Linear Motion Estimation for Systems of Articulated Planes

Ankur Datta Yaser Sheikh Takeo Kanade
Robotics Institute
Carnegie Mellon University
{ankurd,yaser,tk}@cs.cmu.edu

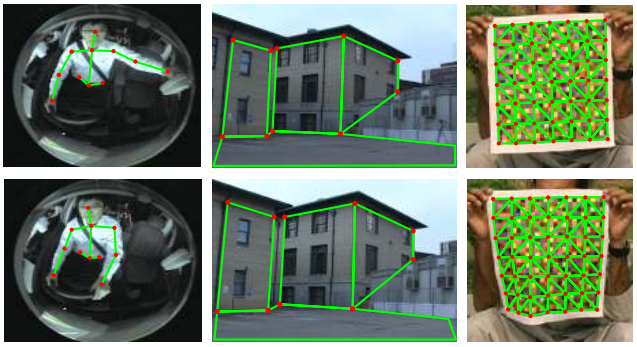
Abstract

In this paper, we describe the explicit application of articulation constraints for estimating the motion of a system of planes. We relate articulations to the relative homography between planes and show that for affine cameras, these articulations translate into linear equality constraints on a linear least squares system, yielding accurate and numerically stable estimates of motion. The global nature of motion estimation allows us to handle areas where there is limited texture information and areas that leave the field of view. Our results demonstrate the accuracy of the algorithm in a variety of cases such as human body tracking, motion estimation of rigid, piecewise planar scenes and motion estimation of triangulated meshes.

1. Introduction

The principal challenge in developing general purpose motion estimation algorithms is the variety of rigid and nonrigid motions encountered in the real world. Consider the three examples shown in Figure 1. In the first image pair, the motion of a human is shown where each limb is able to move with six degrees of freedom. Motion of a rigid scene, shown in the second image pair, is induced by the confluence of the structure of the scene and the motion of the camera. Finally, the motion of a nonrigid object such as the cloth in the third image pair depends on the elasticity of the object and the force acting on it. In computer vision, the problem of motion estimation for varied objects such as these has resulted in the proposition of a large number of algorithms [17, 2, 26, 3, 7, 29, 1, 16, 14]. In particular, due to their wide applicability, layered motion models have gained significant traction over the years [28, 23, 30]. However, existing layers based motion algorithms do not exploit a key constraint that exists in the motion of a large number of real scenes.

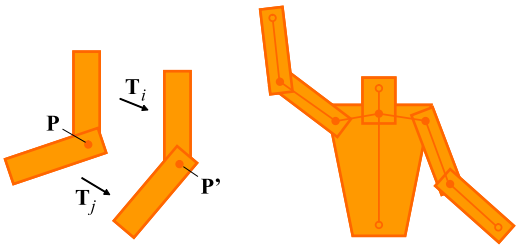
In this paper, we demonstrate that *articulation constraints* are important in many common scenarios for motion estimation and yield useful constraints when taken into account explicitly. Articulation constraints posit the existence of points where the motion of a pair of planes is equal. For instance, even though a human body can move in a variety of complex ways, one constraint that must be followed is that the motion of the upper and lower arm must move the elbow to the same position (Figure 1(a)). Rigid, piecewise planar scenes also observe this constraint because the motion on the line of intersection of any two planes is the same for the two planes. For nonrigid surfaces, a triangulated mesh is a popular representation. Each vertex, shared by multiple triangles, must also move to the same position under the motion of all those triangles.



(a) (b) (c)

Figure 1. Examples of articulated motion (a) Motion of human body limbs are dependent on each other. (b) Motion of the facades of a building are dependent on each other and on the ground plane. (c) A popular choice for parameterizing the motion of a nonrigid surface is a triangulated mesh, where the motion of each triangle is dependent on the its neighboring triangles.

We address the problem of motion estimation for both rigid and nonrigid entities by taking articulation constraints explicitly into account. We study the relationship between articulations and the homographies induced by articulated planes (Section 2). Unlike previous constraints [14, 26], we define exact equality constraints on the motion model of the articulated planes (Section 3), and propose a motion estimation algorithm that solves a linear equality constrained least squares system for the motion of multiple planes simulta-



(a) (b)

Figure 2. Articulation Constraints (a) Articulations transform identically under the transformations of two planes (b) Singly articulated planes as a model for body tracking. Five points connect six body parts.

neously (Section 4). Our results demonstrate that motion is estimated accurately for a variety of settings such as human body tracking, estimating motion in rigid, piecewise planar scenes and estimating the motion of nonrigid surfaces (Section 5).

2. Articulated Planes

Between a pair of planes $(\mathbf{\Pi}_i, \mathbf{\Pi}_j)$ in $\mathbb{R}^3$ undergoing 3D Euclidean transformations $(\mathbf{T}_i, \mathbf{T}_j)$ respectively, an *articulation* $\mathbf{P}$, is a 3D point that moves identically under the action of both $\mathbf{T}_i$ and $\mathbf{T}_j$. There can be at most two such points between planes since if there are three non-collinear articulations the two moving planes are, in fact, the same plane. Singly articulated planar systems are a popular model of the human body [14, 15] (see Figure 2) and what can be considered doubly articulated planar systems have found application in shadow analysis, view synthesis and in scene reconstruction, [18, 12, 21]. Under the action of a projective camera, the motion field induced by a moving plane can be described by a homography,

$$\begin{bmatrix} sx' \\ sy' \\ s \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \qquad (1)$$

that is $\mathbf{x}' \cong \mathbf{H}\mathbf{x}$ where $\mathbf{x}, \mathbf{x}' \in \mathbb{P}^2$ and $\mathbf{H}$ is a nonsingular 3×3 matrix. The motion fields induced between a pair of articulated planes are not independent and their dependencies physically manifest themselves in 2D motion as well. Let $\mathbf{p}$ be the image of $\mathbf{P}$ and let $\mathbf{H}_i$ and $\mathbf{H}_j$ be the respective homographies induced by the motion of the two planes. Since $\mathbf{p}$ is the image of an articulation, it follows that,

$$\mathbf{p}^{'} \cong \mathbf{H}_i\mathbf{p} \cong \mathbf{H}_j\mathbf{p}. \qquad (2)$$

