Linearized Motion Estimation for Articulated Planes

Ankur Datta, Yaser Sheikh, and Takeo Kanade

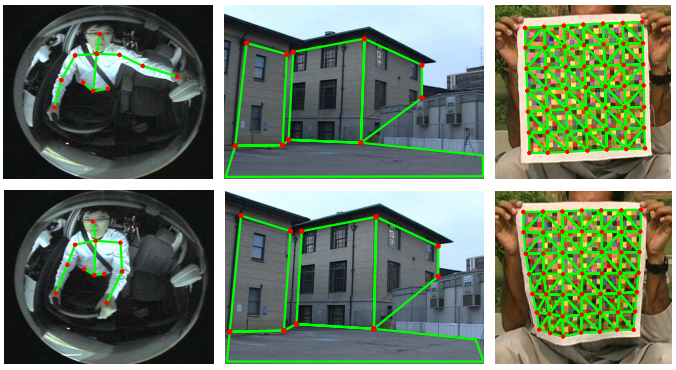
Abstract—In this paper, we describe the explicit application of articulation constraints for estimating the motion of a system of articulated planes. We relate articulations to the relative homography between planes and show that these articulations translate into linearized equality constraints on a linear least squares system, which can be solved efficiently using a Karush-Kuhn-Tucker system. The articulation constraints can be applied for both gradient-based and feature-based motion estimation algorithms and to illustrate we describe a gradient-based motion estimation algorithm for an affine camera and a feature-based motion estimation algorithm for a projective camera that explicitly enforce articulation constraints. We show that explicit application of articulation constraints leads to numerically stable estimates of motion. The simultaneous computation of motion estimates for all the articulated planes in a scene allows us to handle scene areas where there is limited texture information and areas that leave the field of view. Our results demonstrate the wide applicability of the algorithm in a variety of challenging real world cases such as human body tracking, motion estimation of rigid, piecewise planar scenes and motion estimation of triangulated meshes.

Index Terms—I.4.3.d Registration, I.4.8.d Motion, I.4.8.n Tracking

1 Introduction

The principal challenge in developing general purpose motion estimation algorithms is the wide-variety of rigid and nonrigid motions encountered in the real world. Consider the three examples shown in Figure 1. In Figure 1(a), the motion of a human is shown where each limb's motion is dependent on the motion of its connected limbs. Motion of a rigid scene, shown in Figure 1(b), 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 Figure 1(c) depends on the elasticity of the object and the force acting on it. The problem of motion estimation for varied objects such as these has resulted in proposition of a large number of algorithms, for instance [1]–[9]. In particular, due to their wide applicability, layered motion models have gained significant traction over the years [10]–[12]. However, existing layers based motion algorithms do not exploit a key constraint that exists in the motion of a large number of real scenes.

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 of points on the line of intersection of any two planes is the same for the two planes (Figure 1(b)). 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 and can therefore be considered an articulation (Figure 1(c)).



(a) (b) (c)

Fig. 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.

In this paper, we study the relationship between articulations and the homographies induced by articulated planes (Section 3). Unlike previous constraints [3], [9], we define *exact equality* constraints on the motion model of the articulated planes for an affine camera

• *A.Datta, Y. Sheikh, and T. Kanade are with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213.*

• *E-mail: {ankurd,yaser,tk}@cs.cmu.edu*

(Section 4.1). Articulation constraints can be used for motion estimation in both gradient-based and feature-based motion estimation algorithms. As an illustration, for an affine camera, we describe the application of articulation constraints in a gradient-based motion estimation algorithm that solves a linear equality constrained least squares system for estimating motion of multiple planes simultaneously (Section 4.2). For a projective camera, we describe the application of linearized articulation constraints on the motion model (Section 5.1). For feature-based motion estimation algorithms, we show the application of articulation constraints for a projective camera that minimizes linearized least squares transfer error between two images (Section 5.2). Our results demonstrate that the incorporation of explicit articulation constraints in motion estimation algorithms results in accurate estimation of motion in a wide-variety of settings such as human body tracking, estimating motion of rigid, piecewise planar scenes and estimating motion of triangulated meshes (Section 6).

2 Related Work

Motion estimation from image sequences is a core computer vision task. The earliest motion estimation algorithms were pixel-based approaches that used first order image derivatives and were based on image translation [1], [13]–[16]. Pixel-based methods are descriptive, however, they can be unstable due to the corruption of image intensity from real-world effects such as sensor noise during the image acquisition process, aliasing and the "aperture problem". Region-based methods, on the other hand, provide an alternative motion estimation approach to mitigate these effects [17]–[20]. Motion estimation, according to region-based approaches, has been defined as yielding the best fit between image regions at different times [21]. The match score between the regions at different times can be evaluated according to similarity measures such as the normalized correlation or Sum of Squared Differences (SSD). Consequently, region-based approaches are stable but less descriptive than pixel-based approaches since they operate at region granularity and not pixel granularity. The interested reader is directed to excellent surveys on gradient-based optical flow approaches by Fleet *et al.* [22] and also to a survey by Beauchemin *et al.* [21] on other optical flow approaches.

The need to balance model descriptiveness and parameter estimation stability led to the development of the piecewise planar framework for motion estimation by Wang and Adelson in [10]. The layer framework models an image as a collection of independently moving planes that compete for the ownership of image pixels. The layers are ordered by depth and each has an intensity and an alpha map that is used in compositing to explain the underlying image. Sawhney and Ayers [23] introduced a maximum likelihood estimation algorithm for estimating the parameters of the motion layers and their ownership probabilities. The layer framework has been applied successfully to a wide variety of real-world applications [24]–[26].

The traditional layer framework treats each plane in the scene as moving independently. Several methods have been used to enforce the dependencies that exist between moving planes in rigid scenes such as [12], [27]. For articulate object motion, such as human motion, [9], [28] defined articulations to capture relationships between the motion of different planes. Ju *et al.* [9] introduced "Cardboard People" for modeling the human body as a set of connected planar patches or layers. Motion was estimated for all layers simultaneously using gradient-based motion estimation where articulations provide soft constraints on the motion. The motion of the connected layers is enforced to be spatially constrained and a regularization parameter is used to weigh the articulation constraints relative to gradient-based motion estimation. The use of the regularization parameter, however, introduces arbitrariness in the motion estimation algorithm since there is no one consistent value that will work across all applications. Ju *et al.* in [28] extended the cardboard people model to articulated motion estimation for multi-layer framework. The articulated layers models has subsequently been used in a number of papers [29]–[33] and articulation constraints, including constraints on angular velocity and acceleration, have been used for 3D model estimation as well [34]–[40]. Bregler *et al.* in [34], [35] demonstrate recovery of 3D articulated motion using twists and exponential maps. In their formulation, twists are modeled as revolute joints anchored at articulations, that are then propagated to the next time-step under the assumption of isotropic Gaussian noise. Ruf *et al.* in [41] introduced projective formulations for revolute and prismatic joints in the context of an articulate chain with a fixed *base*. Articulated motion for the i^{th} joint is estimated in terms of all the $(i-1)$ joints using twists and exponential maps. Sigal *et al.* in [3], [42] used conditional probability distributions, which can be interpreted as soft articulation constraints, to model the relationships between body joints. Demirdjian *et al.* in [43] imposed kinematic constraints using a linear manifold estimated from the previous body pose. Nonlinear kinematic constraints were then enforced using a learning-based approach via a support vector classifier.

The articulated layers models described above are not descriptive enough to handle nonrigid surface motion. This led to the introduction of specialized models for nonrigid surface motion such as the Thin Plate Spline (TPS) model [44] and triangulated meshes [5], [45]. The principal advantage of using triangulated meshes is that sharp surface creases can be handled by triangulated meshes, but require additional mechanisms with TPS. Sclaroff *et al.* [46] 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. Bartoli *et al.* introduced a direct motion estimation algorithm employing Radial Basis Functions to model nonrigid image deformations [47]. A subsequent feature-driven nonrigid registration method was developed by Gay-Bellile *et al.* [48]. The idea of pairwise nonrigid registration of images was extended in a paper by Cootes *et al.* in [49], where they developed a framework for registering a group of images together using a set of nonlinear diffeomorphic warps employing a regularizing parameter to penalize convoluted deformations. Over the years, many physics-based methods have been introduced for recovering shape and/or tracking of nonrigid surfaces, such as [50]–[55]. A key aspect of physics-based modeling requires making assumptions about the underlying nonlinear-physics of nonrigid surface deformations; these assumptions are hard to justify and are seldom accurate in practice. This has led to the introduction of data-driven priors for modeling nonrigid surface deformations. Recent work in this area includes [56]–[58]. However, a key shortcoming of data-driven approaches is that they require large training data sets which may not be available for nonrigid surfaces.

In contrast with the previous work on articulated and nonrigid motion estimation, which imposes articulation constraints as a soft regularizer or as a "smoothness" term, we introduce the idea of imposing the articulation constraints as exact equality constraints on 2D motion estimation. The proposed constraints facilitate models of motion that are both descriptive and numerically stable to estimate. The articulation constraints can be applied to both gradient-based and feature-based motion estimation algorithms and as an example we have incorporated them in a gradient-based motion estimation algorithm for an affine camera and in a feature-based motion estimation algorithm for a projective camera. We show results on several real world tasks of estimating motion of humans, rigid planar scenes and nonrigid surfaces. In addition, for the nonrigid surfaces, we do not require any physics-based prior i.e. assumptions about the mechanical principles governing the motion of nonrigid surfaces [52] or any data-prior for nonrigid surface registration [58]. The estimation algorithms proposed in this paper have been demonstrated to run upwards of 90 Hz on a Commercial Off-The Shelf (COTS) machine.

3 Articulated Planes

Between a pair of planes $(\mathbf{\Pi}_i, \mathbf{\Pi}_j)$ undergoing Euclidean transformations $(\mathbf{T}_i, \mathbf{T}_j)$ respectively, an *articulation* $\mathbf{P}$, is a point that moves identically under the action of both $\mathbf{T}_i$ and $\mathbf{T}_j$ in $\mathbb{R}^3$. There can be at most two such points between planes since if there are three noncollinear articulations the two moving planes are, in fact, the same plane. Note that this does not exclude collinear points that lie on the line connecting the two articulation points to serve as articulation constraints. Singly articulated planar systems, or planes that share a single articulation, are a popular model of the human body [9], [59] (see Figure 2) and what can be considered doubly articulated planar systems, or planes that share two articulations, have found application in shadow analysis, view synthesis and in scene reconstruction, [60]–[62].

