MOTION CONSTRAINTS FOR VIDEO MOSAICING

Pulkit Parikh, C.V. Jawahar
Center for Visual Information Technology(CVIT)
Internation Institute of Information Technology
Gachibowli, Hyderabad - 500032,India
pulkitparikh@yahoo.com, jawahar@iiit.ac.in

Keywords: Mosaicing, Homography, Video, Motion Model

Abstract

This paper derives a set of constraints relating the homographies in a video sequence. They are derived by assuming the camera motion models. We describe show how they can be used effectively for improving homography-estimation for mosaicing.

1 Introduction

Mosaicing is the process of generating a single, large, integrated image by combining the visual clues from multiple images [4, 8]. Often, this results in larger field of view. In recent years, mosaicing has become an integral part of many video applications. Video mosaics find applications in video compression, virtual environments, panoramic photography, etc. Video capture technology is now available at low prices with poor field of view. Video mosaicing is emerging as a convenient way to capture images without limitations on field of view.

Mosaicing primarily involves image registration and image blending. Accurate registration of constituent images is the key in obtaining geometrically correct mosaics. Primitive image registration algorithms assume the images to be related by a similarity transformation (i.e., translation, rotation and scaling). They estimate the transformation parameters by spatial correlation or transform domain techniques [1, 10, 11]. Subsequently, a six-parameter affine transformation model became popular. Now researchers employ a full projective transformation (*homography*) with 8 free parameters for image alignment. A comprehensive survey of registration techniques can be found in [1, 10, 11]. In most mosaicing schemes, matching is done by numerical computation of the transformation relating the two images. The images are then blended to obtain the mosaic. Conventional video mosaicing algorithms extend the image mosaicing procedure to video by registering and stitching successive frames [4, 8].

Recently, there is a growing interest in developing algorithms for robust global alignment [6, 7, 8, 9]. Global registration (or alignment) refers to the alignment of video frames taking into account the overlapping frames, and not just the consecutive ones. Typically, it is done in two steps. First, homographies are computed for all pairs of images with sufficient overlap. In the second step, a graph based representation is built from this. The problem is then cast as the identification of an optimal path in the graph. In [7], bundle adjustment is employed further to optimize the graph-based solution. However, optimality in the true sense is not obtained in this framework as it is not demanded for the pair-wise homographies in the global context.

In this paper, we derive a set of constraints on homographies relating frames in a video, under constrained motion models. These homographies are shown to be related and this relationship is shown to be an important clue in building mosaics, even if the computation of homography is unreliable for a specific pair.

2 Video Mosaicing

Let H_{12} be the homography relating two different frames 1 and 2, belonging to a planar object. From the fundamentals of multiple view geometry [3], it is well known that

$$X_2 = H_{12} X_1$$

where X_1 and X_2 are the corresponding points in frames 1 and 2, and expressed in homogeneous coordinates as $[x, y, 1]^T$.

If there are N frames with indices $0, 1, \ldots, i, i+1, \ldots, N-1$, then the mosaic could be built

- by mapping all the frames to a reference frame (say 0) with the help of a set of *reference* homographies $H_{i,0}$. Reference frame could be the first, last or any of the intermediate ones.
- by computing all *incremental* (frame-to-next-frame) homographies $H_{i,i+1}$ and imposing a single view point constraint in an appropriate manner.

The quality of mosaics heavily depends on the accuracy in the registration (more specifically, the accuracy of the estimation of the transformation) of images. This often depends on: (i) Accuracy with which one can estimate the features (say interest points) from the images. (ii) Performance of the correspondence computation scheme. (iii) Accuracy of the numerical procedure for robust computation of the homography from the (noisy) correspondences. Though, homography can be computed from eight point correspondences, a larger set is preferred for robustness. Kanatani and Ohta [5] employs a renormalization technique for improving the homography computation. To take care of the outliers in the correspondence computation, RANSAC [2] has been proven to be a good choice.

It is worth noting that in most situations, only consecutive frames are registered for mosaicing. If one of the frame is of poor quality, the entire mosaic can become highly inaccurate. This inaccuracy can also get propagated for all frames succeeding this frame.

In this paper, we argue that, there exists relationships connecting homographies under certain camera motion models. This can help in robustly estimating the ensemble of homographies. Before deriving these relationships, we derive an expression for the homography relating two arbitrary views of a planar object.