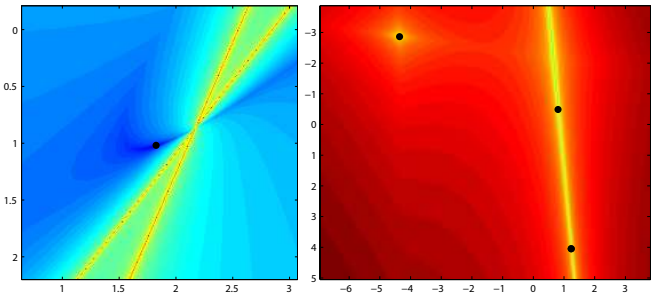
2D articulations can be computed directly from the pair of homographies by noting that they are related to the fixed or united points ([25]) of the relative homography $\mathbf{\Omega}_{ij} = \mathbf{H}_i^{-1}\mathbf{H}_j$. The 2D articulations correspond to eigenvectors of $\mathbf{\Omega}_{ij}$ (and $\mathbf{\Omega}_{ji}$). This can be seen since

$$s_i^k\mathbf{H}_i\mathbf{p}_k = \mathbf{p}'_k, \;\; s_j^k\mathbf{H}_j\mathbf{p}_k = \mathbf{p}'_k. \qquad (3)$$

Since $\mathbf{H}_i$ is non-singular and real,

$$(\mathbf{H}_j^{-1}\mathbf{H}_i - \lambda_k\mathbf{I})\mathbf{p}_k = 0, \qquad (4)$$

where $\lambda_k = \frac{s_j^k}{s_i^k}$ and $\mathbf{I}$ is a 3×3 identity matrix. Thus, given $(\mathbf{H}_1, \mathbf{H}_2)$, finding all $\mathbf{p}$ that satisfy Equation 3 is the generalized eigenvalue problem. From Equation 4, each λ_k is an eigenvalue and each $\mathbf{p}_k$ is an eigenvector of $\mathbf{H}_j^{-1}\mathbf{H}_i$. To illustrate the meaning of articulations in terms of optic motion, the absolute difference in motion fields generated by two homographies is shown in Figure 3. The location of the eigenvectors of the relative homography are marked by black dots. It should be noted that all eigenvectors of the relative homography do not necessarily correspond to 3D articulations. A relevant example is that of a pair of moving planes fixed with respect to each other. The relative homography in this case is a planar homology ([18]). Two eigenvectors are images of points that lie on the fixed line of intersection (which can be considered a stationary articulation) but the third eigenvector does not correspond to any 3D articulation (see Figure 3(b)).



(a) (b)

Figure 3. Magnitude of the difference between the motion fields induced by two homographies. The black dots denote the real eigenvectors of the relative homography. (a) From homographies induced by two identical planes rotating in opposite directions about a common point. (b) From homographies whose relative homography is a planar homology. Note that two points lie on a line of fixed points.

Conversely, knowledge of articulations can be used to constrain the estimation of homographies. To eliminate the effects of scale, we can rewrite equation 2 as,

$$\mathbf{H}_i\mathbf{p} \times \mathbf{H}_j\mathbf{p} = \mathbf{0}. \qquad (5)$$

Equation 5 can be rearranged to yield three relationships,

$$\begin{aligned} \mathbf{p}^T C_1\mathbf{p} &= 0 \\ \mathbf{p}^T C_2\mathbf{p} &= 0 \\ \mathbf{p}^T C_3\mathbf{p} &= 0, \end{aligned}$$

where the conics C_1, C_2 and C_3 are functions of the two homographies H_i and H_j. Each articulation satisfies the three conic equations. Thus, the constraints induced by articulation are quadratic in terms of the elements of $(\mathbf{H}_i, \mathbf{H}_j)$. In the next section we show that if affine cameras are assumed, these constraints are simplified into linear constraints, suitable for numerically stable and accurate estimation of motion.

3. Articulation Constraints for Affine Cameras

For affine cameras, the motion induced between two views of a plane is represented by an affine transformation,

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \tag{6}$$

or equivalently $\mathbf{x}' = \mathbf{A}_i\mathbf{x}$. Between plane $\mathbf{\Pi}_i$ and plane $\mathbf{\Pi}_j$ articulated at $\mathbf{p}$, the articulation constraint takes a particularly simple form,

$$\mathbf{A}_i\mathbf{p} = \mathbf{p}' = \mathbf{A}_j\mathbf{p}. \tag{7}$$

Equation 7 can be rewritten as,

$$(\mathbf{A}_i - \mathbf{A}_j)\mathbf{p} = \mathbf{0}, \tag{8}$$

and therefore the null-vector of $(\mathbf{A}_i - \mathbf{A}_j)$ is $\mathbf{p}$.

