

# TOWARDS POSE INVARIANT GAIT RECONSTRUCTION

Nick Spencer, John Carter {nms00r | jnc}@ecs.soton.ac.uk

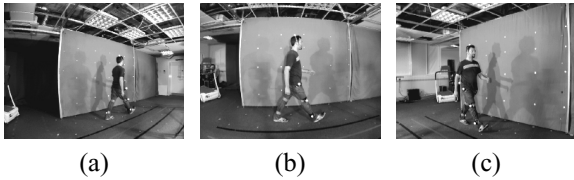
Image, Speech and Intelligent Systems  
University of Southampton, UK

## ABSTRACT

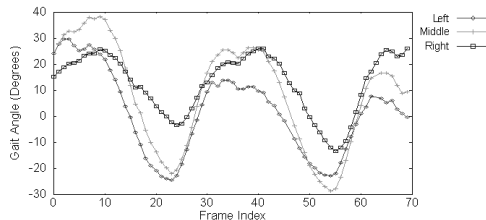
Recently, a lot of research has been conducted into the usefulness of gait for identification at a distance. Since the gait of a person is readily identified when extracted from a canonical side view, most algorithms work with the premise that the motion is frontoparallel in nature, or require some knowledge of the camera calibration. Realistically people will always walk along different trajectories to the camera. In this paper we show that gait has sufficient properties that allows us to exploit the structure of articulated motion within single view sequences, in order to remove the unknown subject pose and reconstruct the underlying gait signature, with no prior knowledge of the camera calibration.

## 1. INTRODUCTION

The effect of varying subject trajectory pose is quite pronounced on the set of measured gait angles.



**Fig. 1.** Three views of a subject walking along a linear trajectory with constant velocity.



**Fig. 2.** Gait angles of upper leg extracted for all three views.

Thanks to Lee Middleton for his most generous help in the lab. We gratefully acknowledge partial support by the European research office of the US Army under contract No. N68171-01-C-9002.

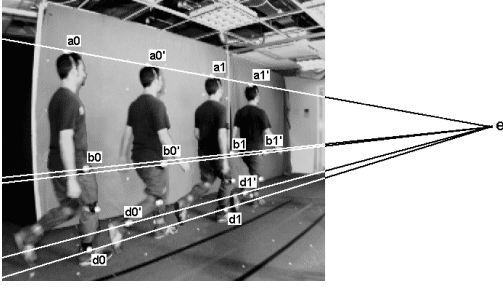
For recognition purposes we must then either use a biometric that is invariant to subject pose or be able to reconstruct the canonical frontoparallel view of gait motion. Researchers have recently given much attention to gait recognition for the canonical view and have successfully shown that identification is possible for different gait speeds [1]. Others have tried to synthesize frontoparallel views by use of structure from motion [2, 3], but require some information about camera calibration. Affine multi-view 3D reconstruction of gait is possible by consideration of rigid bone lengths [4]. In this paper we propose a method to reconstruct gait motion from monocular image sequences by taking advantage of the constraints of articulated limb motions. We assume no prior knowledge of the camera calibration, only that people walk in straight lines with constant velocity and limbs that swing in planes. We assume that the computer vision task of finding limb landmark points and tracking them over all frames in the sequence is solved.

## 2. A STRATIFIED APPROACH TO LINEAR TRAJECTORY GAIT RECONSTRUCTION

Multiple periods of linear gait motion is analogous to a single period viewed from many cameras related by linear translation. The matching landmark correspondences lie in an auto-epipolar configuration with the imaged motion direction  $\mathbf{e}$ , see fig. 3. To find correspondence matches we compute the vector of fitting cost errors for each putative value of period. The epipole is found by computing the intersection of the set of lines formed through each of the putative matched cluster sets. The periodicity error is calculated by applying the stereopsis transformation that maps the epipole  $\mathbf{e}$  to the ideal point  $(1, 0, 0)^T$  and computing the cost based on dot product between matching limb segment vectors. Most poorly fitted values occur in the head and tail segments of the cost vector so the smoothed errors are fitted to a first order harmonic series with Gaussian envelope.