3 Homography Relating Two Arbitrary Views of a Planar Object

We know from [3] that if $K[I|0]$ and $K[R|T]$ are the first and second camera matrices and $\pi = (n^T, d)^T$ represents the world plane (such that, for points P_i on the plane, $n^T P + d = 0$), the homography induced by the plane is given by:

$$H = K\big(R - \frac{Tn^T}{d}\big)K^{-1} \tag{1}$$

However, for the video mosaicing applications, where the two frames are captured from arbitrary poses, we need a generalized relationship. Let $K[R_0|T_0]$ and $K[R|T]$ be the first and second camera matrices and $\pi = (n^T, d)^T$ represent the plane.

Now, to able to use Equation 1, we need to apply a rigid transformation on the world coordinate system such that the first camera pose changes from $[R_0|T_0]$ to $[I|0]$.

After the coordinate transformation, if $[R_n|T_n]$ is the pose of the second camera, R_n is given by RR_0^{-1} and T_n is given by $T - RR_0^{-1}T_0$. This comes from the fundamentals of rigid transformation as shown below.

If P is a point in the original world coordinate system and P_c is the same point in the coordinate system of the second camera $[R|T]$,

$$P_c = RP + T \tag{2}$$

After coordinate transformation, P is transformed to $R_0P + T_0$ and $[R|T]$ gets modified as $[R_n|T_n]$, the new pose of the second camera. Therefore,

$$P_c = R_n(R_0P + T_0) + Tn \tag{3}$$

Simplifing,

$$P_c = \underbrace{R_nR_0P}_{R} + \underbrace{R_nT_0 + Tn}_{T} \tag{4}$$

From equation 2 and 4 after simplification, we get $R_n = RR_0^{-1}$ and $T_n = T - RR_0^{-1}T_0$.

Similarly, the plane parameters also change after the coordinate transformation. Let $(n_n^T, d_n)^T$ be the new parameters. Using them, we have

$$n_n^T(R_0P + T_0) + d1 = 0$$

or

$$\underbrace{(n_n^TR_0)}_{n^T}P + \underbrace{n_n^TT_0 + d1}_{d} = 0$$

Equating 3 with the original plane equation $n^TP + d = 0$, after simplification, we get $n_n^T = n^TR_0^{-1}$ and $d_n = d - n^TR_0^{-1}T_0$.

Substituting R_n,T_n,n_n^T and d_n in Equation 1, we arrive at

$$H = K\left[RR_0^{-1} - (T - RR_0^{-1}T_0)\frac{n^TR_0^{-1}}{d - n^TR_0^{-1}T_0}\right]K^{-1} \tag{5}$$

It may be seen that when the first camera is at $[I|0]$, the above equation reduces to Equation 1.

If the camera motion is only translational and both rotation matrices R and R_0 are I (identity), from Equation 5,

$$H = I - K\left[(T - T_0)\frac{n^T}{d - n^TT_0}\right]K^{-1} \tag{6}$$

4 Homographies under Uniform Velocity Motion

Before considering the general linear motion, we derive constraints on the incremental homographies $H_{i,i+1}$ and the inverse reference homographies $H_{0,i}$, for a purely translational motion. We show that there exist 11 parameter and 9 parameter relationships for the homographies in the case of linear translational motion.

4.1 Incremental Homographies

Theorem 1 *In presence of uniform translational motion, all the incremental homographies $H_{i,i+1}$ are related by a 11 parameter model.*

$$H_{i,i+1} = I + \frac{C}{c_1 + i.c_2} \quad (7)$$

where, C is a 3×3 matrix and c_1 and c_2 are scalars.

Proof 1 *For a purely translational motion, the camera moving with velocity V (and acceleration A) is equivalent to its translational pose T moving with velocity $-V$ (and acceleration $-A$).*
If T_i is the translational pose of the camera for frame I_i, assuming that the time elapsed between the capturing of successive frames is constant (t), we have

$$T_i = T_0 - i.t.V \quad (8)$$

and

$$T_{i+1} - T_i = -t.V$$

Substituting the above two equations in Equation 6,

$$\begin{aligned} H_{i,i+1} &= I - K\left[(-t.V)\frac{n^T}{d - n^T(T_0 - i.t.V)}\right]K^{-1} \\ &= I + \frac{t.KVn^TK^{-1}}{d - n^TT_0 + i(t.n^TV)} \end{aligned}$$

Substituting $C = t.KVn^TK^{-1}$, $c_1 = d - n^TT_0$ and $c_2 = t.n^TV$ in the above equation, directly results in Equation 7.

Corollary 1 *If the camera motion is parallel to the world plane, all the incremental homographies are the same.*

Proof comes from the fact that, in presence of linear motion parallel to the world plane, $n^TV = 0$ leading to c_2 becoming zero in Equation 7 and the incremental homographies becoming identical.