We also observe that for a pair of affine transformations, $(\mathbf{A}_i, \mathbf{A}_j)$, with two articulations, $\mathbf{p}_1$ and $\mathbf{p}_2$, any point on the line defined by $\mathbf{p}_1$ and $\mathbf{p}_2$ is also a articulation. All points that lie on the line defined by the articulations $\mathbf{p}_1$ and $\mathbf{p}_2$ can be expressed through the convex relationship $\mathbf{p}_3 = \alpha\mathbf{p}_1 + (1-\alpha)\mathbf{p}_2$. Since $\mathbf{p}_1$ and $\mathbf{p}_2$ are articulations, from Equation 7,

$$\mathbf{A}_i\mathbf{p}_1 = \mathbf{p}_1' \quad \mathbf{A}_j\mathbf{p}_1 = \mathbf{p}_1',$$
$$\mathbf{A}_i\mathbf{p}_2 = \mathbf{p}_2' \quad \mathbf{A}_j\mathbf{p}_2 = \mathbf{p}_2'.$$

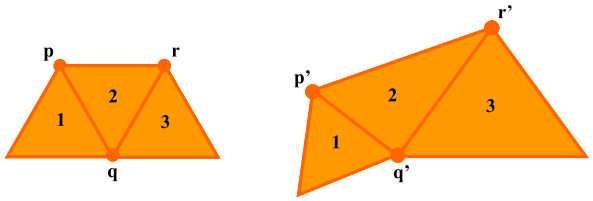
We can see that when $\mathbf{p}_3$ is transformed by $\mathbf{A}_i$ and $\mathbf{A}_j$ we get,

$$\begin{aligned} \mathbf{A}_i\mathbf{p}_3 &= \mathbf{A}_i(\alpha\mathbf{p}_1 + (1-\alpha)\mathbf{p}_2) \\ &= \alpha\mathbf{A}_i\mathbf{p}_1 + (1-\alpha)\mathbf{A}_i\mathbf{p}_2 = \alpha\mathbf{p}_1' + (1-\alpha)\mathbf{p}_2' \\ &= \alpha\mathbf{A}_j\mathbf{p}_1 + (1-\alpha)\mathbf{A}_j\mathbf{p}_2 = \mathbf{A}_j(\alpha\mathbf{p}_1 + (1-\alpha)\mathbf{p}_2) \\ &= \mathbf{A}_j\mathbf{p}_3, \end{aligned}$$

and therefore any point $\mathbf{p}_3$ that lies on the line defined by two articulations of a pair of affine transform is itself an articulation. This property is useful when considering motion estimation over triangulated meshes (Figure 4) as it ensures that tears do not occur while warping the underlying images.

Finally, a remark on the linear dependencies of constraints from articulations between multiple (≥ 3) planes. For a system such as the one shown in Figure 4, there are five unique articulations[1]: $\mathbf{p}_{12}$, $\mathbf{q}_{12}$, $\mathbf{r}_{23}$, $\mathbf{q}_{23}$ and $\mathbf{q}_{13}$. However, there are only four linearly independent constraints since the constraint produced by $\mathbf{q}_{13}$ is linearly dependent on those of $\mathbf{q}_{12}$ and $\mathbf{q}_{23}$.

[1] $\mathbf{a}_{ij}$ refers to the articulation $\mathbf{a}$ between triangles i and j.



(a) (b)

Figure 4. A system of three triangles sharing three articulations (a) before and (b) after motion.

4. Articulated Motion Estimation

In this section, we describe how to use articulation constraints in the estimation algorithm proposed by Bergen *et al*. [2]. By making the brightness constancy assumption between corresponding pixels in consecutive frames, the motion estimation process involves SSD minimization,

$$E(\{\mathbf{a}\}) = \sum_{\mathbf{x}} \Big(I_t(\mathbf{x}) - I_{t+1}\big(W(\mathbf{x}|\{\mathbf{a}\})\big) \Big)^2, \tag{9}$$

where W is a warp function, $\{\mathbf{a}\}$ are the motion parameters. Gauss-Newton minimization is used to estimate the motion parameters. Thus, applying a first order approximation yields the optical flow constraint equation,

$$\nabla I_x u + \nabla I_y v + \nabla I_t = 0, \tag{10}$$

where ∇I_x, ∇I_y and ∇I_t are the spatiotemporal image gradients and $u = x' - x$ and $v = y' - y$ are the horizontal and vertical components of the optical flow vector. Under an affine transformation,

$$x' = a_1x + a_2y + a_3, \tag{11}$$
$$y' = a_4x + a_5y + a_6. \tag{12}$$

Equations 10, 11 and 12 can be combined to create a linear system of equations in the unknown values $\mathbf{a} = [a_1, \cdots a_6]^\top$. Thus, in a system of planes, for the i-th plane we have,

$$\mathbf{\Lambda}_i(\nabla I_x, \nabla I_y, \nabla I_t)\mathbf{a}_i = \mathbf{b}_i(\nabla I_x, \nabla I_y, \nabla I_t), \tag{13}$$

where $\mathbf{\Lambda_i}$ and $\mathbf{b_i}$ define the same linear system as in [2]. For two planes $\mathbf{\Pi}_i$ and $\mathbf{\Pi}_j$, their independent linear systems may be combined by means of a direct sum into a larger system,

$$\begin{bmatrix} \mathbf{\Lambda}_i(\nabla\mathbf{I}) & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_j(\nabla\mathbf{I}) \end{bmatrix} \begin{bmatrix} \mathbf{a}_i \\ \mathbf{a}_j \end{bmatrix} = \begin{bmatrix} \mathbf{b}_i(\nabla\mathbf{I}) \\ \mathbf{b}_j(\nabla\mathbf{I}) \end{bmatrix}. \tag{14}$$

