A SIMPLE PRIOR FREE METHOD FOR NON-RIGID STRUCTURE FROM MOTION FACTORIZATION

Yuchao Dai , Hongdong Li , Mingyi He

April 27, 2013

About :

- This paper proposes a simple prior free method for solving non rigid structure from motion factorization problems.
- The success of the paper lies in the fact that it doesnot assume constraints on non rigidity o camera motions
- The paper discusses about the factorisation problem of measurement matrix and techniques employed to accurately extract non-rigid shape and camera motion.

Introduction :

0.1 Tomasi - kanade Factorization

Suppose we track P feature points over F frames in an image stream. Let the trajectories of image coordinates be (u_{fp}, v_{fp}) where$f = 1, 2, ...F$ and $p = 1, 2, ...P$.

The measurement matrix is formed as follows.

$$W = \begin{bmatrix} u_{11} & u_{12} & . & . & u_{1p} \\ u_{21} & . & . & . & u_{2p} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ v_{11} & v_{12} & . & . & v_{1p} \\ . & . & . & . & . \\ . & . & . & . & . \\ v_{f1} & v_{f2} & v_{f3} & . & v_{fp} \end{bmatrix}$$

- The individual elements are substracted from mean of each frame.
- $u_{ij} = u_{ij} - meanx_i$where $meanx_i = 1/P \sum u_{ij}$

- $v_{ij} = v_{ij} - meany_i$ where $meany_i = 1/P \sum v_{ij}$
- The new matrix is called registered rank matrix.
- Coordinate frame orientation is adjusted that world origin is at centroid of P world points whose P feature points are tracked in image stream.

With such a arrangement W can decomposed into RS where

$$R = \begin{bmatrix} i_1^{'} \\ i_2^{'} \\ . \\ . \\ .^{'} \\ j_1^{'} \\ j_2^{'} \\ . \\ . \\ j_f^{'} \end{bmatrix}$$ where i_1 and j_1 are along scan lines and columns of 1^{st} image in the stream. $(dimension = 2F \times 3)$

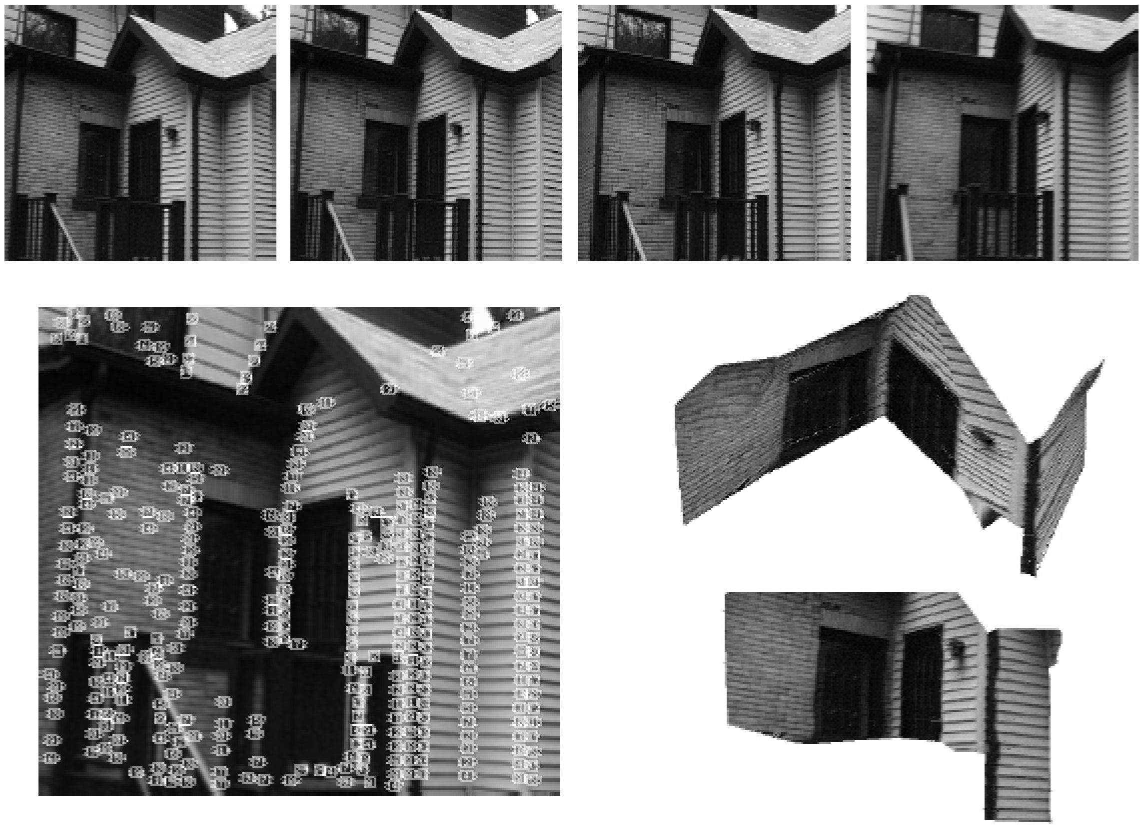
$S = \begin{bmatrix} s_1 & . & . & s_p \end{bmatrix}$

where S is the shape matrix .$(dimension = 3 \times P)$

Now we come most important and powerful theorem of factorization which is rank theorem .