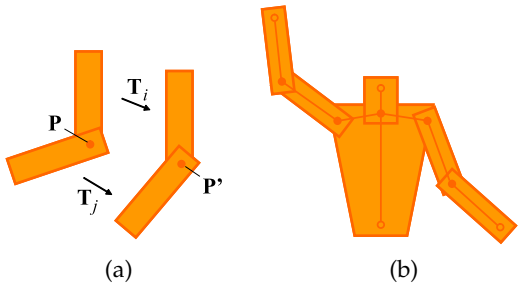


Fig. 2: Articulation Constraints: (a) Articulations move identically under the transformations of two planes (b) Singly articulated planes as a model for body tracking. Five points connect six body parts.

Under the action of a projective camera, the motion field induced by a moving plane can be described by a homography,

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

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

$$\mathbf{p}^{'} \cong \mathbf{H}_i\mathbf{p} \cong \mathbf{H}_j\mathbf{p}, \tag{2}$$

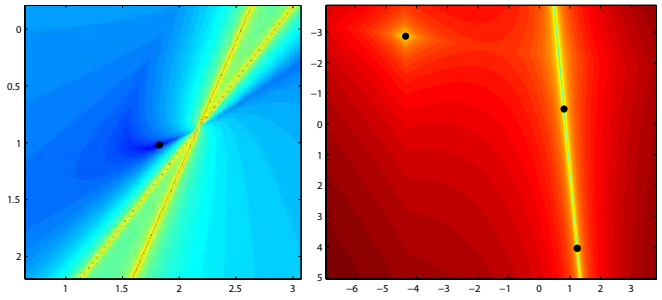
where $\mathbf{p}^{'}$ is the image of the articulation point $\mathbf{P}$ in the second image. 2D articulations can be computed directly from the pair of homographies by noting that they are related to the fixed or united points [63] of the relative homography $\mathbf{\Omega}_{ij} = \mathbf{H}_i^{-1}\mathbf{H}_j$. The 2D articulations, $\mathbf{p}_k$, $k \in \{1, 2, 3\}$, correspond to eigenvectors of $\mathbf{\Omega}_{ij}$ (and $\mathbf{\Omega}_{ji}$). This can be seen from

$$s_i^k\mathbf{H}_i\mathbf{p}_k = \mathbf{p}'_k, \quad s_j^k\mathbf{H}_j\mathbf{p}_k = \mathbf{p}'_k. \tag{3}$$

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

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

where $\lambda_k = \frac{s_j^k}{s_i^k}$ and $\mathbf{I}$ is a 3×3 identity matrix. Thus, given $(\mathbf{H}_i, \mathbf{H}_j)$, 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



(a) (b)

Fig. 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.

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 [60]. 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)).

4 Articulated Motion Model for Affine Cameras

Articulation constraints are constraints that are placed on the articulations of two or more planes that share the articulated points. In this section, we make explicit the constraints placed by the articulated points for an affine camera. As an illustration of the use of articulation constraints in gradient-based or "direct-methods", we describe an algorithm for articulated motion estimation viewed by affine cameras.

4.1 Articulation Constraints

The motion induced between two views of a plane for an affine camera is represented by an affine transformation,

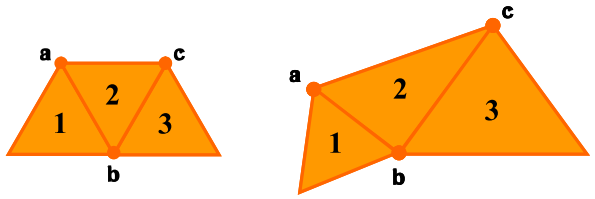
$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \tag{5}$$

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,

$$\mathbf{A}_i\mathbf{p} = \mathbf{p}' = \mathbf{A}_j\mathbf{p}. \tag{6}$$

Equation 6 can be rewritten as,

$$(\mathbf{A}_i - \mathbf{A}_j)\mathbf{p} = \mathbf{0}, \tag{7}$$



(a) (b)

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

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

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 6,

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

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

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

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]: $\mathbf{a}_{12}$, $\mathbf{b}_{12}$, $\mathbf{c}_{23}$, $\mathbf{b}_{23}$ and $\mathbf{b}_{13}$. However, there are only four linearly independent constraints since the constraint produced by $\mathbf{b}_{13}$ is linearly dependent on those of $\mathbf{b}_{12}$ and $\mathbf{b}_{23}$.

4.2 Articulated Motion Estimation

In this section, we describe how to use articulation constraints in the affine parameter estimation algorithm proposed by Bergen *et al.* [2]. Under an affine camera assumption, the imaged motion of planes is described by an affine transform. By making the brightness constancy assumption between corresponding pixels in consecutive frames, the motion estimation process involves SSD minimization,

$$E(\mathbf{a}) = \sum_{\mathbf{x}} \Big(I_t(\mathbf{x}) - I_{t+1}\big(W(\mathbf{x}|\mathbf{a})\big) \Big)^2, \tag{8}$$

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

where W is a warp function, $\mathbf{a} = [a_1, \cdots a_6]^\top$ 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,

$$\nabla I_x u + \nabla I_y v + \Delta I_t = 0, \tag{9}$$

where ∇I_x, ∇I_y and ΔI_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,

$$x' = a_1 x + a_2 y + a_3, \tag{10}$$
$$y' = a_4 x + a_5 y + a_6, \tag{11}$$

or in matrix form as,

$$\mathbf{x}' = \mathbf{X}\mathbf{a}, \tag{12}$$

where

$$\mathbf{X} = \begin{bmatrix} x & y & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x & y & 1 \end{bmatrix}. \tag{13}$$

Equations 9 and 12 can be combined to create a linear system of equations in the unknown motion parameters $\mathbf{a}$. Thus, in a system of planes, for the i^{th} plane we have,

$$\mathbf{\Lambda}_i(\nabla I_x, \nabla I_y)\mathbf{a}_i = \mathbf{b}_i(\nabla I_x, \nabla I_y, \Delta I_t), \tag{14}$$

where $\mathbf{\Lambda}_i = \sum \mathbf{X}^\top(\nabla\mathbf{I})(\nabla\mathbf{I})^\top\mathbf{X}$, $\mathbf{b}_i = -\sum \mathbf{X}^\top(\nabla\mathbf{I})(\Delta I_t)$. 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,

$$\begin{bmatrix} \mathbf{\Lambda}_i(\nabla\mathbf{I}) & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_j(\nabla\mathbf{I}) \end{bmatrix} \begin{bmatrix} \mathbf{a}_i \\ \mathbf{a}_j \end{bmatrix} = \begin{bmatrix} \mathbf{b}_i(\nabla\mathbf{I}, \Delta I_t) \\ \mathbf{b}_j(\nabla\mathbf{I}, \Delta I_t) \end{bmatrix}. \tag{15}$$

Solving the system in Equation 15 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 7. In terms of $[\mathbf{a}_i \;\; \mathbf{a}_j]^\top$ this constraint can be written as,

$$\begin{bmatrix} \mathbf{p}^\top & \mathbf{0} & -\mathbf{p}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{p}^\top & \mathbf{0} & -\mathbf{p}^\top \end{bmatrix} \begin{bmatrix} \mathbf{a}_i \\ \mathbf{a}_j \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{16}$$

or simply $[\theta(\mathbf{p}) \;\; \theta(-\mathbf{p})][\mathbf{a}_i^\top \;\; \mathbf{a}_j^\top]^\top = \mathbf{0}$. Estimating $[\mathbf{a}_i^\top \;\; \mathbf{a}_j^\top]^\top$ from Equations 15 and 16 is a standard equality constrained linear least squares problem which can be solved stably using the Karush-Kuhn-Tucker system as described below or by standard optimization packages (such as `lsqlin` in MATLAB®). For further details on such optimization the interested reader is directed to [64].

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 motion 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 15 would be expanded into a block diagonal matrix with n blocks,

$$\begin{bmatrix} \mathbf{\Lambda}_1 & & \\ & \ddots & \\ & & \mathbf{\Lambda}_n \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_n \end{bmatrix}, \tag{17}$$

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

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

$$\begin{bmatrix} \theta(\mathbf{a}) & \theta(-\mathbf{a}) & \mathbf{0} \\ \theta(\mathbf{b}) & \theta(-\mathbf{b}) & \mathbf{0} \\ \mathbf{0} & \theta(\mathbf{b}) & \theta(-\mathbf{b}) \\ \mathbf{0} & \theta(\mathbf{c}) & \theta(-\mathbf{c}) \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix} = \mathbf{0}. \tag{18}$$

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

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

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

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

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

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

From commutativity, it should be noted that the motion of $\mathbf{a}$ in Figure 4 is not independent of the motion of $\mathbf{c}$ 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.

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 Equation 17 for the system of planes.
2) **Apply Articulation Constraints**: Create the linear equality constraint matrix $\mathbf{\Theta}$ as in Equation 18.
3) **Solve Linearly Constrained Least Squares System**: Solve $\mathbf{\Gamma}\mathcal{A} = \mathcal{B}$ subject to $\mathbf{\Theta}\mathcal{A} = 0$.
4) **Update Source Image**: Warp the source image towards the target image.

Fig. 5: Motion estimation for a system of articulated planes under an affine camera.

5 Articulated Motion Model for Projective Cameras

In this section, we make explicit the constraints placed by the articulated points for a projective camera. As an illustration of the use of articulation constraints in feature-based methods, we describe an algorithm for articulated motion estimation viewed by projective cameras.