Solving the system in Equation 14 is equivalent to solving individually for each plane. However, if $\mathbf{\Pi}_i$ and $\mathbf{\Pi}_j$ share an articulation $\mathbf{p}$, the affine transformations $\mathbf{A}_i$ and $\mathbf{A}_j$ are related as described in Equation 8. In terms of $[\mathbf{a}_i \ \ \mathbf{a}_j]^\top$ this constraint can be written as,

$$\begin{bmatrix} \mathbf{p} & \mathbf{0} & -\mathbf{p} & \mathbf{0} \\ \mathbf{0} & \mathbf{p} & \mathbf{0} & -\mathbf{p} \end{bmatrix} \begin{bmatrix} \mathbf{a}_i \\ \mathbf{a}_j \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad (15)$$

or simply $[\theta(\mathbf{p}) \ \theta(-\mathbf{p})][\mathbf{a}_i \ \ \mathbf{a}_j]^\top = \mathbf{0}$. Estimating $[\mathbf{a}_i \ \ \mathbf{a}_j]^\top$ from Equations 14 and 15 is a standard equality constrained linear least squares problem which can be solved stably as described in Appendix A or by standard optimization packages (such as `lsqlin` in Matlab). For further details on such optimization the interested reader is directed to [9].

For more than two planes with pairwise articulations, such as the case in Figure 2 (b), this analysis can be used to globally constrain the estimate of the planes. Each pairwise articulation introduces a pair of constraints on the affine parameters of the system. For n planes with k articulations, we have $6n$ affine parameters and $2k$ equality constraints. The matrix in Equation 14 would be expanded into a block diagonal matrix with n blocks.

$$\begin{bmatrix} \mathbf{\Lambda}_1 & & \\ & \ddots & \\ & & \mathbf{\Lambda}_n \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_n \end{bmatrix}. \qquad (16)$$

Each of the k articulations would provide two constraints that can be directly encoded in a single matrix. As an illustration, consider the following linear equations for the system in Figure 4,

$$\begin{bmatrix} \mathbf{\Lambda}_1 & 0 & 0 \\ 0 & \mathbf{\Lambda}_2 & 0 \\ 0 & 0 & \mathbf{\Lambda}_3 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix}. \qquad (17)$$

or in matrix form, $\mathbf{\Gamma}\mathcal{A} = \mathcal{B}$.

The corresponding constraint equations for the system would be,

$$\begin{bmatrix} \theta(\mathbf{p}) & \theta(-\mathbf{p}) & \mathbf{0} \\ \theta(\mathbf{q}) & \theta(-\mathbf{q}) & \mathbf{0} \\ \mathbf{0} & \theta(\mathbf{q}) & \theta(-\mathbf{q}) \\ \mathbf{0} & \theta(\mathbf{r}) & \theta(-\mathbf{r}) \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix} = \mathbf{0}. \qquad (18)$$

or in matrix form, $\mathbf{\Theta}\mathcal{A} = 0$.

From commutativity, it should be noted that the motion of $\mathbf{p}$ is not independent of the motion of $\mathbf{\Pi}_3$ even though an explicit connection is not present. The network of articulations place a constraint on the global motion estimation of the system of planes.

5. Applications

We have conducted several experiments to evaluate our motion estimation algorithm for a wide variety of motions that occur in real scenes. In particular, we evaluated our algorithm on the specific tasks of estimating the motion of the upper body of a human, estimating the motion of rigid, piecewise planar scenes with low texture planes, and finally on estimating the motion of several nonrigid surfaces.

Objective
Given 2 images, P articulations and the support of each of the N planes, estimate the motion of the system of articulated planes.