RANK THEOREM:

- Without noise the registered measrement matrix is atmost rank 3.Even if the matrix is noisy the best possible shape and rotation estimate is obtained by considering only the three greatest singular values of W.
- By using SVD decomposition $W^{'} = U^{'}D^{'}V^{'}$ .Then we can assign $R^{'} = U^{'}D^{'}$ and $S^{'} = V^{'}$. This solution is true only up to 3×3 matrix Q because $R = R^{'}Q$ and $S = Q^{-1}S^{'}$ are also possible solutions.
- To find Q we introduce metric constraints for ex : $i_1 \times i_1^{'} = 1$ and $i_1 \times i_2^{'} = 0$.
- Even now the solution is determined to a 3×3 rotation matrix hence another constraint that world frame is initially parallel to first frame.
- Finally rotation matrix and shape matrix can be found for each frame and the results were as follows:



0.2 Bregler's work:

Abstract :

- This paper addresses the problem of recovering 3D non-rigid shape models from image sequences.
- Here 3D shapes in each frame is a linear combination of a set of basis.
- This paper demonstrates a new technique for recovering 3D non-rigid shape models from 2D image sequences recorded with a single camera.
- All existing methods for nonrigid 3D shapes either require an a-priori model or require multiple views.
- This is the first algorithm that can tackle this problem without the use of a prior model and without multiple view or other 3D input.

Explanation:

- The shape of the non-rigid object is described as a keyframe basis setS_1, S_2, S_3........

- The shape of a specific configuration is a linear combination of basis set.

$S = \sum_{i=1}^{k} L_i S_i (dimension = 3 \times P)$

$\begin{bmatrix} u1 & u2 & . & . & up \\ v1 & v2 & . & . & vp \end{bmatrix} = R \times S + T(u_i, v_i)$are 2-D image points in a frame.

where$R = \begin{bmatrix} r1 & r2 & r3 \\ r4 & r5 & r6 \end{bmatrix}$contains first two rows of rotation matrix.

- We eliminate T by subtracting the mean of all 2D points, and henceforth can assume that S is centered at the origin.

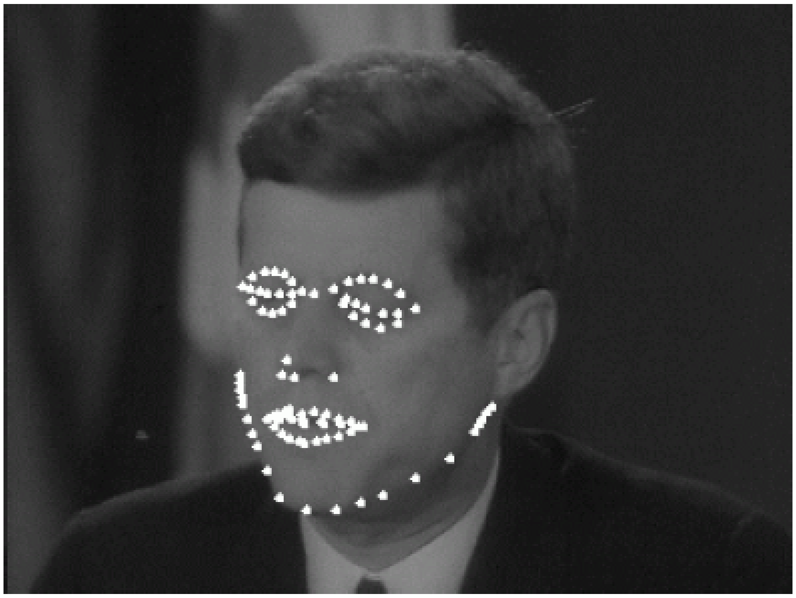
- After adding temporal index to each frame we form measurement matrix W.