5.1 Articulation Constraints

For projective cameras, the motion induced between two views of a plane is represented by a homography. Let $\mathbf{H}_i$ be the homography that maps the plane $\mathbf{\Pi}_i$ from the first image to the second image of the same plane and similarly let $\mathbf{H}_j$ be the homography for plane $\mathbf{\Pi}_j$, where $\mathbf{H} = [\mathbf{h}_1^T, \mathbf{h}_2^T, \mathbf{h}_3^T]^T$ and $\mathbf{h}_i^T$ represents the i^{th}-row of the homography matrix $\mathbf{H}$. Between plane $\mathbf{\Pi}_i$ and plane $\mathbf{\Pi}_j$ articulated at homogeneous image point $\mathbf{p} \in \mathbb{R}^3$, the articulation constraint takes the form,

$$\mathbf{H}_i\mathbf{p} \cong \mathbf{p}' \cong \mathbf{H}_j\mathbf{p}. \tag{24}$$

Given a point in one image and its corresponding point in the other image, the transfer error [65] is defined as the squared Euclidean distance between the projection of the point using homography and its corresponding point. We define a variant of the transfer error to impose articulation constraints for homography estimation. This constraint measures the difference in the projection of an articulation point using the homographies of the planes which share that point,

$$\phi(\mathbf{H}_i, \mathbf{H}_j; \mathbf{p}) = \mathbf{0}, \tag{25}$$

where

$$\phi(\mathbf{H}_i, \mathbf{H}_j; \mathbf{p}) = \begin{bmatrix} \left(\frac{\mathbf{h}_{i1}^T\mathbf{p}}{\mathbf{h}_{i3}^T\mathbf{p}} - \frac{\mathbf{h}_{j1}^T\mathbf{p}}{\mathbf{h}_{j3}^T\mathbf{p}}\right) \\ \left(\frac{\mathbf{h}_{i2}^T\mathbf{p}}{\mathbf{h}_{i3}^T\mathbf{p}} - \frac{\mathbf{h}_{j2}^T\mathbf{p}}{\mathbf{h}_{j3}^T\mathbf{p}}\right) \end{bmatrix}. \tag{26}$$

A single articulation point $\mathbf{p}$, therefore, results in a vector with 2 rows, $\phi(\mathbf{H}_i, \mathbf{H}_j; \mathbf{p})$, imposing constraints on the articulated motion. Rigidly connected planar systems, such as in Figure 1(b), share two articulated points that define the line of intersection. In such a case, a vector with 4 rows of articulation constrains can be formed,

$$\Phi(\mathbf{H}_i, \mathbf{H}_j; \mathbf{p}_1, \mathbf{p}_2) = \mathbf{0}, \tag{27}$$

where

$$\Phi(\mathbf{H}_i, \mathbf{H}_j; \mathbf{p}_1, \mathbf{p}_2) = \begin{bmatrix} \phi(\mathbf{H}_i, \mathbf{H}_j; \mathbf{p}_1)^2 \\ \phi(\mathbf{H}_i, \mathbf{H}_j; \mathbf{p}_2)^2 \end{bmatrix}, \tag{28}$$

where $\mathbf{p}_1$ and $\mathbf{p}_2$ are the two articulated points and $\phi(\mathbf{H}_i, \mathbf{H}_j; \mathbf{p})^2$ denotes element-wise square operation on the vector.

5.2 Articulated Motion Estimation

We now describe how to use the articulation constraints for motion estimation for a projective camera. The articulation constraints can be applied to both direct and feature-based algorithms, and as an example, we present a feature-based articulated motion estimation algorithm in this section.

Consider a doubly articulated planar system consisting of two planes, $\mathbf{\Pi}_i$ and $\mathbf{\Pi}_j$, with their respective homographies, $\mathbf{H}_i$ and $\mathbf{H}_j$, that map the images of the planes from the first image to the second image. Given the feature correspondences for each plane between the two views, we can use the transfer error [65] to measure the accuracy of point transfer from one image to the other using homography,

$$\Psi(\mathbf{H}_i, \mathbf{H}_j; \{\mathbf{x}_i, \mathbf{x}_i^{'}\}, \{\mathbf{x}_j, \mathbf{x}_j^{'}\}) = \begin{bmatrix} \psi(\mathbf{H}_i; \{\mathbf{x}_i, \mathbf{x}_i^{'}\})^2 \\ \psi(\mathbf{H}_j; \{\mathbf{x}_j, \mathbf{x}_j^{'}\})^2 \end{bmatrix}, \tag{29}$$

where

$$\psi(\mathbf{H}; \{\mathbf{x}, \mathbf{x}^{'}\}) = \begin{bmatrix} \left(f(\mathbf{H}; \mathbf{x}^1) - \mathbf{x}^{1'}\right) \\ \vdots \\ \left(f(\mathbf{H}; \mathbf{x}^n) - \mathbf{x}^{n'}\right) \end{bmatrix}, \tag{30}$$

where $(f(\mathbf{H}; \mathbf{x}^j) - \mathbf{x}^{j'})$ is a vector with 2 rows, $f(\mathbf{H}; \mathbf{x}^j)$ is transfer of the j^{th} point, $\mathbf{x}^j$, from the first image to the second image using homography $\mathbf{H}$ and $\mathbf{x}^{j'}$ denotes its corresponding feature match in the second image. $\{\mathbf{x}_i, \mathbf{x}_i^{'}\}$ is the set of homogeneous feature-point matches in the first and the second image respectively for plane $\mathbf{\Pi}_i$ and similarly $\{\mathbf{x}_j, \mathbf{x}_j^{'}\}$ is the set of homogeneous feature-point matches for plane $\mathbf{\Pi}_j$.

Equation 29 is a nonlinear equation in the motion parameters of the two planes, $\mathbf{H}_i$ and $\mathbf{H}_j$, and can be minimized using the Gauss-Newton gradient descent algorithm. Taking the Taylor series expansion, we get,

$$\Psi(\mathbf{H}_i, \mathbf{H}_j; \{\mathbf{x}_i, \mathbf{x}_i^{'}\}, \{\mathbf{x}_j, \mathbf{x}_j^{'}\}) \approx \left(\begin{bmatrix} \psi(\mathbf{H}_{i0}; \{\mathbf{x}_i, \mathbf{x}_i^{'}\}) \\ \psi(\mathbf{H}_{j0}; \{\mathbf{x}_j, \mathbf{x}_j^{'}\}) \end{bmatrix} + \begin{bmatrix} J_{f\mathbf{H}_{i0}} & \mathbf{0} \\ \mathbf{0} & J_{f\mathbf{H}_{j0}} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{H}_i \\ \Delta\mathbf{H}_j \end{bmatrix} \right)^2, \tag{31}$$

where $\mathbf{H}_{i0}$ and $\mathbf{H}_{j0}$ are initial homography estimates for planes $\mathbf{\Pi}_i$ and $\mathbf{\Pi}_j$ that can be obtained from the Direct Linear Transform algorithm [65], $J_{f\mathbf{H}_{i0}}$ and $J_{f\mathbf{H}_{j0}}$ are the corresponding Jacobians for the transfer error function f and $\Delta\mathbf{H}_i$, $\Delta\mathbf{H}_j$ are the respective motion parameter updates.

Taking the derivative with respect to $\Delta\mathbf{H}_i$ and $\Delta\mathbf{H}_j$, and equating to zero, we get,

$$Hess_{f\mathbf{H}_{i0}\mathbf{H}_{j0}} \begin{bmatrix} \Delta\mathbf{H}_i \\ \Delta\mathbf{H}_j \end{bmatrix} = -J_{f\mathbf{H}_{i0},\mathbf{H}_{j0}}^{\top} \begin{bmatrix} \psi(\mathbf{H}_{i0}; \{\mathbf{x}_i, \mathbf{x}_i^{'}\}) \\ \psi(\mathbf{H}_{j0}; \{\mathbf{x}_j, \mathbf{x}_j^{'}\}) \end{bmatrix}, \tag{32}$$

where $J_{f\mathbf{H}_{i0}\mathbf{H}_{j0}}$ is the combined Jacobian and $Hess_{f\mathbf{H}_{i0}\mathbf{H}_{j0}}$ is the combined Hessian,

$$J_{f\mathbf{H}_{i0}\mathbf{H}_{j0}} = \begin{bmatrix} J_{f\mathbf{H}_{i0}} & \mathbf{0} \\ \mathbf{0} & J_{f\mathbf{H}_{j0}} \end{bmatrix}, \tag{33}$$

$$Hess_{f\mathbf{H}_{i0}\mathbf{H}_{j0}} = \begin{bmatrix} Hess_{f\mathbf{H}_{i0}} & \mathbf{0} \\ \mathbf{0} & Hess_{f\mathbf{H}_{j0}} \end{bmatrix}. \quad (34)$$

The Jacobian $J_{f\mathbf{H}}$ of the transfer error function f is described in Appendix A and $Hess_{f\mathbf{H}}$ is the Gauss-Newton approximation to the Hessian,

$$Hess_{f\mathbf{H}} = J_{f\mathbf{H}}^{\top} J_{f\mathbf{H}}. \quad (35)$$

Equation 32 can be rewritten in matrix form as $\Gamma_{\mathbf{H}}\mathcal{H} = \mathcal{B}_{\mathbf{H}}$.

Doubly articulated planar system consist of two rigidly connected planes that share two points of articulations; these points serve as articulation constraints on the motion of the planes. Equation 27, which is a nonlinear articulated motion constraint, can be linearized using Taylor series and used to estimate articulated motion,

$$\Phi(\mathbf{H}_i, \mathbf{H}_j; \mathbf{p}_1, \mathbf{p}_2) \approx \left(\begin{bmatrix} \phi(\mathbf{H}_{i0}, \mathbf{H}_{j0}; \mathbf{p}_1) \\ \phi(\mathbf{H}_{i0}, \mathbf{H}_{j0}; \mathbf{p}_2) \end{bmatrix} + J_{\Phi}(\mathbf{H}_{i0}, \mathbf{H}_{j0}; \mathbf{p}_1, \mathbf{p}_2) \begin{bmatrix} \Delta\mathbf{H}_i \\ \Delta\mathbf{H}_j \end{bmatrix} \right)^2, \quad (36)$$

where $J_{\Phi}(\mathbf{H}_{i0}, \mathbf{H}_{j0}; \mathbf{p}_1, \mathbf{p}_2)$ is the Jacobian of Φ and is described in Appendix B. Since the articulation constraint is of the form, $\Phi(\mathbf{H}_i, \mathbf{H}_j; \mathbf{p}_1, \mathbf{p}_2) = \mathbf{0}$, we get,

$$J_{\Phi}^{\top} J_{\Phi} \begin{bmatrix} \Delta\mathbf{H}_i \\ \Delta\mathbf{H}_j \end{bmatrix} = -J_{\Phi}^{\top} \begin{bmatrix} \phi(\mathbf{H}_{i0}, \mathbf{H}_{j0}; \mathbf{p}_1) \\ \phi(\mathbf{H}_{i0}, \mathbf{H}_{j0}; \mathbf{p}_2) \end{bmatrix}. \quad (37)$$

Equation 37 can be rewritten in matrix form as, $\Theta_{\mathbf{H}}\mathcal{H} = \mathcal{C}$.