Algorithm
Do until convergence

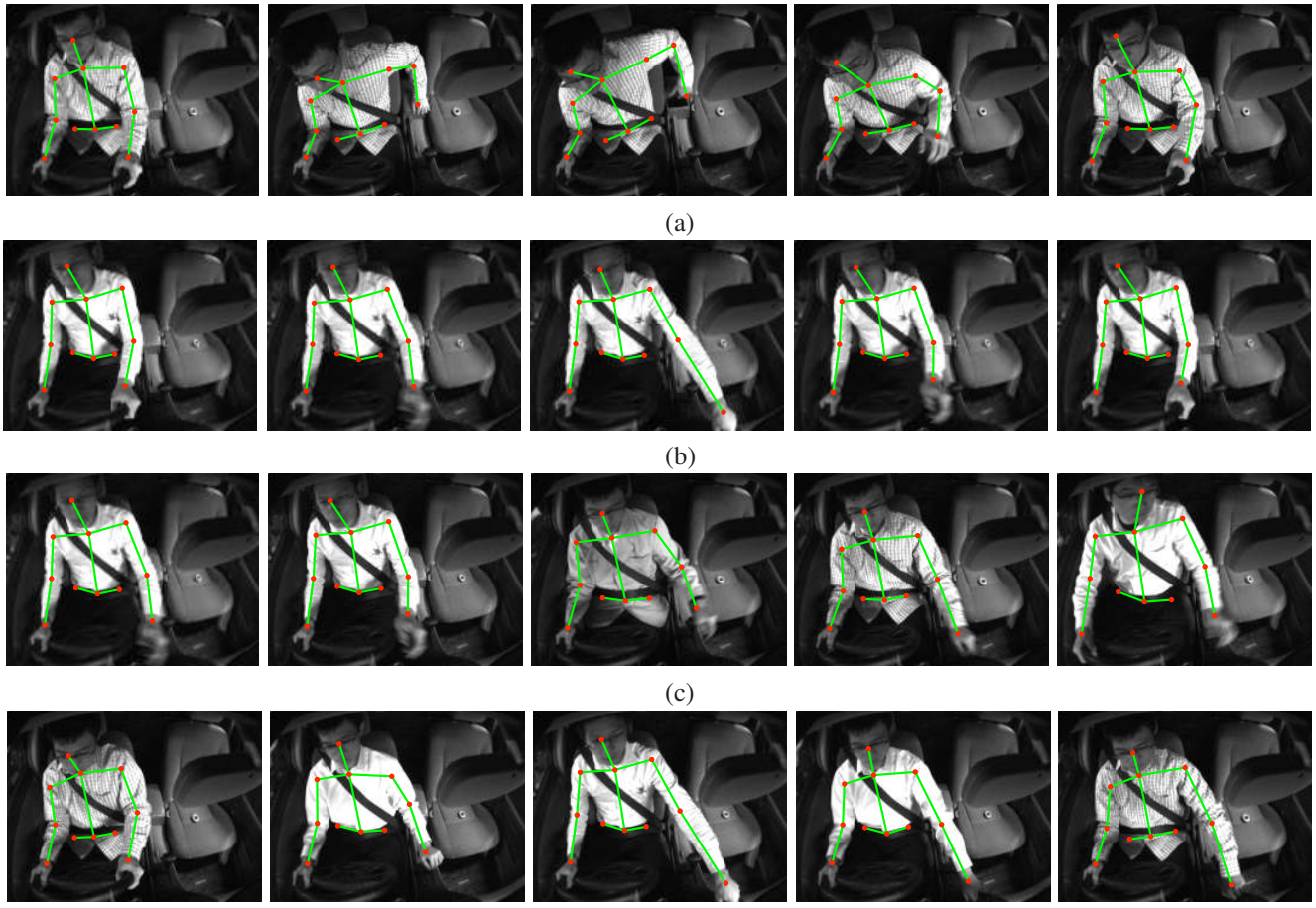
1. **Create Linear System**: Create a block diagonal matrix $\mathbf{\Gamma}$ and a vector $\mathcal{B}$ as in 16 for the system of planes.
2. **Apply Articulation Constraints**: Create the linear equality constraint matrix $\mathbf{\Theta}$ as in Equation 15.
3. **Solve Linearly Constrained Least Squares System**: Solve $\mathbf{\Gamma}\mathcal{A} = \mathcal{B}$ subject to $\mathbf{\Theta}\mathcal{A} = 0$ (See Appendix A).
4. **Update Source Image**: Warp the source image towards the target image.

Figure 5. Motion Estimation for Systems of Articulated Planes.

5.1. Human Body Tracking

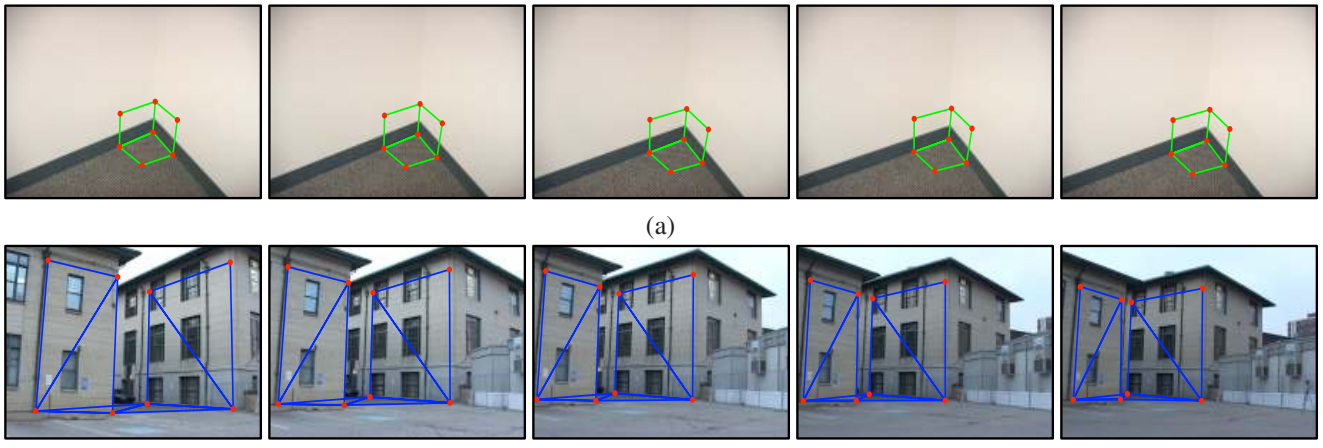
A human body can be modeled as a system of singly articulated planes, where each limb shares one articulation with an attached limb. We collected a large data set of 11,000 frames of 3 people wearing 5 different types of clothing, over a period of several imaging sessions. This data set has on the order of about 25 human activities, with each activity roughly 400 frames long at 30 frames per second.

We manually initialized eleven points on the upper body, of which five were articulations. Based on these points, a rectangular box around each limb is obtained and the pixels lying in each box are used to construct $\mathbf{\Lambda}_i$ and $\mathbf{b}_i$ for that plane. The articulations are used to set up the linear constraint matrix, $\mathbf{\Theta}$. Thus, a system of 36 unknowns with 10 constraint equations is constructed, which is solved using the algorithm outlined in Figure 5 at an average speed of 4 seconds per frame in MATLAB. We conducted several tests on a variety of activities such as reaching for the glove box, changing gears, and reaching into the center console. Several results are shown in Figure 6. An interesting point can be made about tracking through motion blur (Figure 6(c)). Since our tracking algorithm uses articulations, therefore, even though the information content locally around the blurred area is low, the tracker is able to incorporate information from the connected limbs to successfully track the blurred object. During experimentation the principal sources of failure were strong occlusions and the presence of strong background gradient during severe blurring.