### 2.1. Recovering Affine Structure

We first determine the normalization transform  $\mathbf{K}_n$  with isotropic scaling that maps the set of landmark point tracks



**Fig. 3.** Imaged matching similar phase poses. Landmark points **a** – **d** shown with matched poses. Each landmark point has  $T_p$  phase positions **a**, **a'**,  $\dots$ , **a\***, with each set of matching phase clusters  $(\mathbf{a}_0, \dots, \mathbf{a}_n), (\mathbf{a}'_0, \dots, \mathbf{a}'_n), \dots$  in auto-epipolar configuration with the motion epipole **e**.

to the unit square  $\mathbf{x}' = \mathbf{K}_n \mathbf{x}$ , then transform the epipole  $\mathbf{e}' = \mathbf{K}_n \mathbf{e}$  and re-normalize to unit norm  $\|\mathbf{e}'\| = 1$ . We then compute the interpolated set of re-sampled landmark point tracks that approximate the similar phase matches by assuming linear velocity between frames. For each set of matching phase clusters, see fig. 3, find the optimal point estimates  $\hat{\mathbf{x}}_i$  that lie on the epipolar line satisfying the condition  $\hat{\mathbf{x}}_i^\top [\mathbf{e}']_\times \hat{\mathbf{x}}_j = 0$ . The back projected rays formed from a set of optimal point estimates intersect in a single workspace point. The back projection of all sets of point estimates generates the cluster of 3D point tracks for an assumed single period of reconstructed gait motion. We use the Direct Linear Transform,  $([\hat{\mathbf{x}}_k]_\times \mathbf{P}_k) \cdot \mathbf{X} = \mathbf{0}$ , with the set of camera projection matrices  $\mathbf{P}_k = [\mathbf{R}_e^\top \mid -k \cdot \mathbf{e}']$  to compute the 3D points  $\mathbf{X}$ , where  $\mathbf{R}_e$  is the 3 by 3 rotation matrix that aligns the epipolar vector  $\mathbf{e}'$  with the X axis, and  $k$  an integer describing the camera periodicity translation.

We can now apply the assumption that articulated limb motion is planar, and proceed to fit the 3D limb points to two planes. Since we have aligned the epipolar vector with the X axis, one such point that must lie on each of the workspace planes is the ideal point  $(1, 0, 0, 0)^\top$ . The pencil of planes that intersect this ideal point have the form  $\pi = (0, v_2, v_3, v_4)^\top$ , hence the problem reduces to that of finding two lines within the YZ plane cross section data. Both workspace limb swing planes are approximately parallel. Due to experimental noise the cross section data points are best fitted to two lines with a common normal.

We evaluate the mean  $(\bar{y}, \bar{z})^\top$  of the cross section point distribution and apply a translation  $\mathbf{H}_t$  that maps this point to the origin. The two cross section plane lines  $\mathbf{l}_1$  and  $\mathbf{l}_2$ , computed by orthogonal regression, are then aligned parallel with the Y axis by applying the rotation  $\mathbf{H}_r$ . The transformed lines  $\mathbf{l}'_i = \mathbf{H}_r^{-\top} \mathbf{l}_i$  can be re-normalized such that  $\mathbf{l}'_1 = (0, 1, -c_1)^\top$ ,  $\mathbf{l}'_2 = (0, 1, -c_2)^\top$  in order that we can find the points at which they cut the Z axis  $(c_1, c_2)$ . We then apply a further similarity transform  $\mathbf{H}_s$  that translates the

mid point  $(c_2 + c_1)/2$  to the origin and scales in the Z direction to rectify the lines to the form  $\mathbf{l} = (0, 1, \pm 1)^\top$ . Application of a plane selection transform  $\mathbf{H}_\beta$  translates by  $\pm 1$  mapping the selected set of points onto the  $z = 0$  plane.

$$\mathbf{H}_v = \mathbf{H}_\beta \mathbf{H}_s \mathbf{H}_r \mathbf{H}_t \quad (1)$$

The projection transform mapping the computed  $z = 0$  plane points  $\mathbf{W}$  to normalized image points  $\hat{\mathbf{x}}$  is given by.

$$\hat{\mathbf{x}}(k) = [\mathbf{R}_e^\top \mid -k \cdot \mathbf{e}'] \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{H}_v^{-1} \end{pmatrix} \mathbf{W} \quad (2)$$