We wish to solve,

$$\min_{\mathcal{H}} \|\mathcal{B}_{\mathbf{H}} - \Gamma_{\mathbf{H}}\mathcal{H}\|_2 \quad \text{subject to} \quad \Theta_{\mathbf{H}}\mathcal{H} = \mathcal{C}. \quad (38)$$

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

$$\begin{bmatrix} \Gamma_{\mathbf{H}} & \Theta_{\mathbf{H}}^{T} \\ \Theta_{\mathbf{H}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{H} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathcal{B}_{\mathbf{H}} \\ \mathcal{C} \end{bmatrix}, \quad (39)$$

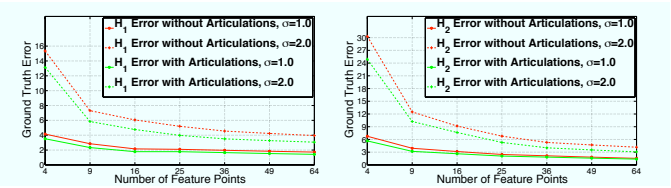
where $\Delta\mathbf{H}$ represents the update parameters for the homographies, $\mathbf{H}_i$ and $\mathbf{H}_j$ and λ is the Lagrange multiplier.

6 Results

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

6.1 Quantitative Evaluation

In this section, we quantitatively evaluate feature-based motion estimation under projective camera with articulation constraints. In addition, we evaluate the accuracy of gradient-based human motion estimation under affine camera against known ground-truth and also compare the proposed approach of exact equality articulation constraints against soft articulation constraints [9], [28].



(a) H_v Ground Truth Error (b) H_g Ground Truth Error

Fig. 6: Estimating homographies with varying number of noisy feature points with a Gaussian noise. The estimated homographies with and without the articulation constraint are compared against the ground truth homography and we can see that inclusion of articulation constraints consistently leads to low errors in homography estimation.

6.1.1 Do Articulation Constraints Help?

An interesting question that arises is whether incorporation of articulation constraints leads to accurate estimation of motion compared with the approach of not including the articulation constraints during motion estimation. Consider a doubly articulated planar system composed of two planes, a ground plane $\mathbf{\Pi}_g$ and a vertical plane $\mathbf{\Pi}_v$, that share two articulation points that defines their line of intersection. The planes $\mathbf{\Pi}_g$ and $\mathbf{\Pi}_v$ are viewed from a random pair of projective cameras, $\mathbf{P}_1$ and $\mathbf{P}_2$. Since in this setup, the plane equations and the camera matrices are known, we can estimate ground-truth homography associated with each plane [65]. We next select k feature point matches on each plane between the two views and add Gaussian noise σ to them to simulate feature matching noise. Homography can then be estimated with the articulation constraints using Equation 39. Homography without the articulation constraints can be estimated by setting the Lagrange multiplier to zero in Equation 39. Figure 6 shows the result of homography estimation for different number of feature matches k with and without the articulation constraints against the ground truth homographies for two different levels of Gaussian noise ($\sigma = 1.0$ and $\sigma = 2.0$) under 500 trials. To compare the ground truth homography against the estimated homography, we used, $||I - \mathbf{H}_{gt}^{-1}\mathbf{H}_{est}||^2$, where I is a 3×3 identity matrix, $\mathbf{H}_{gt}$ is the ground truth homography and $\mathbf{H}_{est}$ is the estimated homography [12]. It can be observed that inclusion of articulation constraints consistently leads to lower error in motion estimation, especially for feature points with higher Gaussian noise. This is to be expected since the inclusion of the articulation constraints helps to make the homographies of the two planes, $\mathbf{\Pi}_g$ and $\mathbf{\Pi}_v$, consistent with each other, thereby reducing the effect of noisy feature points.

6.1.2 Human Body Tracking: Comparison with Ground Truth

We manually labeled 300 frames (image size: 640×480) of a human performing a challenging activity recorded from a 30 frames per second (fps) camera. Motion was then estimated using a gradient-based algorithm under

an affine camera. The motion estimation algorithm was initialized in the first frame at the ground-truth location, and human motion tracked for the remainder of the sequences. Figure 7(a) compares the recovered body-part locations against the ground-truth for six body-parts. Figure 7(b) plots body-part location error as a percentage of the total distance moved by the body-part respectively. We can note that even though left wrist (LWrist) seems to have the highest absolute location error, as a percentage of distance moved, however, it has good accuracy. Figure 8 shows the ground-truth in green and the tracked body posture in red. Note the presence of motion blur due to a 30 fps camera and the presence of self-occlusion in the test sequence that makes motion estimation challenging.

6.1.3 Human Body Tracking: Drift Analysis

Motion estimation algorithms suffer from the problem of drift for a variety of reasons such as accumulation of subpixel errors or changes in object appearance. It is therefore important to quantify the magnitude of drift in motion estimation algorithms. We manually labeled another sequence of 100 frames of a person reaching for the glove box for the purpose of drift analysis. Drift was computed by first estimating motion in the forward direction, followed by, estimating motion in the reverse direction and comparing the difference in the initial position (first frame) versus the recovered initial position. Motion estimation was done using a gradient-based algorithm under an affine camera. For frame to frame motion, the accuracy of the estimated motion is at times subpixel. Figure 7(c) plots the drift error (in pixels) of individual articulations. Figure 7(d) plots drift error as a percentage of the total distance moved by the entire system of planes. As expected as the motion increases, the drift increases too, but remains under control across the sequence.

6.1.4 Comparing Proposed Exact Equality Articulations with Soft Articulations and No Articulations

In this subsection, we quantitatively compare the proposed exact equality articulation constraints for human body tracking using a gradient-based algorithm under an affine camera against the soft articulation constraints proposed by Ju *et al.* in [9], [28] and no articulation constraints. Given a corpus of labeled training data of $2,000$ frames for each of the 3 people, we estimated motion using the exact equality articulation constraint using Equation 23. Soft constraints proposed by Ju *et al.* in [9], [28] were also used to estimate motion. Motion without articulation constraints was estimated by setting the Lagrange multiplier to zero in Equation 23. Figure 9 qualitatively compares the three different articulation constraints for motion estimation. Top row for each person is motion estimation without any articulation constraints, the middle row for each person shows recovered motion using soft articulation constraints, while the last row for each person shows motion estimates using exact equality articulation constraints. It can be observed that in all the three cases imposition of exact equality constraints leads to accurate recovery of motion. Table 1 reports quantitative comparison for motion estimates in Figure 9. In the table, **W**, **E**, **S**, refer to the average root-mean square error across the image sequence for the two wrists, two elbows, and two shoulders respectively. We can observe that soft articulation constraints on an average lead to a 70% reduction in error against ground-truth as compared to no articulation constraints. Exact equality articulation constraints lead to on an average a further reduction of 65% over the soft articulation constraints.

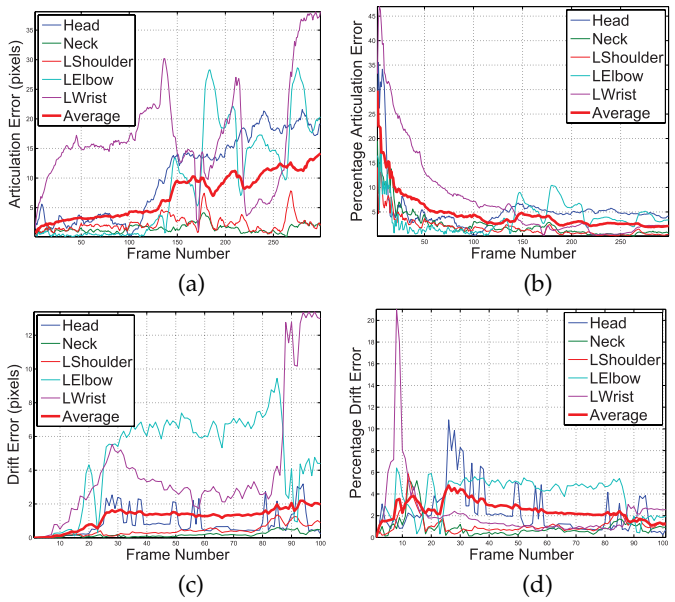


Fig. 7: Comparison of tracked joint position against the manually labeled data of a human steering the wheel. (a) Plots the absolute position error in pixels for individual articulations on a 640×480 image. (b) Plots tracking error as a percentage of the total distance moved by the entire system of planes. Measuring drift of motion estimation on a video sequence of a person reaching for the glove box. (c) Plots the absolute drift as measured in pixels for individual articulations. (d) Plots drift as a percentage of the total distance moved by the entire system of planes. The motion estimation algorithm has low error and low drift on complex activities.

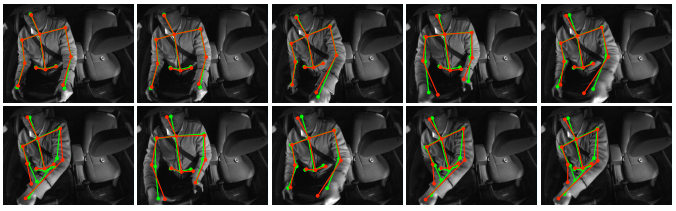


Fig. 8: Images corresponding to the evaluation in Figure 7. Green lines represent manually labeled ground-truth over the course of 300 frames of the video sequence. Red lines represent the tracked human body posture. Note the presence of fast-motion of wrist that cause motion blur and the presence of occlusion between the two wrists that present challenging scenarios for motion estimation.

	Exact Equality Constraints				Soft Constraints (Ju *et al.* [9], [28])				No Articulation Constraints			
Subject	**W**	**E**	**S**	**Average**	**W**	**E**	**S**	**Average**	**W**	**E**	**S**	**Average**
$\mathbf{S}_1$	0.011	0.029	0.065	**0.035**	0.259	0.125	0.113	0.166	0.650	0.889	0.362	0.634
$\mathbf{S}_2$	0.083	0.042	0.004	**0.043**	0.117	0.169	0.216	0.167	0.205	0.499	0.503	0.402
$\mathbf{S}_3$	0.091	0.086	0.132	**0.103**	0.170	0.127	0.270	0.189	0.644	0.949	0.423	0.672
Average	**0.061**	**0.052**	**0.067**	**0.060**	0.182	0.140	0.199	0.174	0.499	0.779	0.429	0.569

TABLE 1: Driving Sequence Quantitative Comparison: Imposition of exact equality articulation constraints for motion estimation results in more accurate motion estimation as compared to the soft and no articulation constraints. 'W' corresponds to the average RMS error for the left and the right wrist across the image sequence, similarly 'E' corresponds to elbow, and 'S' corresponds to shoulder. 'Average' corresponds to the average error across the row or the column as appropriate.