4.2 Inverse Reference Homographies

Theorem 2 *In presence of uniform translational motion, all the inverse reference homographies $H_{0,i}$ are related by a 9 parameter model.*

$$H_{0,i} = I + iK_c$$

where, K_c is a 3×3 matrix.

Proof 2 *Substituting Equation 8 in Equation 6,*

$$H_{0,i} = I - K\left[-i.t.V\frac{n^T}{d^*}\right]K^{-1} \quad (9)$$

where, $d^* = d - n^TT_0$.

After simplification,

$$H_{0,i} = I + i\left[\underbrace{t.KV\frac{n^T}{d^*}K^{-1}}_{K_c}\right]$$

Corollary 2 *If the camera moves horizontally, imaging objects lying on a horizontal plane, the first order difference between the normalized inverse reference homographies is constant. By normalization, we mean that H(3,3) is set to 1.*

Proof 3 *After normalization of the homographies (i.e.,after H(3,3) is set to unity),*

$$H_{0,i} = \frac{I + iK_c}{1 + iK_c(3,3)} \quad (10)$$

As the camera moves horizontally, the z component of V is zero. Therefore, V is of the form $[x \;\; x \;\; 0]^T$. Moreover, the world plane is horizontal so vector n is of the form $[0 \;\; 0 \;\; x]^T$. We also know that the calibration matrix K is affine and so is its inverse.

Now, from Equation 2, K_c is a multiplication of three matrices - K, Vn^T and K^{-1} and two scalars. From that, it directly follows that $K_c(3,3)$ is zero. Substituting that in 10, after simplification,

$\Delta H_i = K_c = constant$

where, the first order difference $\Delta H_i = H_{0,i+1} - H_{0,i}$.

Note that when the camera moves parallel to an arbitrary plane and images objects on that plane, this corollary holds only when the camera's Z axis (which is usually the viewing axis) is perpendicular to the plane.

4.3 General Linear Motion

In the previous subsections, we derived constraints on incremental and inverse reference homographies for translational motion. In the case of general linear motion, we need to consider rotational velocity as well. We express the rotation in terms of the three Euler angles ϕ, θ and ψ represented in the vector form as L. The rotational velocity is denoted by V_r. L_0 denotes the initial configuration of Euler angels. Rewriting Equation 5 for the inverse reference homography of frame I_i,

$$H_{0,i} = K\left[R_i R_0^{-1} - (T_i - R_i R_0^{-1} T_0)\frac{n^T R_0^{-1}}{d - n^T R_0^{-1} T_0}\right] K^{-1}$$

where,

$$\begin{array}{rcll} K &=& f_K(\alpha,\beta,\vartheta,u_0,v_0) & \text{- 5 parameters} \\ R_i &=& f_R(L,V_r,i) & \text{- 6 parameters} \\ R_0 &=& f_r(L) & \\ T_i &=& f_T(T_0,V,i) & \text{- 6 parameters} \\ \pi &=& (n^T,d)^T & \text{- } \underbrace{\text{4 parameters}}_{\text{21 parameters}} \end{array}$$

Thus, all inverse reference homographies are related by a 21 parameter model. In the same way, there exists a 21 parameter model for incremental homographies as well. The important point to note is that for general linear motion, the relationships or the constraints are non-linear whereas, for a purely translational case, the constraints derived are linear in the model parameters. Since the purely translational case is a special case of the general linear motion, there also exists a non-linear relationship between homographies for that case. However, linear relationships are always preferred as they can be exploited more robustly for video mosaicing.

It can be easily deduced from the results given in this section that under general linear motion, the homography between any pair of frames I_i and I_j can be computed from a 21 parameter model.

5 Homographies under Uniform Acceleration Motion

Here again, we derive constraints on the incremental homographies and the inverse reference homographies, for a purely translational motion and then discuss the general acceleration motion.

5.1 Incremental Homographies

Theorem 3 *In presence of uniform acceleration motion, all the incremental homographies $H_{i,i+1}$ are related by a 21 parameter model.*

$$H_{i,i+1} = I - \left[\frac{E_1 + i.E_2}{b_1 + i.b_2 + i^2.b_3}\right] \tag{11}$$

where, E_1,E_2 are 3×3 matrices and b_1,b_2 and b_3 are scalars.