The augmented 4 by 4 matrix inverse  $\tilde{\mathbf{H}}_v^{-1}$  has the form:

$$\left[ \begin{array}{c|ccc} 1 & \mathbf{m}_2 & \mathbf{m}_3 & \mathbf{m}_4 \\ \hline \mathbf{0} & 0 & 0 & 1 \end{array} \right] \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -\beta \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (3)$$

The corresponding projection transformation of workspace points  $\mathbf{W} = (u, v, 0, w)^\top$  into the image is then given by the 3 by 3 homography mapping  $\hat{\mathbf{x}} = \mathbf{H}_p \cdot (u, v, w)^\top$ .

$$\mathbf{H}_p = [\mathbf{e}' \quad \mathbf{m}'_2 \quad (\mathbf{m}'_4 - \beta \cdot \mathbf{m}'_3 - k \cdot \mathbf{e}')] \quad (4)$$

where  $\mathbf{m}'_i = \mathbf{R}_e^\top \mathbf{m}_i$  and  $\mathbf{e}' = \mathbf{R}_e^\top \cdot (1, 0, 0)^\top$

We finally find both sets of optimal  $z = 0$  plane points by solution of the D.L.T.  $([\hat{\mathbf{x}}_{k,\beta}]_\times \mathbf{H}_p(k, \beta)) \cdot \hat{\mathbf{u}} = \mathbf{0}$ , for each point  $\hat{\mathbf{u}}$  in order to minimize image reprojection error.

## 2.2. Recovering Metric Structure

Structure on the  $z = 0$  plane has been recovered up to an affine ambiguity. We need to find the affine transform  $\mathbf{H}_a$  that maps the imaged circular points  $(1, \mu \pm i \cdot \lambda, 0)^\top$  back to their canonical positions  $(1, \pm i, 0)^\top$ .

$$\mathbf{H}_\mu = \begin{pmatrix} 1 & 0 & 0 \\ -\mu/\lambda & 1/\lambda & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (5)$$

We can recover metric structure on the plane by using the known ratios of lengths [5] between articulated limb landmark points over all reconstructed frame poses. The skeletal structure is rigid hence the length ratio of a limb segment should remain fixed (unity) over all frames. The squared distance between any two limb segment endpoints  $\mathbf{x}_0, \mathbf{x}_1$  can be written  $d^2 = \Delta \mathbf{x}^\top \Delta \mathbf{x}$ , where  $\Delta \mathbf{x} = (u_1 - u_0, v_1 - v_0)^\top$ . If  $\Delta \mathbf{x}_1$  and  $\Delta \mathbf{x}_2$  are the pose difference vectors for a corresponding limb segment at two different frames then the equal limb length constraint can be written.

$$\Delta \mathbf{x}_1^\top \mathbf{H}^\top \mathbf{H} \Delta \mathbf{x}_1 = \Delta \mathbf{x}_2^\top \mathbf{H}^\top \mathbf{H} \Delta \mathbf{x}_2 \quad (6)$$

Writing  $\Delta \mathbf{x}_i = (\delta x_i, \delta y_i)^\top$  and elements of the symmetric matrix  $\mathbf{M} = \mathbf{H}^\top \mathbf{H}$  as  $\mathbf{m} = (\mathbf{M}_{11}, \mathbf{M}_{12}, \mathbf{M}_{22})^\top$ .

$$\begin{bmatrix} \delta x_1^2 - \delta x_2^2 & 2(\delta x_1^2 \delta y_1^2 - \delta x_2^2 \delta y_2^2) & \delta y_1^2 - \delta y_2^2 \end{bmatrix} \mathbf{m} = \mathbf{0} \quad (7)$$

Since  $\mathbf{m}$  is defined up to scale then a minimum of two such corresponding pose constraints are required. We stack all constraints formed from all sets of combinations of same limb frame poses on each swing plane. The rectification matrix  $\mathbf{H}_\mu$  is formed from the extracted parameters of  $\mathbf{H}^\top \mathbf{H}$ , where  $\mu = -\frac{m_2}{m_3}$  and  $\lambda = \sqrt{\frac{m_1}{m_3} - \mu^2}$ . The ideal epipole  $(1, 0, 0)^\top$  is mapped by  $\mathbf{H}_\mu$  to  $(1, -\mu/\lambda, 0)^\top$  so we must also apply a rotation  $\mathbf{H}_r$  to align the epipole back along the X axis such that  $\mathbf{H}_a = \mathbf{H}_r \mathbf{H}_\mu$  is the affine transform that recovers metric angles and length ratios on both planes. Points on the metric plane  $\hat{\mathbf{w}}$  are then mapped into the image as:

$$\hat{\mathbf{x}} = \mathbf{H}_p \mathbf{H}_a^{-1} (\mathbf{H}_a \hat{\mathbf{u}}) = \tilde{\mathbf{H}} \hat{\mathbf{w}} \quad (8)$$