Fig. 9: Images corresponding to the results in Table 1. Top row for each person: motion estimation with no articulation constraints. Middle row for each person: motion estimation using Ju *et al.* [9], [28]. Bottom row for each person: motion estimation using exact equality articulation constraints.

6.2 Qualitative Evaluation

In this section, we perform qualitative evaluation of the proposed motion estimation algorithm using exact equality articulation constraints for human upper-body motion estimation, estimating the motion of rigid piecewise planar scenes, and estimating motion of nonrigid surfaces.

6.2.1 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 12,000 frames of 5 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 500 frames long at 60 fps.

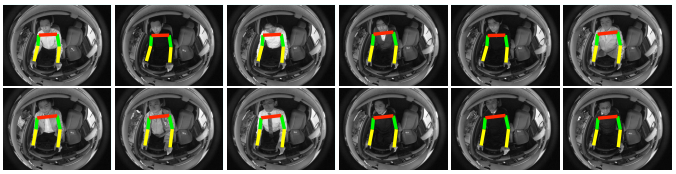


Fig. 10: Posture detection results for 9 different people with a variety of clothing and variations in the neutral posture.

Detection of Human Posture

Human posture was detected in images using the approach described in [66]. Given a large corpus of labeled training data, we learn posture specific parts-based appearance model for the human body. We focus on building a posture detector for the neutral posture of holding the steering wheel, since during a driving scenario it is the most common posture. Given several labeled instances of the neutral posture, we first use procrustes analysis to align the data. Then for each body part, a two dimensional normalized histogram, $\mathcal{A}_i$, is built that captures the frequency of observing image gradient over the normalized (x, y) coordinates. Given a collection of part appearance models, $\mathcal{A} = \{\mathcal{A}_1, \ldots, \mathcal{A}_N\}$, "candidate proposals" for neutral posture can be evaluated as,

$$f(\mathbf{X}|\{\mathcal{A}\}; \Delta\mathbf{I}_t) = \prod_{(x,y)} e^{\Delta\mathbf{I}_t(x,y) \times \mathcal{A}_\mathbf{X}(x,y)}, \qquad (40)$$

where $\Delta\mathbf{I}_t$ is the gradient magnitude map of $\mathbf{I}_t$, and $\mathcal{A}_\mathbf{X}$ is the expected sketch of the configuration $\mathbf{X}$. The sketch $\mathcal{A}_\mathbf{X}$ takes the histograms of individual body part frequencies and transforms them to the location of the body part defined by $\mathbf{X}$. Since the neutral posture resides in $\mathbb{R}^{12}$, an efficient search of the space of "candidate proposals" is important. We use the labeled data set of the neutral posture to learn a low dimensional, usually one, search space using PCA. Figure 10 shows the neutral posture detection result for 9 different people with a variety of clothing and variations in the neutral posture. For further details on the posture detector, we refer the interested reader to Sheikh *et al.* [66].

Given a detected neutral posture, 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 for gradient-based motion estimation under an affine camera. The articulations, which are initialized at the joint points of the detected posture, are used to set up

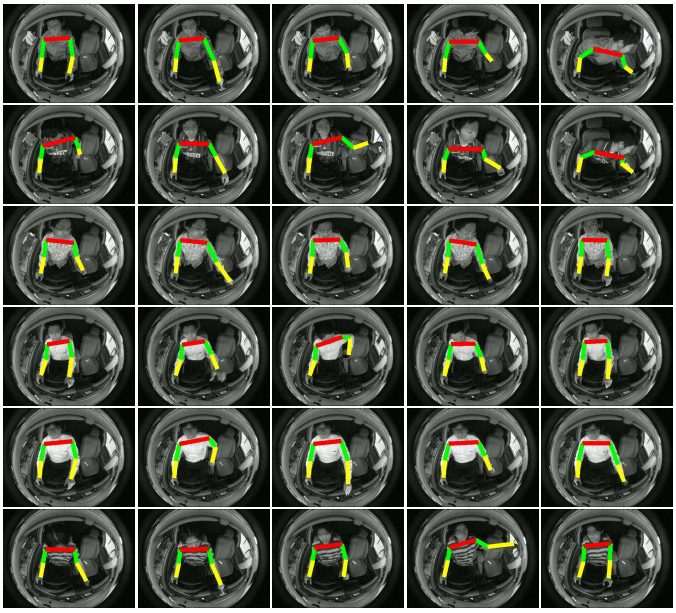
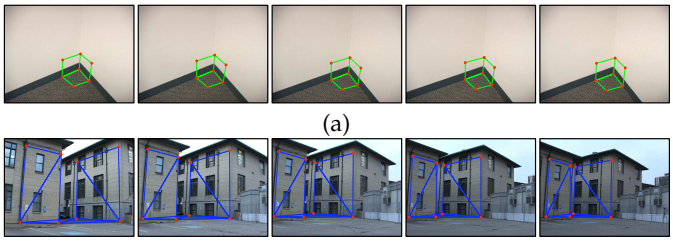


Fig. 11: Motion estimation for human upper-body. Our results demonstrate successful tracking of different individuals with different clothing performing a variety of activities inside a car. Note that since the entire network of articulation constraints bears down on the motion estimation of the body parts, we are able to use information from other articulations to estimate motion for blurred body parts during motion.

the linear constraint matrix, Θ. Thus, a system of 30 unknowns with 8 constraint equations is constructed, which is solved using the algorithm outlined in Figure 5 at an average speed of 90 fps on a COTS machine. 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 11. An interesting point can be made about tracking through motion blur, which is present in the video sequences. Since our tracking algorithm estimates the motion of each limb using all the 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 self-occlusions and the presence of strong background gradient from the center console box which looks just like an arm.

6.2.2 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 [61]. Figure 12 shows result of gradient-based motion estimation under an affine camera in scenes containing multiple planes fixed with respect to each other (in 3D). The outline of the planes in the images were manually initialized. It can be observed



(a)

(b)

Fig. 12: 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.

from the images that due to the articulation constraints, planes which have little or no texture can also be tracked. For example in Figure 12(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 12(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.

We can also observe the difference in homography estimation with and without the articulation constraints using feature-based motion estimation under projective camera in Figure 13. SIFT [67] features were obtained between the current image and the past image to perform feature driven tracking of rigid piecewise planar scenes in these images. Given the feature correspondences, we estimated warp parameters for homography with articulation constraints using non-linear gradient descent parameter updates as outlined in Equation 39. Homography without articulation constraints was also estimated using the same SIFT matches by setting the Lagrange multiplier to zero in Equation 39. Figure 13 shows that the incorporation of articulation constraints help to recover accurate homography over long image sequences as demonstrated by correct alignment of the planar image masks to the real planar facades.

6.2.3 Motion Estimation of Triangulated Meshes

Given a mesh constructed from Harris corner points, we set up a linear system using the pixels contained within each triangle as described in Section 4.2. The articulation 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 to obtain motion estimates for each triangle in the triangulated mesh, which is then used to propagate the feature points to the next frame with the same mesh connectivity. Figure 14 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

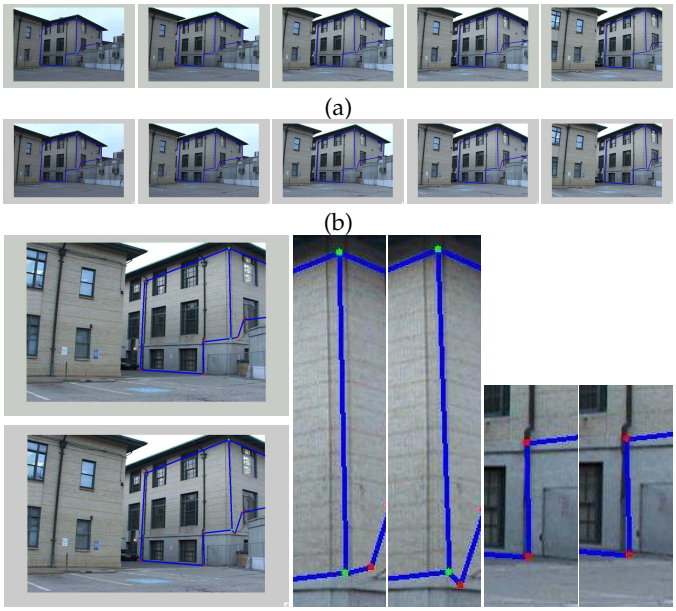


Fig. 13: Difference between homography estimation with and without the articulation constraints. The articulations are plotted in green and the red dots mark the outline of the tracked planes. (a) Tracking using homography *with* articulation constraints estimated using feature matches between images at consecutive temporal instances. (b) Tracking using homography *without* articulation constraints. (c) Top row shows tracked articulated planes using homography with articulation constraints, while bottom row shows otherwise. Column 2 and 3 show zoomed in differences between the tracking of the articulated planes. Left picture in each column corresponds to tracking using homography with articulation constraints and right picture otherwise. It can be observed that alignment in both the cases is better for the homography estimated using articulation constraints.

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 14(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 texture) and the resolution of triangulations (which allows greater expression of nonrigid motion).

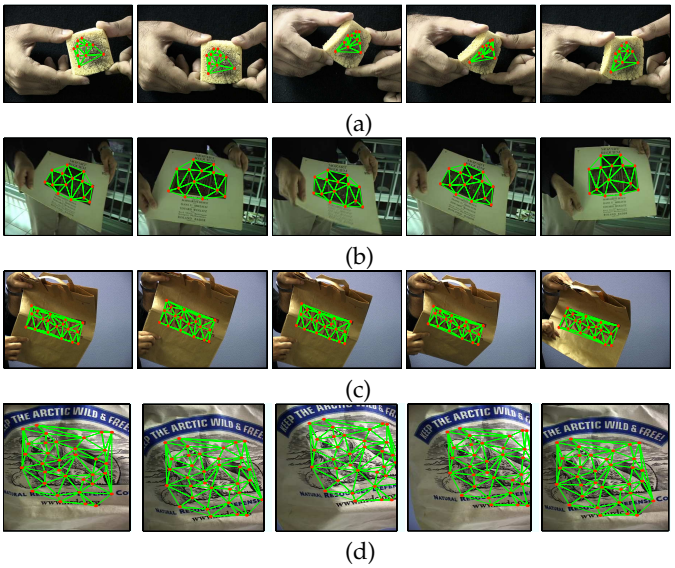


Fig. 14: Result of tracking a triangulated mesh on a variety of nonrigid surfaces. (a) Snapshots of large illumination change resistant tracking of a sponge. (b) Tracking a paper being moved in a wave-like manner. (c) Tracking large deformations on a paper bag. (d) Robust tracking of a cloth bag, where the points on the right side of the 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.

