature, i.e., less weight is given to the more noisy trajectories, the filtering across space is improved and the weighted estimate of the 3D motion is much more accurate than the non-weighted estimate. In contrast, the weights are constant across time, thus filtering across time is insensitive to the different weights, and the weighted estimate of the 3D shape is similar to the non-weighted one.

The filtering analogies in the previous paragraph explain the experimental results in a coarse way. By examining in detail the estimates of the relative depth in both plots of Figure 2, we see that the weighted estimate of the 3D shape is slightly more accurate than the non-weighted estimate – compare the very accurate weighted estimates of the relative depths of the first 10 features (the subset of features observed with lower level of noise) in the bottom plot of Figure 2 with the non-weighted estimates in the top plot of Figure 2. The accuracy of the weighted estimate of the relative depth of a given feature reflects the level of the observation noise for the trajectory of the projection of that feature – note that the weighted estimates of the relative depths of the subset of features observed with higher level of noise (the last 11 features in the bottom plot of Figure 2) are less accurate than the estimates of the relative depths of the first subset (the first 10 features in the bottom plot of Figure 2).

The experiment just described demonstrates the overall improvement on the accuracy of the 3D structure estimates when the proposed *weighted factorization* method is used.

## 5. REFERENCES

[1] Carlo Tomasi and Takeo Kanade, "Shape and motion from image streams under orthography: a factorization method," *IJCV*, vol. 9, no. 2, pp. 137–154, 1992.

[2] Pedro M. Q. Aguiar and José M. F. Moura, "A fast algorithm for rigid structure from image sequences," in *IEEE ICIP*, Kobe, Japan, October 1999.

[3] Conrad J. Poelman, *A Paraperspective Factorization Method For Shape and Motion Recovery*, Ph.D. thesis, Carnegie Mellon University, USA, 1995.