(a) (b) (c) (d)

Figure 6. Human body tracking (a) Key frames of tracking a human performing a complete activity (reaching for the center console box). (b) Key frames of tracking a human reaching for the center instrument panel. (c) Key frames of successful tracking of blurred body parts. Since, we use articulation constraints, therefore even though the information content locally around the blurred area is low, we are able to use the information from other articulations. (d) Key frames of tracking a human performing miscellaneous activities.



(a) (b)

Figure 7. Tracking a rigid, piecewise planar scene with low texture layers. Note that it is challenging to track points on low texture walls and the ground plane without the use of articulation constraints.

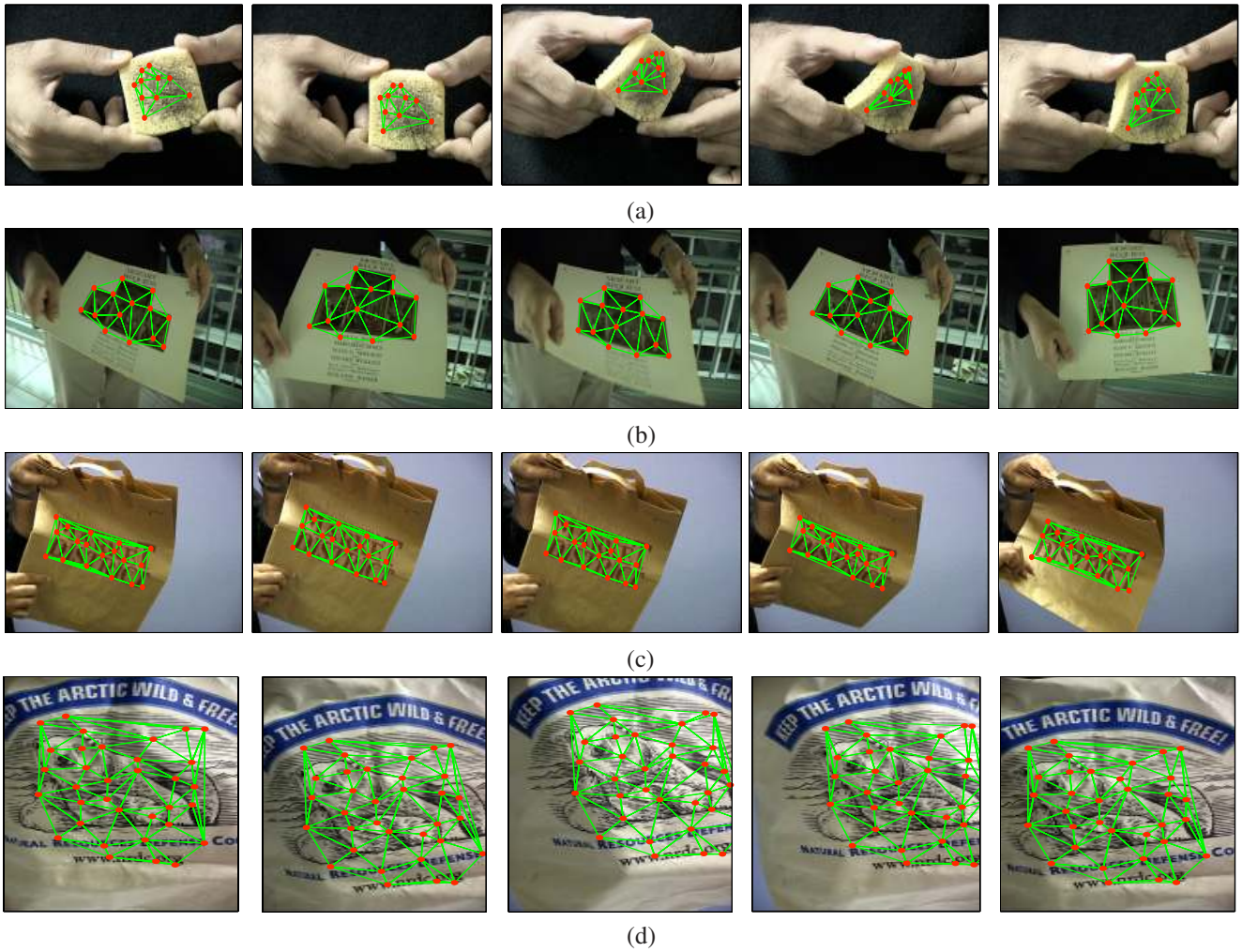


Figure 8. Result of tracking a triangulated mesh on a variety of nonrigid surfaces. (a) Snapshots of large illumination change resistant tracking of a sponge. (b) Tracking a paper being moved in a wave-like manner. (c) Tracking large deformations on a paper bag. (d) Robust tracking of a cloth bag, where the points on the right side of picture, disappear and then reappear in the field of view. Notice that when the points reappear, they are at their correct locations. Despite not having any gradient information, they are tracked correctly because of the articulation constraints from the neighboring points. We initialize the points (mesh vertices) in the first frame using the Harris corner detector and track the points in the consecutive frames.