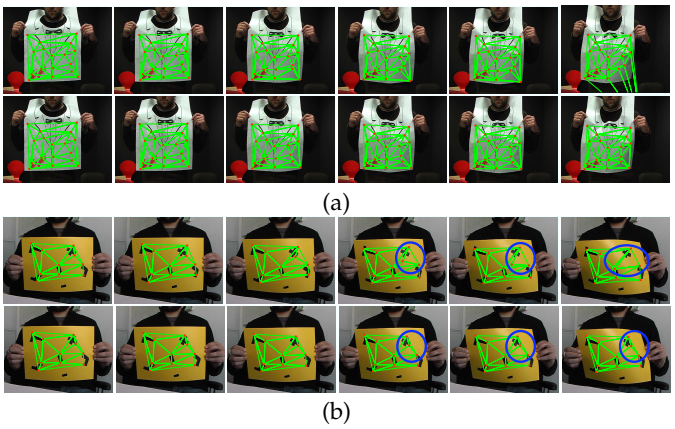


Fig. 15: Motion estimation for low texture nonrigid surfaces. (a & b) Top row: Motion estimation with no articulation constraints. (a & b) Bottom row: Motion estimation with exact equality articulation constraints. Imposition of articulation constraints leads to stable and accurate recovery of motion.

Motion estimation of low texture nonrigid surfaces

Imposition of exact equality articulation constraints on motion estimation helps to recover stable and accurate estimates of motion. Figure 15 shows two example of low texture surfaces: napkin and cardboard, that present challenging scenarios for recovery of motion due to the presence of low texture nonrigid surfaces. Figure 15 (a & b) top row shows recovered motion estimates using a gradient-based algorithm under an affine camera with no articulation constraints, while Figure 15 (a & b) bot-

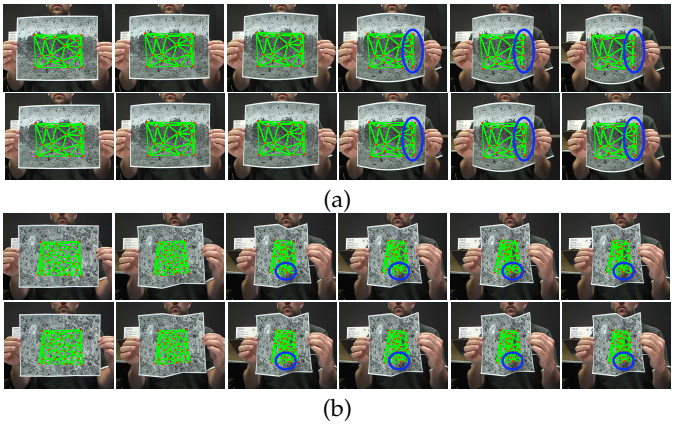


Fig. 16: Motion estimation for nonrigid surface with large deformations. (a & b) Top row: Motion estimation with no articulation constraints. (a & b) Bottom row: Motion estimation with exact equality articulation constraints. Since we solve for the motion of the mesh globally, all articulation constraints are used for motion estimation. Figure best seen when zoomed in.

tom row imposes exact equality articulation constraints and shows the recovered motion. Imposition of exact equality articulation constraints reduces the degree of freedom for the motion of the triangulated mesh and helps to recover stable and accurate motion estimates under low texture scenarios.

Motion estimation of nonrigid surfaces with large deformations

Figure 16 shows an example of a nonrigid surface undergoing large deformation that presents a challenging scenario for motion estimation. Figure 16 (a & b) top row shows recovered motion estimates with no articulation constraints, while Figure 16 (a & b) bottom row shows motion estimation after imposition of exact equality articulation constraints. Since, we solve for the motion of the entire triangulated mesh at the same time, all the articulation constraints are used for motion estimation. This leads to stable and accurate recovery of nonrigid surface motion.

7 Conclusion

In this paper, we have presented the explicit application of articulation constraints to motion estimation algorithms for an affine and a projective camera to recover a variety of real world motions. The motion estimation algorithm constructs an over-constrained system of linear equations subject to linear, exact equality constraints, in case of an affine camera, and linearized articulation constraints are used, in the case of a projective camera, to solve for the motion of multiple layers simultaneously. Since, we solve for the motion of all layers 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.

During experimentation, we noted two primary sources of error. The first source of error is self-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 underlying surface 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.

The value of our framework 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 reduces arbitrariness and points to broad applicability of the framework to a variety of real-world motion estimation tasks as demonstrated in this paper. We do not require the inclusion of any physics-based prior or data prior for motion estimation, even though, they may be added but perhaps with the loss of the linear formulation of the presented motion estimation algorithm. Investigating inclusion of such shape and/or motion priors within a linear framework for stable and efficient motion estimation remains as a source of future research.

Appendix A

The Jacobian $J_{f\mathbf{H}_\Pi}$ of the transfer error function f is a $(2N) \times 9$ matrix, where N is the number of feature matches for plane $\mathbf{\Pi}$. The $(2 \times i) - 1^{\text{th}}$ and $2 \times i^{\text{th}}$-row of the matrix for a pair of corresponding feature points $(\mathbf{p}_i, \mathbf{p}_i^{'})$ are as follows, where $i \in \{1, \ldots, N\}$,

$$\frac{1}{\mathbf{h}_3^\top \mathbf{p}_i} \begin{bmatrix} -2\mathbf{p}_{1x}\eta_1(\mathbf{p}_{1x}, \mathbf{p}_{1x}^{'}) & 0 \\ -2\mathbf{p}_{1y}\eta_1(\mathbf{p}_{1x}, \mathbf{p}_{1x}) & 0 \\ -2\eta_1(\mathbf{p}_{1x}, \mathbf{p}_{1x}^{'}) & 0 \\ 0 & -2\mathbf{p}_{1x}\eta_2(\mathbf{p}_{1y}, \mathbf{p}_{1y}^{'}) \\ 0 & -2\mathbf{p}_{1y}\eta_2(\mathbf{p}_{1y}, \mathbf{p}_{1y}) \\ 0 & -2\eta_2(\mathbf{p}_{1y}, \mathbf{p}_{1y}^{'}) \\ \frac{2\mathbf{p}_{1x}\mathbf{h}_1^\top \mathbf{p}_1 \eta_1(\mathbf{p}_{1x}, \mathbf{p}_{1x}^{'})}{\mathbf{h}_3^\top \mathbf{p}_1} & \frac{2\mathbf{p}_{1x}\mathbf{h}_2^\top \mathbf{p}_1 \eta_2(\mathbf{p}_{1y}, \mathbf{p}_{1y}^{'})}{\mathbf{h}_3^\top \mathbf{p}_1} \\ \frac{2\mathbf{p}_{1y}\mathbf{h}_1^\top \mathbf{p}_1 \eta_1(\mathbf{p}_{1x}, \mathbf{p}_{1x}^{'})}{\mathbf{h}_3^\top \mathbf{p}_1} & \frac{2\mathbf{p}_{1y}\mathbf{h}_2^\top \mathbf{p}_1 \eta_2(\mathbf{p}_{1y}, \mathbf{p}_{1y}^{'})}{\mathbf{h}_3^\top \mathbf{p}_1} \\ \frac{2\mathbf{h}_1^\top \mathbf{p}_1 \eta_1(\mathbf{p}_{1x}, \mathbf{p}_{1x}^{'})}{\mathbf{h}_3^\top \mathbf{p}_1} & \frac{2\mathbf{h}_2^\top \mathbf{p}_1 \eta_2(\mathbf{p}_{1y}, \mathbf{p}_{1y}^{'})}{\mathbf{h}_3^\top \mathbf{p}_1} \end{bmatrix}^\top,$$

where

$$\eta_1(\mathbf{p}_{ix}, \mathbf{p}_{ix}^{'}) = \mathbf{p}_{ix}^{'} - \frac{\mathbf{h}_1^\top \mathbf{p}_i}{\mathbf{h}_3^\top \mathbf{p}_i}, \tag{41}$$

$$\eta_2(\mathbf{p}_{iy}, \mathbf{p}_{iy}^{'}) = \mathbf{p}_{iy}^{'} - \frac{\mathbf{h}_2^\top \mathbf{p}_i}{\mathbf{h}_3^\top \mathbf{p}_i}. \tag{42}$$

APPENDIX B

The Jacobian $J_\Phi(\mathbf{H}_i, \mathbf{H}_j; \mathbf{p}_1, \mathbf{p}_2)$ is a 4×18 matrix which is as follows:

$$2\begin{bmatrix}
\frac{\mathbf{p}_{1x}\eta_1(\mathbf{p}_1)}{\mathbf{h}_{3i}^\top\mathbf{p}_1} & 0 & \frac{\mathbf{p}_{2x}\eta_1(\mathbf{p}_2)}{\mathbf{h}_{3i}^\top\mathbf{p}_2} & 0 \\
\frac{\mathbf{p}_{1y}\eta_1(\mathbf{p}_1)}{\mathbf{h}_{3i}^\top\mathbf{p}_1} & 0 & \frac{\mathbf{p}_{2y}\eta_1(\mathbf{p}_2)}{\mathbf{h}_{3i}^\top\mathbf{p}_2} & 0 \\
\frac{\eta_1(\mathbf{p}_1)}{\mathbf{h}_{3i}^\top\mathbf{p}_1} & 0 & \frac{\eta_1(\mathbf{p}_2)}{\mathbf{h}_{3i}^\top\mathbf{p}_2} & 0 \\
0 & \frac{\mathbf{p}_{1x}\eta_2(\mathbf{p}_1)}{\mathbf{h}_{3i}^\top\mathbf{p}_1} & 0 & \frac{\mathbf{p}_{2x}\eta_2(\mathbf{p}_2)}{\mathbf{h}_{3i}^\top\mathbf{p}_2} \\
0 & \frac{\mathbf{p}_{1y}\eta_2(\mathbf{p}_1)}{\mathbf{h}_{3i}^\top\mathbf{p}_1} & 0 & \frac{\mathbf{p}_{2y}\eta_2(\mathbf{p}_2)}{\mathbf{h}_{3i}^\top\mathbf{p}_2} \\
0 & \frac{\eta_2(\mathbf{p}_1)}{\mathbf{h}_{3i}^\top\mathbf{p}_1} & 0 & \frac{\eta_2(\mathbf{p}_2)}{\mathbf{h}_{3i}^\top\mathbf{p}_2} \\
\frac{-\mathbf{p}_{1x}\mathbf{h}_{1i}^\top\mathbf{p}_1\eta_1(\mathbf{p}_1)}{(\mathbf{h}_{3i}^\top\mathbf{p}_1)^2} & \frac{-\mathbf{p}_{1x}\mathbf{h}_{2i}^\top\mathbf{p}_1\eta_2(\mathbf{p}_1)}{(\mathbf{h}_{3i}^\top\mathbf{p}_1)^2} & \frac{-\mathbf{p}_{2x}\mathbf{h}_{1i}^\top\mathbf{p}_2\eta_1(\mathbf{p}_2)}{(\mathbf{h}_{3i}^\top\mathbf{p}_2)^2} & \frac{-\mathbf{p}_{2x}\mathbf{h}_{2i}^\top\mathbf{p}_2\eta_2(\mathbf{p}_2)}{(\mathbf{h}_{3i}^\top\mathbf{p}_2)^2} \\
\frac{-\mathbf{p}_{1y}\mathbf{h}_{1i}^\top\mathbf{p}_1\eta_1(\mathbf{p}_1)}{(\mathbf{h}_{3i}^\top\mathbf{p}_1)^2} & \frac{-\mathbf{p}_{1y}\mathbf{h}_{2i}^\top\mathbf{p}_1\eta_2(\mathbf{p}_1)}{(\mathbf{h}_{3i}^\top\mathbf{p}_1)^2} & \frac{-\mathbf{p}_{2y}\mathbf{h}_{1i}^\top\mathbf{p}_2\eta_1(\mathbf{p}_2)}{(\mathbf{h}_{3i}^\top\mathbf{p}_2)^2} & \frac{-\mathbf{p}_{2y}\mathbf{h}_{2i}^\top\mathbf{p}_2\eta_2(\mathbf{p}_2)}{(\mathbf{h}_{3i}^\top\mathbf{p}_2)^2} \\
\frac{-\mathbf{h}_{1i}^\top\mathbf{p}_1\eta_1(\mathbf{p}_1)}{(\mathbf{h}_{3i}^\top\mathbf{p}_1)^2} & \frac{-\mathbf{h}_{2i}^\top\mathbf{p}_1\eta_2(\mathbf{p}_1)}{(\mathbf{h}_{3i}^\top\mathbf{p}_1)^2} & \frac{-\mathbf{h}_{1i}^\top\mathbf{p}_2\eta_1(\mathbf{p}_2)}{(\mathbf{h}_{3i}^\top\mathbf{p}_2)^2} & \frac{-\mathbf{h}_{2i}^\top\mathbf{p}_2\eta_2(\mathbf{p}_2)}{(\mathbf{h}_{3i}^\top\mathbf{p}_2)^2} \\
\frac{-\mathbf{p}_{1x}\eta_1(\mathbf{p}_1)}{\mathbf{h}_{3j}^\top\mathbf{p}_1} & 0 & \frac{-\mathbf{p}_{2x}\eta_1(\mathbf{p}_2)}{\mathbf{h}_{3j}^\top\mathbf{p}_2} & 0 \\
\frac{-\mathbf{p}_{1y}\eta_1(\mathbf{p}_1)}{\mathbf{h}_{3j}^\top\mathbf{p}_1} & 0 & \frac{-\mathbf{p}_{2y}\eta_1(\mathbf{p}_2)}{\mathbf{h}_{3j}^\top\mathbf{p}_2} & 0 \\
\frac{-\eta_1(\mathbf{p}_1)}{\mathbf{h}_{3j}^\top\mathbf{p}_1} & 0 & \frac{-\eta_1(\mathbf{p}_2)}{\mathbf{h}_{3j}^\top\mathbf{p}_2} & 0 \\
0 & \frac{-\mathbf{p}_{1x}\eta_2(\mathbf{p}_1)}{\mathbf{h}_{3j}^\top\mathbf{p}_1} & 0 & \frac{-\mathbf{p}_{2x}\eta_2(\mathbf{p}_2)}{\mathbf{h}_{3j}^\top\mathbf{p}_2} \\
0 & \frac{-\mathbf{p}_{1y}\eta_2(\mathbf{p}_1)}{\mathbf{h}_{3j}^\top\mathbf{p}_1} & 0 & \frac{-\mathbf{p}_{2y}\eta_2(\mathbf{p}_2)}{\mathbf{h}_{3j}^\top\mathbf{p}_2} \\
0 & \frac{-\eta_2(\mathbf{p}_1)}{\mathbf{h}_{3j}^\top\mathbf{p}_1} & 0 & \frac{-\eta_2(\mathbf{p}_2)}{\mathbf{h}_{3j}^\top\mathbf{p}_2} \\
\frac{\mathbf{p}_{1x}\mathbf{h}_{1j}^\top\mathbf{p}_1\eta_1(\mathbf{p}_1)}{(\mathbf{h}_{3j}^\top\mathbf{p}_1)^2} & \frac{\mathbf{p}_{1x}\mathbf{h}_{2j}^\top\mathbf{p}_1\eta_2(\mathbf{p}_1)}{(\mathbf{h}_{3j}^\top\mathbf{p}_1)^2} & \frac{\mathbf{p}_{2x}\mathbf{h}_{1j}^\top\mathbf{p}_2\eta_1(\mathbf{p}_2)}{(\mathbf{h}_{3j}^\top\mathbf{p}_2)^2} & \frac{\mathbf{p}_{2x}\mathbf{h}_{2j}^\top\mathbf{p}_2\eta_2(\mathbf{p}_2)}{(\mathbf{h}_{3j}^\top\mathbf{p}_2)^2} \\
\frac{\mathbf{p}_{1y}\mathbf{h}_{1j}^\top\mathbf{p}_1\eta_1(\mathbf{p}_1)}{(\mathbf{h}_{3j}^\top\mathbf{p}_1)^2} & \frac{\mathbf{p}_{1y}\mathbf{h}_{2j}^\top\mathbf{p}_1\eta_2(\mathbf{p}_1)}{(\mathbf{h}_{3j}^\top\mathbf{p}_1)^2} & \frac{\mathbf{p}_{2y}\mathbf{h}_{1j}^\top\mathbf{p}_2\eta_1(\mathbf{p}_2)}{(\mathbf{h}_{3j}^\top\mathbf{p}_2)^2} & \frac{\mathbf{p}_{2y}\mathbf{h}_{2j}^\top\mathbf{p}_2\eta_2(\mathbf{p}_2)}{(\mathbf{h}_{3j}^\top\mathbf{p}_2)^2} \\
\frac{\mathbf{h}_{1j}^\top\mathbf{p}_1\eta_1(\mathbf{p}_1)}{(\mathbf{h}_{3j}^\top\mathbf{p}_1)^2} & \frac{\mathbf{h}_{2j}^\top\mathbf{p}_1\eta_2(\mathbf{p}_1)}{(\mathbf{h}_{3j}^\top\mathbf{p}_1)^2} & \frac{\mathbf{h}_{1j}^\top\mathbf{p}_2\eta_1(\mathbf{p}_2)}{(\mathbf{h}_{3j}^\top\mathbf{p}_2)^2} & \frac{\mathbf{h}_{2j}^\top\mathbf{p}_2\eta_2(\mathbf{p}_2)}{(\mathbf{h}_{3j}^\top\mathbf{p}_2)^2}
\end{bmatrix}^\top$$

where $\{\mathbf{p}_x, \mathbf{p}_y\}$ represents the (x, y) component respectively of an articulation point $\mathbf{p}$ and

$$\eta_1(\mathbf{p}_i; \mathbf{H}_i, \mathbf{H}_j) = \frac{\mathbf{h}_{1i}^\top\mathbf{p}_i}{\mathbf{h}_{3i}^\top\mathbf{p}_i} - \frac{\mathbf{h}_{1j}^\top\mathbf{p}_i}{\mathbf{h}_{3j}^\top\mathbf{p}_i}, \tag{43}$$

$$\eta_2(\mathbf{p}_i; \mathbf{H}_i, \mathbf{H}_j) = \frac{\mathbf{h}_{2i}^\top\mathbf{p}_i}{\mathbf{h}_{3i}^\top\mathbf{p}_i} - \frac{\mathbf{h}_{2j}^\top\mathbf{p}_i}{\mathbf{h}_{3j}^\top\mathbf{p}_i}. \tag{44}$$

ACKNOWLEDGMENT

The research described in this paper was supported by the DENSO Corporation, Japan. We thank Adrien Bartoli for sharing data used during experimentation. We also thank Mathieu Salzmann for sharing datasets for nonrigid surface reconstruction.