Proof 4 *If V is the velocity of the camera and A is the acceleration of the camera,*

$$T_i = T_0 - V.(i.t) - \frac{1}{2}.A(i.t)^2 \tag{12}$$

and

$$T_{i+1} - T_i = \underbrace{-(t.V + \frac{t^2}{2}.A)}_{D_1} + i.\underbrace{(-t^2.A)}_{D_2}$$

After simplification, we also have

$$d - n^T T_i = \underbrace{d - n^T T_0}_{b_1} + i.\underbrace{(t.n^T V)}_{b_2} + i^2\underbrace{(\frac{t^2}{2}.n^T A)}_{b_3}$$

Substituting $T_{i+1} - T_i$ and $d - n^T T_i$ in Equation 6,

$$\begin{aligned} H_{i,i+1} &= I - \left[\frac{K(D_1 + i.D_2)n^T K^{-1}}{b_1 + i.b_2 + i^2.b_3}\right] \\ &= I - \left[\frac{(KD_1 n^T K^{-1}) + i.(KD_2 n^T K^{-1})}{b_1 + i.b_2 + i^2.b_3}\right] \end{aligned}$$

Setting $E_1 = KD_1 n^T K^{-1}$ and $E_2 = KD_2 n^T K^{-1}$, we arrive at Equation 11.

Corollary 3 *If the camera moves horizontally, imaging objects lying on a horizontal plane, the first order difference between the normalized incremental homographies is constant.*

Proof 5 *In presence of acceleration motion parallel to the world plane, $n^T V = 0$ and $n^T A = 0$. Using them, the normalized incremental homographies are given by:*

$$H_{i,i+1} = \frac{I - E_1 - i.E_2}{1 - E_1(3,3) - i.E_2(3,3)} \tag{13}$$

As the camera moves horizontally and the world plane is horizontal, $E_1(3,3)$ is zero. Similarly, $E_2(3,3)$ can also be shown to be zero. Substituting them in 13, after simplification,

$$\Delta H_i = E_2 = constant$$

where, the first order difference $\Delta H_i = H_{i+1,i+2} - H_{i,i+1}$.

5.2 Inverse Reference Homographies

Theorem 4 *In presence of uniform acceleration motion, all the inverse reference homographies $H_{i,i+1}$ are related by a 18 parameter model.*

$$H_{0,i} = I + i.M_1 + i^2.M_2$$

where, M_1,M_2 are 3×3 matrices.

Proof 6 *Substituting Equation 12 in Equation 6,*

$$H_{0,i} = I - K\left[(i.(-t.V) + i^2.(-\frac{t^2}{2}.A)).\frac{n^T}{d}\right]K^{-1}$$

where,

$$d^* = d - n^T T_0$$

Simplifying,

$$H_{0,i} = I + i\underbrace{\left[\frac{t.KVn^TK^{-1}}{d^*}\right]}_{M_1} + i^2\underbrace{\left[\frac{t^2.KAn^TK^{-1}}{2.d^*}\right]}_{M_2}$$

Corollary 4 *If the camera moves horizontally, imaging objects lying on a horizontal plane, the second order difference between the normalized inverse reference homographies is constant.*

Proof 7 *After normalization,*

$$H_{0,i} = \frac{I + i.M_1 + i^2.M_2}{1 + i.M_1(3,3) + i^2.M_2(3,3)}$$

As the camera motion is horizontal and the world plane is also horizontal, $M_1(3,3)$ and $M_2(3,3)$ are zero. Using that, the first order differences

$$\Delta H_i = (M_1 + M_2) + i(2M_2)$$

and the second order differences

$$\Delta H_{i+1} - \Delta H_i = 2M_2 = \textit{constant.}$$

5.3 General Acceleration Motion

Under general uniform acceleration motion comprising rotation and translation, all inverse reference homographies are related by a 27 parameter model. Similarly, there is a 27 parameter model relating incremental homographies as well.

The proof is identical to the general linear motion case except that 3 additional parameters come in for rotational acceleration and 3 for translational acceleration. This constraint, like the one for the general linear motion, is non-linear. However, it can be made linear by introducing more parameters which are functions of one or more these 27 parameters. This is helpful when we want to use a linear technique to compute the model parameters in closed form.

6 Estimating Homographies using Motion Constraints

In this section, we discuss how the relationships between homographies, derived in the previous section, can be used to estimate homographies. The conventional techniques use the point correspondence between successive frames to individually compute the incremental homographies. However, we argue that these homographies are not independent of each other and the relationship between them can be exploited to achieve robustness.

Recollecting from the previous sections, all the incremental homographies can be expressed in terms of a fixed parameter model. In other words, given these parameters, the series of incremental homographies can be trivially computed from them. Thus, the problem of robust estimation of the ensemble of homographies boils down to accurate computation of the model parameters. For that, we describe a simple method which uses correspondences from all consecutive pairs of frames pooled in. From equation 7, for uniform translational motion,

$$X_{i+1}^j = \left[I + \frac{C}{c_1 + i.c_2}\right]X_i^j$$

where, $i \in 0, 1, \ldots, N-2$ denotes frame number and $j \in 0, 1, \ldots, m-1$ is the feature point index.