5.2. Tracking Rigid Piecewise Planar Scenes

An important manifestation of doubly articulated planes occurs between the rigid faces of a building in urban scenes. As the camera moves, the motion of connected facades of a building are dependent on each other. Accurate motion estimation that ensures connectivity leads to application in 3D scene reconstruction and view synthesis of rigid scenes [12]. Figure 7 shows results of motion estimation in scenes containing multiple planes fixed with respect to each other (in 3D). It can be observed from the images that due to the articulation constraints, planes which have little or no texture can also be tracked. For example in Figure 7(a) two of the planar faces have unidirectional texture. Despite this, the articulation constraints allow the ground plane to anchor the motion of the other two planes. This ability is even more apparent in Figure 7(b), where the ground plane has barely any texture at all. This is a common phenomenon in real urban scenes, and articulations provide a solution for estimating ground plane motion robustly.

5.3. Motion Estimation of Triangulated Meshes

Consider the problem of estimating the motion of a nonrigid surface such as a piece of cloth or paper. The underlying motion of such a surface cannot be captured by a single, globally defined parametric motion model and hence must take on more sophisticated representations such as Thin Plate Splines (TPS) or triangulated meshes. The principal advantage of using triangulated meshes is that surface

discontinuities can be handled by triangulated meshes, but require additional mechanisms with TPS.

Given a mesh constructed, for example, out of Harris corner points or uniformly sampled points, we set up the linear system using the pixels contained within each triangle. The constraint system is setup so that mesh vertices are transferred to the same location by all the triangles sharing that point. This system is then solved using the algorithm outlined earlier in Figure 5. Figure 8 presents result on different nonrigid surfaces on which we applied our algorithm. Note that we do *not* require point correspondences to estimate motion.

Several interesting observation can be made about the results. We are able to robustly estimate the motion of the nonrigid surface through large illumination changes in part because the motion of the triangles which lie in saturated areas of the image is well-constrained by the other neighboring triangles through the articulation constraints. This is the same reason as to why we are able to accurately recover the motion of triangles even after part of the triangulated mesh has left the field of view. This is evident in several results, in particular the Cloth Bag sequence (Figure 8(d)) — note the accurate localization of the vertex on the last "E" of "DEFENSE". This happens because a large number of articulation constraints are placed by the triangulated mesh on each triangle and hence even if the triangles, or some parts of the triangles are not visible, the neighboring triangles can accurately constrain their positions.

The principal source of error in these experiments was the inability of the triangulated mesh to express the underlying motion of the surface. There is a tradeoff between the size of the triangles (which ensures that each triangle contains sufficient gradients) and the resolution of triangulations (which allows greater expression of nonrigid motion).

6. Related Work

In this paper, we describe the use of articulation constraints on direct algorithms and demonstrate their applicability in a variety of applications. The study of articulation has a long history in the field. Since the introduction of spring constraints in the seminal work of Fischler and Elschlager [8] in 1973, motion estimation algorithms have modeled articulation constraints in many different ways to capture the space of physically realizable set of motions [10, 22]. Nishihara and Marr [19] represented the body as a hierarchical collection of cylinders. Each component cylinder was connected to other cylinders using *adjunct relations*, which were predefined relations that specify the location of the component cylinder relative to the torso. Rourke *et al*. [20] introduced constraints on human body models such as distance constraints or joint angle limits to refine the 3D joint positions. Johansson in [13] and Lee *et al*. in [15] introduced and made popular the stick figure model for understanding and analyzing human body motion. This model was later extended by Ju [14], where each body limb was modeled by a planar patch and a set of constraints, two per limb, were introduced as representing a "smoothness" term. More recently, Bregler [6] modeled human body motion constraints as a product of exponential maps in a kinematic chain, where each articulation is modeled as a twist. Sigal [26] introduced conditional probabilistic modeling of limb articulations, where the limb articulation constraints are learnt from the motion capture training data.

Methods in the tradition of the Lucas-Kanade algorithm ([17]), also called "direct" algorithms [11], have been proposed for many different parametric motion models such as the affine transformation [2] and the homography [27]. In addition, several direct methods that utilize appearance information for estimating non-rigid deformation have been proposed in the literature [4, 7]. Bookstein in [5] introduced Thin Plate Spline (TPS) model for warping points between two frames to estimate nonrigid motion. This idea was further explored by others [16]. Sclaroff *et al*. [24] employed texture-mapped triangulated meshes, *active blobs*, for tracking deformable shapes in images. Active blobs, similar in spirit to the TPS model, solve an energy minimization problem with an application dependent regularization parameter to perform nonrigid tracking. One limitation of TPS model is their inability to handle surface discontinuities such as the ones encountered in Figure 7.

Our goal in this work is to consider articulation constraints, not as a form of soft regularization or "smoothness", but as linear, exact equality constraints that are placed on the 2D motion estimation task. We show that these translate into linear constraints that enable us to depart from estimating or hard-coding potentially nonlinear spring constraints and regularization weights. Since there are no application dependent parameters, our motion estimation framework allows us the flexibility to employ the algorithm for a variety of tasks without parameter tuning.

7. Conclusion

In this paper, we have presented a motion estimation algorithm that explicitly employs articulation constraints to recover a variety of real world motions. The algorithm constructs an over-constrained system of linear equations subject to linear, exact equality constraints to solve for the motion of multiple entities simultaneously. Since, we solve for the motion of all entities simultaneously, therefore the entire set of constraints bears on the motion parameters for all the entities. In some cases, this enables the algorithm to track parts of the object even if they have left the field of view and when there is little gradient information available for that plane.

The value of our algorithm lies in its ability to compute motion estimates for systems of articulated planes without

the use of any application dependent regularization parameters or smoothness terms. This points to broad applicability of the algorithm to a variety of real-world motion estimation tasks as demonstrated in this paper.