REFERENCES

[1] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Image Understanding Workshop*, 1981.
[2] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," in *Second ECCV*, 1992.
[3] L. Sigal, S. Bhatia, S. Roth, M. J. Black, and M. Isard, "Tracking loose-limbed people," in *CVPR*, 2004.
[4] M. Black and A. Jepson, "Eigentracking: Robust matching and tracking of articulated objects using a view-based representation," in *IJCV*, 1998.
[5] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," in *ECCV*, 1998.
[6] Y. Weiss, "Smoothness in layers: Motion segmentation using nonparametric mixture estimation," in *CVPR*, 1997.
[7] V. G. Bellile, M. Perriollat, A. Bartoli, and P. Sayd, "Image registration by combining thin-plate splines with a 3D morphable model," in *Internation Conference on Image Processing*, 2006.
[8] J. Lim and M. H. Yang, "A direct method for modeling non-rigid motion with thin plate spline," in *CVPR*, 2005.
[9] S. Ju, M. Black, and Y. Yacoob, "Cardboard people: A parameterized model of articulated image motion," in *Face and Gesture Recognition*, 1996.
[10] J. Y. A. Wang and E. H. Adelson, "Representing moving images with layers," *Transactions on Image Processing*, vol. 3, no. 5, 1994.
[11] H. S. Sawhney and S. Ayer, "Compact representations of videos through dominant and multiple motion estimation," *PAMI*, vol. 18, no. 8, 1996.
[12] L. Zelnik-Manor and M. Irani, "Multiview constraints on homographies," in *PAMI*, 2002.
[13] H. Nagel, "Displacement vectors derived from second-order intensity variations in image sequences," *Graphical Model and Image Processing*, vol. 21, no. 1, pp. 85–117, January 1983.
[14] C. Fennema and W. Thompson, "Velocity determination in scenes containing several moving objects," *Graphical Model and Image Processing*, vol. 9, no. 4, pp. 301–315, April 1979.
[15] B. Schunck and B. Horn, "Determining optical flow," in *MIT AI Memo*, 1980.
[16] S. Uras, F. Girosi, A. Verri, and V. Torre, "A computational approach to motion perception," *BioCyber*, vol. 60, pp. 79–87, 1989.
[17] F. Glazer, G. Reynold, and P. Anandan, "Scene matching through hierarchical correlation," in *In Proc. CVPR*, 1983, pp. 432–441.
[18] P. Anandan, "A unified perspective on computational techniques for the measurement of visual motion," in *ICCV*, 1987, pp. 219–230.
[19] P. J. Burt, C. Yen, and X. Xu, "Multiresolution flow through motion analysis," in *In Proc. CVPR*, 1983.
[20] J. J. Little, H. H. Bulthoff, and T. A. Poggio, "Parallel optical flow using local voting," in *In Proc. ICCV*, 1988, pp. 454–459.
[21] S. S. Beauchemin and J. L. Barron, "The computation of optical flow," *ACM Comput. Surv.*, vol. 27, no. 3, pp. 433–466, 1995.
[22] D. J. Fleet and Y. Weiss, "Optical flow estimation," *Handbook of Mathematical Models in Computer Vision*, 2006.
[23] S. Ayer and H. Sawhney, "Layered representation of motion video using robust maximum-likelihood estimation of mixture models and mdl encoding," in *ICCV*, 1995.
[24] Y. Weiss, "Smoothness in layers: Motion segmentation using nonparametric mixture estimation," in *CVPR*, 1997, pp. 520–526.
[25] P. Anandan, R. Szeliski, and P. Torr, "An integrated bayesian approach to layer extraction from image sequences," in *ICCV*, 1999, pp. 983–990.
[26] R. Szeliski, S. Avidan, and P. Anandan, "Layer extraction from multiple images containing reflections and transparency," in *CVPR*, 2000, pp. I: 246–253.
[27] Q. Ke and T. Kanade, "A subspace approach to layer extraction," in *CVPR*, 2001.
[28] S. Ju, M. Black, and A. Jepson, "Skin and bones: Multi-layer, locally affine, optical flow and regularization with transparency," in *CVPR*, 1996.
[29] N. R. Howe, M. E. Leventon, and W. T. Freeman, "Bayesian reconstruction of 3-d human motion from single-camera video," in *Advances in Neural Information Processing Systems (NIPS)*, 1999.
[30] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4s: A real-time system detecting and tracking people in 2 1/2d," in *ECCV*, 1998, pp. 877–892.
[31] Y. Huang and T. S. Huang, "Model-based human body tracking," in *ICPR*, 2002.
[32] T. jen Cham and J. M. Rehg, "A multiple hypothesis approach to figure tracking," in *CVPR*, 1999.
[33] A. Agarwal and B. Triggs, "Tracking articulated motion using a mixture of autoregressive models," in *ECCV*, 2004, pp. 54–65.
[34] C. Bregler, J. Malik, and K. Pullen, "Twist based acquisition and tracking of animal and human kinematics," *IJCV*, vol. 56, no. 3, 2004.
[35] C. Bregler and J. Malik, "Tracking people with twists and exponential maps," in *CVPR*, 1998.
[36] J. M. Rehg and T. Kanade, "Model-based tracking of self-occluding articulated objects," in *ICCV*, 1995, pp. 612–617.
[37] D. M. Gavrila and L. S. Davis, "3-d model-based tracking of humans in action: a multi-view approach," in *CVPR*, 1996.
[38] Y. Yacoob and L. S. Davis, "Learned models for estimation of rigid and articulated human motion from stationary or moving camera," *IJCV*, vol. 36, no. 1, pp. 5–30, 2000.
[39] I. A. Kakadiaris and D. N. Metaxas, "Model-based estimation of 3d human motion," *PAMI*, vol. 22, no. 12, pp. 1453–1459, 2000.

[40] M. Yamamoto and K. Yagishita, "Scene constraints-aided tracking of human body," in *CVPR*, 2000.
[41] A. Ruf and R. Horaud, "Rigid and articulated motion seen with an uncalibrated stereo rig," in *ICCV*, 1999.
[42] L. Sigal and M. J. Black, "Measure locally, reason globally: Occlusion-sensitive articulated pose estimation," in *CVPR*, 2006.
[43] D. Demirdjian, T. Ko, and T. Darrell, "Constraining human body tracking," in *ICCV*, 2003.
[44] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *PAMI*, vol. 11, no. 6, 1989.
[45] M. J. Black and Y. Yacoob, "Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion," in *ICCV*, 1995.
[46] S. Sclaroff and J. Isidoro, "Active blobs," in *ICCV*, 1998.
[47] A. Bartoli and A. Zisserman, "Direct estimation of non-rigid registration," in *In Proc. BMVC*, 2004.
[48] V. Gay Bellile, A. Bartoli, and P. Sayd, "Feature-driven direct non-rigid image registration," in *In Proc. BMVC*, 2007.
[49] T. F. Cootes, S. Marsl, C. J. Twining, K. Smith, and C. J. Taylor, "Groupwise diffeomorphic non-rigid registration for automatic model building," in *In Proc. ECCV*. Springer, 2004, pp. 316–327.
[50] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *IJCV*, vol. V1, no. 4, pp. 321–331, January 1988.
[51] T. Mcinerney and D. Terzopoulos, "A finite element model for 3d shape reconstruction and nonrigid motion tracking," in *ICCV*, 1993, pp. 518–523.
[52] D. Metaxas and D. Terzopoulos, "Shape and nonrigid motion estimation through physics-based synthesis," *PAMI*, vol. 15, no. 6, 1993.
[53] A. P. Pentland, "Automatic extraction of deformable part models," *IJCV*, vol. 4, no. 2, pp. 107–126, 1990.
[54] L. D. Cohen and I. Cohen, "Finite-element methods for active contour models and balloons for 2-d and 3-d images," *PAMI*, vol. 15, no. 11, pp. 1131–1147, 1993.
[55] H. Delingette, M. Hebert, and K. Ikeuchi, "Deformable surfaces: A free-form shape representation," in *Proc. SPIE Vol 1570: Geometric Methods in Computer Vision*, 1991, pp. 21–30.
[56] X. Llado, A. Del Bue, and L. Agapito, "Non-rigid 3d factorization for projective reconstruction," in *BMVC*, Sept 2005.
[57] L. T. Stanford, A. Hertzmann, and C. Bregler, "Learning non-rigid 3d shape from 2d motion," in *NIPS*, 2003, pp. 1555–1562.
[58] M. Salzmann, R. Urtasun, and P. Fua, "Local deformation models for monocular 3d shape recovery," in *CVPR*, 2008.
[59] H. J. Lee and Z. Chen, "Determination of 3D human body postures from a single view," *Graphical Model and Image Processing*, vol. 30, 1985.
[60] 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.
[61] B. Johansson, "View synthesis and 3D reconstruction of piecewise planar scenes using intersection lines between the planes," in *ICCV*, 1999.
[62] P. Pritchett and A. Zisserman, "Matching and reconstruction from widely separated views," in *3D Structure from Multiple Images of Large-Scale Environments*, 1998.
[63] J. Semple and G. Kneebone, "Algebraic projective geometry," *Oxford University Press*, 1952.
[64] P. Gill, W. Murray, and M. Wright, "Practical optimization," in *Academic Press*, 1981.
[65] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.
[66] Y. A. Sheikh, A. Datta, and T. Kanade, "On the sustained tracking of human motion," in *8th IEEE International Conference on Automatic Face and Gesture Recognition*, September 2008.
[67] D. Lowe, "Distinctive image features from scale-invariant keypoints," in *IJCV*, vol. 20, 2003.

Ankur Datta Ankur Datta received PhD degree in Robotics from Carnegie Mellon University on a NSF Graduate Fellowship in 2010 and BSc degree in Computer Science with University Honors and Honors in the Major from University of Central Florida in 2004. He received the best paper award at the ICCV THEMIS workshop in 2009 for work on human motion estimation. His current research interest are in human motion estimation, camera calibration, and structured learning.

Yaser Sheikh Yaser Sheikh is an Assistant Research Professor at the Robotics Institute and Adjunct Faculty at the Department of Mechanical Engineering, at Carnegie Mellon University. His research area is computer vision, primarily in analyzing dynamic scenes in 3D, including human activity analysis, dynamic scene reconstruction, mobile camera networks, and nonrigid motion estimation. He obtained his doctoral degree from the University of Central Florida in 2006. He has served as Associate Editor for IAPR International Conference on Pattern Recognition 2010, Program Committee Member for Three Dimensional Information Extraction for Video Analysis and Mining, 2010 (held in conjunction with CVPR 2010), and Track Chair for International Conference on Multimedia Computing and Information Technology, 2010. He won the best paper award at the ICCV THEMIS workshop in 2009 and the Hillman Fellowship in 2004.

Takeo Kanade Takeo Kanade is the U. A. and Helen Whitaker University Professor of Computer Science and Robotics and the director of Quality of Life Technology Engineering Research Center at Carnegie Mellon University. He is also the director of Digital Human Research Center in Tokyo, which he founded in 2001. He received his Doctoral degree in Electrical Engineering from Kyoto University, Japan, in 1974. After holding a faculty position in the Department of Information Science, Kyoto University, he joined Carnegie Mellon University in 1980, where he was the Director of the Robotics Institute from 1992 to 2001.

Dr. Kanade works in multiple areas of robotics: computer vision, multimedia, manipulators, autonomous mobile robots, medical robotics and sensors. He has written more than 300 technical papers and reports in these areas, and holds more than 20 patents. He has been the principal investigator of more than a dozen major vision and robotics projects at Carnegie Mellon.

Dr. Kanade has been elected to the National Academy of Engineering (1997) and the American Academy of Arts and Sciences (2004). He is a Fellow of the IEEE, a Fellow of the ACM, a Founding Fellow of American Association of Artificial Intelligence (AAAI), and the former and founding editor of International Journal of Computer Vision. He has received many awards, including the Franklin Institute Bower Prize, Okawa Award, C&C Award, Joseph Engelberger Award, IEEE Robotics and Automation Society Pioneer Award, FIT Accomplishment Award, and IEEE PAMI-TC Azriel Rosenfeld Lifetime Accomplishment Award. Dr. Kanade has served on advisory or consultant committees for government, industry and university, including the Aeronautics and Space Engineering Board (ASEB) of National Research Council, NASA's Advanced Technology Advisory Committee, PITAC Panel for Transforming Healthcare Panel, and the Advisory Board of Canadian Institute for Advanced Research.