We are only interested in limb length ratios, so scalings are applied to both planes in order to transform each first limb segment to unit length. In practice many data points may be missing due to occlusion. Even in the ideal case where motion is frontoparallel the hip point on the occluded side of the body may never be imaged. To robustly compute the scaling transforms we first compute  $\mathbf{H}_\tau$ , the scaling between both swing planes. We evaluate the mean set of limb lengths for both planes  $\mathbf{d}, \mathbf{d}'$ . These lengths are related by the inter-plane scaling:  $\mathbf{d}_i = \tau \cdot \mathbf{d}'_i$ . A minimal solution to this trivial set of linear equations requires at least one valid length correspondence within the set of limb segments. With  $\mathbf{H}_\tau$  now known we find the optimal first limb segment length  $\hat{\mathbf{d}}_1$  on the first plane. We then compute the scaling transform  $\mathbf{H}_s$  that maps  $\hat{\mathbf{d}}_1$  to the unit length and update both sets of points and projection homographies.

$$\mathbf{H}_1 = \tilde{\mathbf{H}}_1 \mathbf{H}_s^{-1} = \begin{bmatrix} \mathbf{p}_1/s & \mathbf{p}_2/s & \mathbf{p}_3 \end{bmatrix} \quad (9)$$

$$\begin{aligned} \tilde{\mathbf{H}}_2 \mathbf{H}_\tau^{-1} \mathbf{H}_s^{-1} &= \begin{bmatrix} \mathbf{p}_1/s \cdot \tau & \mathbf{p}_2/s \cdot \tau & \mathbf{p}_3' \end{bmatrix} \\ \mathbf{H}_2 &= \begin{bmatrix} \mathbf{p}_1/s & \mathbf{p}_2/s & \tau \cdot \mathbf{p}_3' \end{bmatrix} \end{aligned} \quad (10)$$

where  $\mathbf{p}_i$  are the column vectors of the swing plane transforms  $\mathbf{H}_p \mathbf{H}_a^{-1}$  with  $k = 0$  and  $\beta = \mp 1$  within eqn. 4.

### 2.3. Recovering Gait Dynamics

Recompute the true metric structure  $\mathbf{w}_i$  from the real normalized image points  $\mathbf{x}'_i$  by applying the inverse mappings  $\mathbf{w}_i = \mathbf{H}_1^{-1} \mathbf{x}'_i$  and  $\mathbf{w}'_i = \mathbf{H}_2^{-1} \mathbf{x}'_i$ . Resolve the four-fold X,Y

reflection ambiguity of the metric plane by consideration of the gross spatiotemporal motion structure. Two smoothed data vectors  $\tilde{\mathbf{u}}, \tilde{\mathbf{u}}'$ , generated from the mean X coordinate positions of limb points over a centred 3 frame window, are computed and fitted to a linear velocity model with a pair of simultaneous equations:  $\tilde{u}_i = v_x \cdot i + u_0$  and  $\tilde{u}'_i = v_x \cdot i + u'_0$ .

We choose to normalize gait sequences to emulate a left to right walk, so ensure that  $v_x$  is positive by applying a reflection about the Y axis and update both points  $\mathbf{w}_i, \mathbf{w}'_i$  and homographies. The reflection about the X axis, to ensure that the sky is upward, is determined from the Y coordinate ordering (*hip*  $\rightarrow$  *knee*  $\rightarrow$  *ankle*) of the means of each limb point over all frames. The only remaining ambiguity is the translation between both sets of plane points.