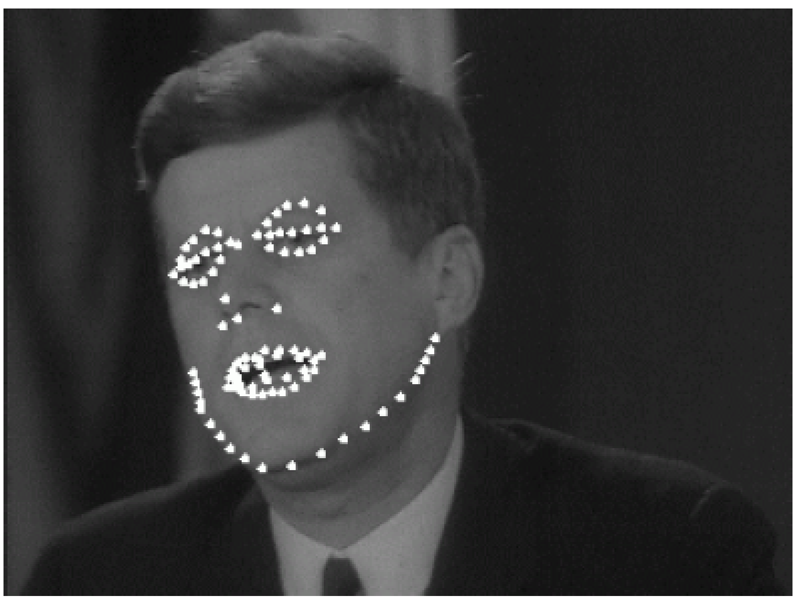
$$W = \begin{bmatrix} l_1^1 R^1 & . & . & . & . & l_k^1 R^1 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ l_1^N R^N & . & . & . & . & l_k^N R^N \end{bmatrix} \times \begin{bmatrix} S_1 \\ S_2 \\ . \\ . \\ S_K \end{bmatrix} \text{for N frames}$$

$= QB(dimension of Q = 2N \times 3K and B = 3K \times P)$

Q can be factorised into configuration weights and pose matrix in each frame .

The transformation matrix G maps all R_t into an orthonormal$R_t = R_t^{'} \times G$. which is found using least square methods.





Track the eye brows, upper and lower eye lids, 5 nose points, outer and inner boundary of the lips, and the chin contour using

apperance based 2D tracking technique.

Paper work:

- In a 2004 paper, Xiao et al. proved that the problem itself is indeed ill-posed or under-constrained, in the sense that, based on the orthonormality constraint alone, one cannot recover the non-rigid shape bases and the corresponding shape coefficients uniquely.
- To resolve this ambiguity, Xiao et al. suggested to add extraneous "basis constraints" so as to make the system well-constrained.
- Akhter et al. made an important theoretical progress in [2] which reveals that: although the ambiguity in shape basis is inherent, the 3D shape itself can be recovered uniquely without ambiguity.
- Based on the linear combination model , the non-rigid shape Si belongs to $R(3 \times P)$ can be represented as a linear combination of K shape bases Bk belong to $R(3 \times P)$ with shape coefficients cik as$S_i = \sum cikBk$
- Under orthographic camera model, the coordinates of the 2D image points observed at frame i are given by $W_i = R_i S_i$ where R_i contains first two rows of ith camera rotation.

$$W = \begin{bmatrix} C_{11}R_1 & \dots & C_{1k}R_1 \\ \vdots & \ddots & \vdots \\ C_{f1}R_1 & \cdots & C_{fk}R_f \end{bmatrix} \times \begin{bmatrix} B_1 \\ \vdots \\ \vdots \\ B_k \end{bmatrix}$$

$= R(C \times I_3)B$ where $R = blkdiag(R_1, R_2, R_3............)$
$= \Pi \times B$
(dimension of $\Pi = 3 \times k$ and dimension of $B = 3 \times k$)

However this decomposition is not unique as any non singular matrix G can be inserted between as $\Pi G G^{-1} B$. This G which rectifies Π is called Euclidean corrective matrix .

Orthonormality constraints in the Π matrix can be imposed to recover a Gram Matrix Q_k such that $Q_k = G_k {G^T}_k$.(dimension of $Q_k = 3k \times 3k$)

- Akhter et al showed that the fundamental ambiguity does not necessarily lead to an ambiguous shape . Orthonormality constraints alone is in fact sufficient to recover a unique (unambiguous) non-rigid shape (provided that a previously-overlooked rank-3 constraint on Q_k)
- Denote ith double rows of $\widehat{\Pi}$ as $\widehat{\Pi}_{2i-1:2i}$and kth triplet og G as G_k.

$\Pi_{2i-1} * Q_k * \Pi'_{2i-1} = {}_{2i} * Q_k * {\Pi'}_{2i}, \Pi_{2i-1} * Q_k * {\Pi'}_{2i} = 0$——- (Eq)

Theorem:

All the solutions of $Q_k$to linear system of (Eq) form a linear subspace of dimension $2K^2 - K$.

The above equation is converted into the form $Aq_k = 0$ where $q_k = vec(Q_k)$ is $9k^2$column vector . q_k is the nullspace of the vector A .