During experimentation, we noted two primary sources of error. The first source of error is occlusion. For cases such as the human body, this is an important consideration where self-occlusion is a fairly common phenomenon. The second type of error occurs in nonrigid surface tracking, when the resolution of the model is unable to represent the motion. This raises an important open question of what is an appropriate triangulation of a nonrigid surface and should the mesh be constructed out of feature detectors or uniformly or perhaps affected by the underlying motion of the nonrigid surface. Developing occlusion handling mechanisms and resolving the question of triangulation coverage and resolution will be the focus of future research.

Acknowledgements

The research described in this paper was supported by the DENSO Corporation, Japan.

References

[1] V. G. Bellile, M. Perriollat, A. Bartoli, and P. Sayd. Image registration by combining thin-plate splines with a 3D morphable model. In *ICIP*, 2006.

[2] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *Second ECCV*, 1992.

[3] M. Black and A. Jepson. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. In *IJCV*, 1998.

[4] M. J. Black and Y. Yacoob. Tracking and recognizing rigid and nonrigid facial motions using local parametric models of image motion. In *ICCV*, 1995.

[5] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *PAMI*, 11(6), 1989.

[6] C. Bregler and J. Malik. Tracking people with twists and exponential maps. In *CVPR*, 1998.

[7] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *ECCV*, 1998.

[8] M. A. Fischler and R. A. Elschlager. The representation and matching of pictorial structures. *Transactions on Computer*, 22(1), 1973.

[9] P. Gill, W. Murray, and M. Wright. Practical optimization. In *Academic Press*, 1981.

[10] D. Hogg. Model-based vision: a program to see a walking person. *Image and Vision Computing*, 1(1), 1983.

[11] B. Horn and E. Weldon. Direct methods for recovering motion. In *IJCV*, 1988.

[12] B. Johansson. View synthesis and 3D reconstruction of piecewise planar scenes using intersection lines between the planes. In *ICCV*, 1999.

[13] G. Johansson. Visual motion perception. *Scientific American*, 232, 1976.

[14] S. Ju, M. Black, and Y. Yacoob. Cardboard people: A parameterized model of articulated image motion. In *FGR*, 1996.

[15] H. J. Lee and Z. Chen. Determination of 3D human body postures from a single view. *CVGIP*, 30, 1985.

[16] J. Lim and M. H. Yang. A direct method for modeling non-rigid motion with thin plate spline. In *CVPR*, 2005.

[17] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Image Understanding Workshop*, 1981.

[18] L. Van Gool, L. Proesmans, and A. Zisserman. Grouping and invariants using planar homologies. In *Workshop on Geometric Modeling and Invariants for Computer Vision*, 1995.

[19] H. K. Nishihara and D. Marr. Representation and recognition of the spatial organization of three-dimensional shapes. In *MIT AI Memo*, 1976.

[20] J. O'Rourke and N. I. Badler. Model-based image analysis of human motion using constraint propagation. *PAMI*, 2(6), 1980.

[21] P. Pritchett and A. Zisserman. Matching and reconstruction from widely separated views. In *3D Structure from Multiple Images of Large-Scale Environments*, 1998.

[22] K. Rohr. Towards model-based recognition of human movements in image sequences. *CVGIP: Image Understanding*, 59(1), 1994.

[23] H. S. Sawhney and S. Ayer. Compact representations of videos through dominant and multiple motion estimation. *PAMI*, 18(8), 1996.

[24] S. Sclaroff and J. Isidoro. Active blobs. In *ICCV*, 1998.

[25] J. Semple and G. Kneebone. Algebraic projective geometry. *Oxford University Press*, 1952.

[26] L. Sigal, S. Bhatia, S. Roth, M. J. Black, and M. Isard. Tracking loose-limbed people. In *CVPR*, 2004.

[27] R. Szeliski. Image mosaicing for tele-reality applications. In *WACV*, 1994.

[28] J. Y. A. Wang and E. H. Adelson. Representing moving images with layer. *Transactions on Image Processing*, 3(5), 1994.

[29] Y. Weiss. Smoothness in layers: Motion segmentation using nonparametric mixture estimation. In *CVPR*, 1997.

[30] L. Zelnik-Manor and M. Irani. Multiview constraints on homographies. In *PAMI*, 2002.

Appendix A. Least Squares with Linear Equality Constraints

We wish to solve,

$$\min_{\mathcal{A}} \|\mathcal{B} - \mathbf{\Gamma}\mathcal{A}\|_2 \text{ subject to } \mathbf{\Theta}\mathcal{A} = 0, \tag{19}$$

where $\mathbf{\Gamma}$ is an $M \times N$ matrix, $\mathcal{B}$ is a M-vector, $\mathbf{\Theta}$ is a $C \times N$ matrix and $C \leq N \leq M$. Using Lagrange Multipliers,

$$f(\mathcal{A}|\lambda) = \|\mathcal{B} - \mathbf{\Gamma}\mathcal{A}\|_2^2 + 2\lambda^T \mathbf{\Theta}\mathcal{A}. \tag{20}$$

The gradient of $f(\mathcal{A}|\lambda)$ equals zero when,

$$\mathbf{\Gamma}^T\mathbf{\Gamma}\mathcal{A} + \mathbf{\Theta}^T\lambda = \mathbf{\Gamma}^T\mathcal{B}, \tag{21}$$

and

$$\mathbf{\Theta}\mathcal{A} = 0. \tag{22}$$

This can be written and solved as a Karush-Kuhn-Tucker system,

$$\begin{bmatrix} \mathbf{\Gamma}^T\mathbf{\Gamma} & \mathbf{\Theta}^T \\ \mathbf{\Theta} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathcal{A} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{\Gamma}^T\mathcal{B} \\ \mathbf{0} \end{bmatrix}. \tag{23}$$