Normal gait is bilaterally symmetric with a half phase shift. For each limb segment compute, both plane limb angle sets and their corresponding time sample vectors. We concatenate the angle vectors  $\mathbf{A} = (\mathbf{a}^\top, \mathbf{a}'^\top)^\top$  and time sample vectors  $\mathbf{S} = (\mathbf{t}^\top, \mathbf{t}'^\top + \frac{1}{2}\mathbf{T}^\top)^\top$  then determine, with fixed fundamental frequency  $f_0$ , the Fourier series representation of the limb angle function. With the knowledge of the normalized limb lengths  $\mathbf{D}$  we can find by back substitution both sets of origin limb points  $\mathbf{o}, \mathbf{o}'$ . We then compute two vectors of smoothed X origin limb data generated from the mean positions over a centred 3 frame window, and fit the linear velocity model to the pair of simultaneous equations in  $t$ . This gives a reasonable estimate of the linear velocity component and initial X offset points  $(u_0, u'_0)$  of gait on the metric plane. We now compute a partitioned bilateral Fourier series representation of the origin point displacement function with sample data  $\mathbf{o}, \mathbf{o}'$ .

$$\begin{aligned} u(t) &= v_x t + u_0 + \sum_{k=1}^n A_k \cos(2\pi k f_0 t + \phi_k) \\ u'(t) &= v_x t + u'_0 + \sum_{k=1}^n A_k \cos(2\pi k f_0 (t + \frac{1}{2}T) + \phi_k) \end{aligned} \quad (11)$$

We make an initial first harmonic approximation by partitioning the parameter vector  $\mathbf{P}_1 = (v_x, A_1, \phi_1 \mid u_0, u'_0)^\top$ . The computed estimates of  $\mathbf{P}_1$  are then used to bootstrap the full partitioned parameterization.

$$\mathbf{P} = (v_x, A_1, \phi_1, \dots, A_n, \phi_n \mid u_0, u'_0)^\top \quad (12)$$

The Y component origin limb point displacement function is similar, though  $v_y$  is held fixed (zero). Both are computed using a partitioned Levenberg-Marquardt algorithm with fixed fundamental frequency  $f_0$ . The translations  $\mathbf{H}_o, \mathbf{H}'_o$  then map the starting origin limb point displacements  $(u_0, v_0)^\top, (u'_0, v'_0)^\top$  to the origin. We finally apply the inverse normalization transform to the updated homography mappings  $\mathbf{K}_n^{-1} \mathbf{H}'_i$ . The homography projection functions then map metric plane points into the image as:

$$\begin{aligned} \mathbf{x}_1(t) &= [\mathbf{h}_1 \quad \mathbf{h}_2 \quad \mathbf{h}_3] g(t : f_0, \mathbf{D}, \mathbf{X}, \mathbf{Y}, \mathbf{V}) \quad (13) \\ \mathbf{x}_2(t) &= [\mathbf{h}_1 \quad \mathbf{h}_2 \quad \mathbf{h}'_3] g(t + T/2 : f_0, \mathbf{D}, \mathbf{X}, \mathbf{Y}, \mathbf{V}) \end{aligned}$$

where  $g(t)$  is the bilateral Fourier series function,  $\mathbf{X}, \mathbf{Y}$  are the velocity and Fourier coefficients of the metric plane origin limb displacement functions and  $\mathbf{V}$  the Fourier coefficients of the set of limb pose angle functions.

### 3. MAXIMUM LIKELIHOOD ESTIMATION

We can better model a limb plane pose by first applying a rotation  $\mathbf{H}_\alpha$  about the X axis to facilitate the swing plane inclination to the vertical, then apply the plane selection translation  $\mathbf{H}_\beta$  to map the required hip point to  $\mp 1$ . This is followed by a scaling  $\mathbf{H}_\tau$  in the Z direction that generates the correct distance between both hip points. Initially  $\alpha = 0$  and the set of limb plane pose projections  $\mathbf{H}$  have the form

$$\begin{aligned} \mathbf{H}_1 &= [\mathbf{m}_1 \quad (c_\alpha \cdot \mathbf{m}_2 - s_\alpha \cdot \mathbf{m}_3) \quad (\mathbf{m}_4 - \mathbf{m}_3)] \\ \mathbf{H}_2 &= [\mathbf{m}_1 \quad (c_\alpha \cdot \mathbf{m}_2 + s_\alpha \cdot \mathbf{m}_3) \quad (\mathbf{m}_4 + \mathbf{m}_3)] \end{aligned} \quad (14)$$

where  $c_\alpha = \cos \alpha$ ,  $s_\alpha = \sin \alpha$ ,  $\mathbf{m}_1 = \mathbf{h}_1$ ,  $\mathbf{m}_2 = \mathbf{h}_2$ ,  $\mathbf{m}_3 = \frac{1}{2}(\mathbf{h}'_3 - \mathbf{h}_3)$  and  $\mathbf{m}_4 = \frac{1}{2}(\mathbf{h}'_3 + \mathbf{h}_3)$  from eqn. 13.