Intersection theorem:

Under nondegenerate and noise-free conditions, any correct solution of Qk (i.e. the Gram matrix of a column-triplet of the true Euclidean corrective matrix Gk) must lie in the intersection of the (2K2–K) dimensional null-space of A and a rank-3 positive semi-definite matrix cone, i.e., Qk belongs to

$Avec(Q_k) = 0 \bigcap Qk \succ 0 \bigcap rank(Q_k) = 3$

The three steps to solve the above algorithm are

- Measurement noise may increase rank of Q_k and the problem is to minimise $rank(Q_k)$. Nuclear norm minimization is the tightest convex relaxation for the rank minimization problem (above) which is NP-hard. Now $trace(Q_k^*) = rank(Q_k^*)$. So the problem reduces to $min(trace(Q_k))$ such that $Avec(Q_k) = 0$ and Q_k is semi-definite matrix.This is a standard semi definite programming problem (SDP) of fixed size $(2k^2 - k)$.

Once Gk find similarly $G_1, G_2, ...$ and form G matrix.

- Once G_k is solved, the rotation at every frame $i = 1, \cdots, F$ can be solved by using

$\hat{\Pi}_{2i-1:2i} G_k = C_{ik} R_i$, $i = 1, 2, ...F$

- Estimate S by Rank Minimisation

valid solution to shape matrix is formulated as following rank minisization problem:

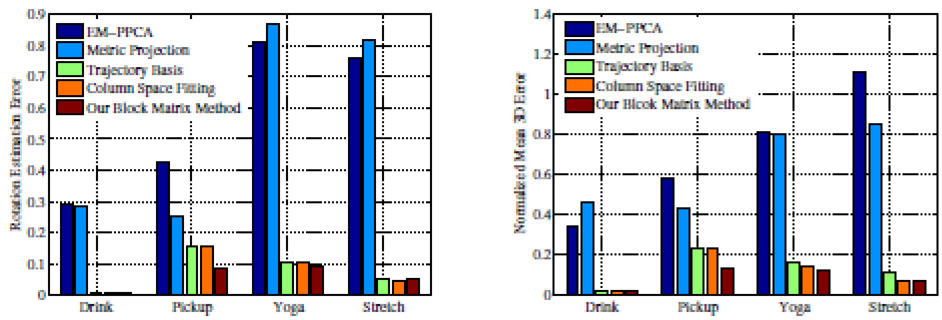
$rank(s) \leq 3 \times k$

$W = RS$

The solution to the above problem is unique and is given as

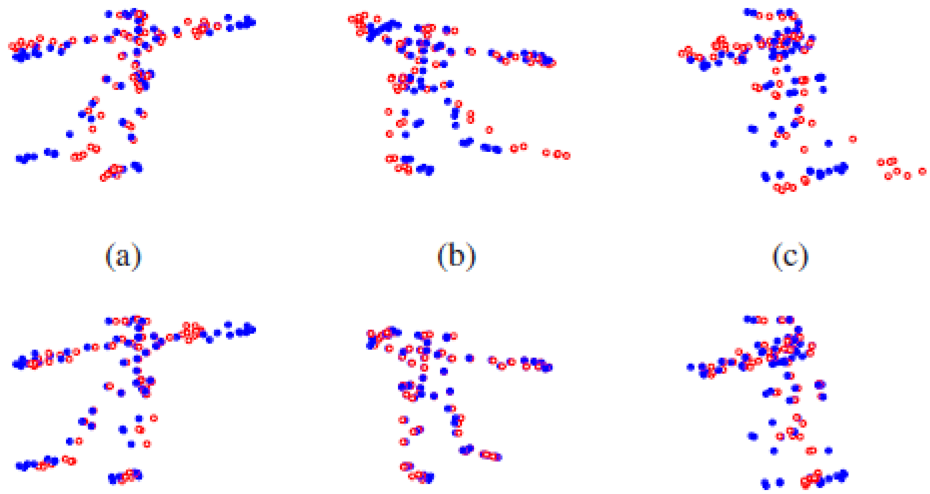
$S = R^\dagger W = (R^T (RR^T)^{-1})W.$

Experiments:



(a) (b)

Figure 3. Motion capture data experimental results. **Left**: Rotation estimation error; **Right**: 3D reconstruction error.



(a) (b) (c)

(d) (e) (f)

Figure 6. Comparison of the 3D reconstruction results on the Dance sequence. The blue dots are the ground truth 3D points, and the red circles show the reconstructed points. **Top row:** results by the trajectory basis method [3], where the 3D errors are **0.3011, 0.2827, 0.2814** for the 3 frames. **Bottom row:** our result by the block matrix method, where the 3D errors are **0.2228, 0.0355, 0.1389** for the 3 frames.

Conclusion:

This paper advocates a novel prior-free approach to nonrigid factorization. Our method is purely convex, very easy to implement, and is guaranteed to converge to an optimal solution

(at least approximately up to certain relaxation).