It can be seen from the above equation that each corresponding pair of points gives two linearly independent equations in the 9 parameters of C and c_1, c_2. In summary, we have $m(N-1)$ correspondences, giving us $2m(N-1)$ equations. Thus, what we have is an over-determined linear system of equations in the model parameters which can be solved using Singular Value Decomposition (SVD). The same ideas can be extended for general uniform motion and acceleration motion. Direct estimation of inverse reference homographies follows the same structure.

6.1 Results and Discussions

We have verified the constraints derived in this paper by thorough experimentation with camera motions of different velocities and accelerations. We have also implemented and tested the linear approach mentioned in section 6 to estimate homographies using these constraints. We also estimate homographies using the conventional technique to facilitate comparisons. We use RANSAC to eliminate outlier pairs of corresponding points and then, use SVD to compute the 8 homography parameters from the inlier set of correspondence. We add uniform noise to the image points to test the sensitivity of our approach. To compare the homographies computed using our approach and the conventional approach, we compute the average re-projection error for the noise-free points.

$$E = \sum_{i,j} \|(X_i^j - \Pi(H_{0,i}X_0^j))\|^2$$

In Figure 1, we show the plot of the average re-projection error against the level of noise, for the conventional approach and our approach. It is easy to see that for our approach, the error increases at a much slower rate with the level of noise, compared to the conventional approach.

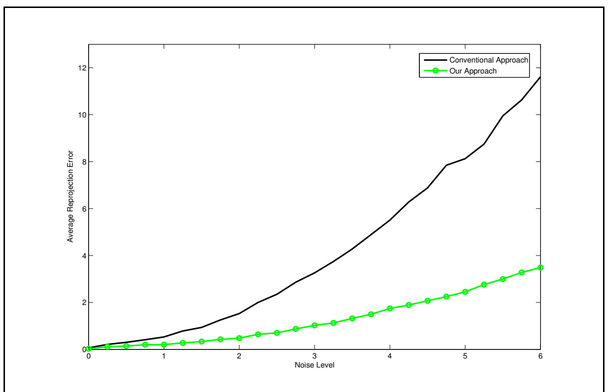

Figure 1: Average Re-projection Error Vs Level of Noise, Solid: Conventional Approach, Circled: Our Approach

For our approach to homography-estimation, even if one of the frames is of very bad quality, the average re-projection error is not affected as much as it does for the conventioal approach, as shown in Table 1.

	Conventional Approach	Our Approach
Average Re-projection Error	5.7898	1.7625

Table 1: Average Re-projection Error when one frame is of poor quality

7 Conclusions

We have derived a set of constraints on incremental and inverse reference homographies when the camera motion follows certain models. We also demonstrate how these constraints can be exploited for homography-estimation. Experiments have been conducted to verify our claims and to illustrate how homographies can be estimated more accurately using our approach.

References

[1] L.G. Brown. A survey of image registration techniques. *ACM Computing Surveys 24*, 24(4):325–376, December 1992.

[2] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. In *RCV87*, pages 726–740, 1987.

[3] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2000.

[4] Michal Irani, P. Anandan, and Steve Hsu. Mosaic based representations of video sequences and their applications. In *ICCV*, pages 605–611, 1995.

[5] K. Kanatani and N. Ohta. Accuracy bounds and optimal computation of homography for image mosaicing applications. In *ICCV99*, pages 73–78, 1999.

[6] Eun-Young Kang, Isaac Cohen, and Gérard G. Medioni. A graph-based global registration for 2d mosaics. In *ICPR*, pages 1257–1260, 2000.

[7] Roberto Marzotto, Andrea Fusiello, and Vittorio Murino. High resolution video mosaicing with global alignment. In *CVPR (1)*, pages 692–698, 2004.

[8] Harpreet S. Sawhney, Steven C. Hsu, and Rakesh Kumar. Robust video mosaicing through topology inference and local to global alignment. In *ECCV (2)*, pages 103–119, 1998.

[9] Heung-Yeung Shum and Richard Szeliski. Systems and experiment paper: Construction of panoramic image mosaics with global and local alignment. *Int. J. Comput. Vision*, 36(2):101–130, 2000.

[10] M. Svedlow, C.D. McGillem, and P.E. Anuta. Image registration: Simlarity measure and processing method comparisons. *AeroSys*, 14(1):141–149, 1978.

[11] B. Zitova and J. Flusser. Image registration methods: a survey. *IVC*, 21(11):977–1000, October 2003.