As a final optimization step we perform a bundle adjustment procedure that minimizes reprojection error with respect to all parameters  $\mathbf{P}$  of the gait projection function. Good reconstruction results have been achieved with  $n = 5$  Fourier coefficients and cameras with frame rates  $\geq 25$  frames per second. Writing  $\mathbf{M} = (\mathbf{m}_1^\top, \mathbf{m}_2^\top, \mathbf{m}_3^\top, \mathbf{m}_4^\top, \alpha)^\top$

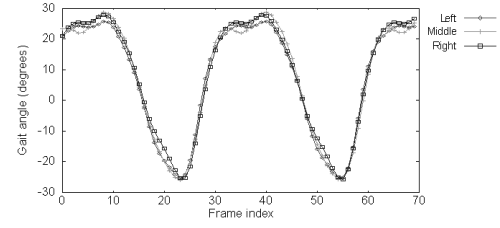
$$\mathbf{P} = (\mathbf{M}^\top, f_0, \mathbf{D}^\top, \mathbf{X}^\top, \mathbf{Y}^\top, \mathbf{V}^\top)^\top \quad (15)$$

### 4. RECONSTRUCTION RESULTS

A synchronized, uncalibrated three camera system (30 fps) is set up, each with a different type of lens. Four subjects, whom have retro-reflective marker balls attached to the principle joint features of the skeletal system, are told to walk along a linear trajectory through the field of view of all cameras. The acquired image sequences are manually marked and the reconstruction process performed. Since image acquisition is simultaneous and gait dynamics consistent in all three views the only parameters that differ between image sequences are the camera intrinsic and extrinsic parameters.

### 5. CONCLUSIONS

We have demonstrated that gait has sufficient properties that allows us to exploit the structure of articulated motion in or-



**Fig. 4.** Reconstruction of upper leg gait angles for subject 1. The corresponding unrectified angles are shown in fig. 2.

	<i>L</i>	<i>M</i>	<i>R</i>		<i>L</i>	<i>M</i>	<i>R</i>
1	7.48	6.17	9.03	1	1.63	2.54	2.39
2	8.54	5.48	8.37	2	1.69	2.81	2.47
3	8.79	9.89	9.89	3	2.37	2.36	2.54
4	3.5	3.71	6.73	4	1.41	1.95	2.15

(a) Stratified

(b) M.L.E.

**Table 1.** Comparison of root mean square pixel reprojection errors for the reconstructed gait motion in the left, middle and right camera views (*L, M, R*).

der to remove the unknown camera and pose ambiguities and reconstruct the underlying gait signature. In the context of biometric gait analysis this is a positive step toward making already established and future techniques more robust to changes in pose. Further analysis needs to be done to validate its potential usefulness for identification at a distance.

### 6. REFERENCES

- [1] C. Y. Yam, M. S. Nixon, and J. N. Carter, "Extended model based automatic gait recognition of walking and running," in *3rd Proc. AVBPA*, June 2001, pp. 278–283.
- [2] N. M. Spencer and J. N. Carter, "Viewpoint invariance in automatic gait recognition," in *Proc. Third IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'02)*, 2002, pp. 1–6.
- [3] A. Kale, R. Chowdhury, and R. Chellappa, "Towards a view invariant gait recognition algorithm," in *Proc. IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS'03)*, 2003, pp. 143–150.
- [4] D. Liebowitz, *Camera calibration and reconstruction of geometry from images*, Ph.D. thesis, University of Oxford, Dept. Eng. Science, June 2001, D.Phil. thesis.
- [5] D. Liebowitz and A. Zisserman, "Metric rectification for perspective images of planes," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Santa Barbara, CA, U.S.A., June 1998, pp. 482–488.