Image Inpainting using Two-View Epipolar Geometry

Prasanna Rangarajan    Panos Papamichalis    Marc Christensen

Abstract

We take a fresh look at the problem of removing occluders in an image using inpainting. We examine a geometric method that utilizes a second image of the scene from a different viewpoint, to identify the occluded objects. We recover the missing intensities by using the geometric relationship between corresponding points in the two images. The relationship is generally specified by the “epipolar line constraint”, and degenerates to a projective homography under special circumstances. For this reason, we use the plane + parallax approach to estimate the epipolar geometry relating the two images. We fill-in missing pixels by copying information from the respective epipolar lines in the second image.

The success of the proposed framework hinges on the ability to estimate the “epipolar line constraint” from noisy correspondences. To this end, we analyze the uncertainty in estimating a homography from noisy correspondences. We rely on the knowledge of this uncertainty to identify parallax vectors best suited for estimating the epipolar geometry.

We do not make any explicit assumptions about the nature or the extent of camera motion, only requiring that the occluded objects are static and undergo limited perspective change. To illustrate the effectiveness of the proposed approach, we “inpaint” occluders in wide-baseline situations.

1. Introduction

The objective of image inpainting is to produce a modified image where the inpainted region merges seamlessly with the rest of the image, in a manner not obvious to the untrained eye. Applications of image inpainting include occluding object removal, wire/rig removal, retouching of damaged paintings & photographs and error concealment. Traditional image inpainting methods[1, 2, 5, 12] infer missing intensities using basic principles like continuation of edges (level lines), and texture synthesis. The quality of the inpainting depends on the size of the missing region, the geometry of the occluded objects, and the fill-order[1]. Small regions are inpainted effortlessly, while inpainting large regions is difficult and often produces unrealistic results.

The proposed approach extends the scope of current image inpainting methods, by using the epipolar line constraint to infer missing intensities from an additional image (exemplar image). The use of an exemplar image permits us to inpaint large regions, regardless of the geometry of the occluded objects and the fill-order[1].

The use of an exemplar image in inpainting is a recent development. Kang et al. [11] recover the missing pixels by assuming that the two images are affinely related. Jain[9] assumes that the occluded objects are planar in nature, and recovers the missing intensities using a projective homography. Also, there are spatio-temporal techniques that attempt to recover missing pixels using several images. Please refer to [15] for a review of such techniques. In this paper, we focus our attention on image inpainting using a single exemplar image, for arbitrary camera motion.

The novelty of our approach lies in the use of camera and/or occluder motion to reveal the missing intensities. Motion in the occluder either fully or partially reveals the occluded objects in the exemplar image. In the case of static occluders, camera motion induces an apparent motion in the occluders, depending on their depth. As a result, previously occluded objects are revealed fully or partially in the exemplar image. In either situation, the occluded objects can experience significant foreshortening due to camera motion. As this makes the task of inpainting difficult, we require that occluded objects undergo limited perspective change.

From now on, we will refer to the image to be inpainted as view-1 and the exemplar image as view-2.

2. Proposed Approach

The challenge in the proposed inpainting framework, lies in accurately estimating the geometric relationship between views-1 & 2. The relationship is completely specified by a homography in these situations

- occluded object/region belongs to planar facet or
- small baseline and distant occluders or
- small change in camera perspective between views-1 & 2

We inpaint the missing region in view-1, by warping view-2 into view-1, using the inverse homography. In all other situations, the “epipolar line constraint” describes the geometric relationship between corresponding points in views-1 & 2. We inpaint missing pixels in view-1, by copying intensities from the respective epipolar lines in view-2.

In either situation, we need to identify putative point correspondences in the two views. We start by identifying corners with pixel precision using the fast corner detection algorithm[14], and use log-polar normalized cross correlation (LP-NCC) [16] to match the neighborhood of each cor-

\footnote{fill-order refers to the order in which the missing pixels are inpainted.}
The objective is to inpaint parts of the pillar-like structure (transparent red region) in Fig 1(a), using an image of the scene (Fig 1(d)), from a different viewpoint. We inpaint the missing pixels in Fig 1(a) by copying intensities from the respective epipolar lines in Fig 1(e). Epipolar lines sharing the same color in Figs 1(b), 1(e) are in correspondence. The result is shown in Fig 1(f).

The challenge is to accurately estimate the epipolar geometry relating Fig 1(a) and Fig 1(d), from a limited number of noisy point correspondences. The proposed approach relies on the uncertainty in estimating homographies and parallax vectors from noisy point correspondences, to accurately estimate the epipolar geometry.

Figure 1. Image Inpainting using Epipolar Geometry for the Valbonne church scene (courtesy INRIA Robotvis Group)

The success of the proposed inpainting framework hinges on the ability to estimate the epipolar geometry from noisy correspondences. So we start by analyzing the uncertainty in estimating a homography from noisy correspondences, in Sec–3.2. In Sec–3.3.2, we propagate the uncertainty in estimating a homography to parallax vectors. In Sec–4 we examine the use of uncertainty analysis in inpainting, and provide examples of inpainting using the proposed approach in Sec–5.

3. Uncertainty Analysis

Different structures for the uncertainty in a homography, have been proposed by Criminisi[4], Suter[3], Kanatani[10], and Hartley [7]. Unfortunately, these struc-
tures cannot be used directly to find: **parallax vectors and planar facets** that are best suited for estimating epipolar geometry from noisy point correspondences. With this objective in mind, we re-examine the problem of determining the uncertainty in a homography. The utility of the proposed structure of uncertainty lies in the ease with which it can be extended to different normalization schemes, and propagated through projective entities such as **algebraic transfer error, parallax vectors, lines and epipoles.**

### 3.1. Notation

- **boldface** symbols are **vectors**, **underlined** symbols are **homogeneous vectors**, and **CAPITAL** letters represent matrices
- The ordered pair \((x, y)\) represents a point in Euclidian 2-space \((\mathbb{R}^2)\). The 3-vector \(x = (x, y, 1)^T\) is the corresponding point in homogeneous coordinates (Projective 2-space \(\mathbb{P}^2\)).
- \(A(i,:)\) represents the \(i^{th}\) row of the matrix \(A\)
- \(\text{diag}(v)\) is a diagonal matrix with entries of the vector \(v\) along the principal diagonal
- \(I_d\) denotes a \(d \times d\) identity matrix
- \(\text{vec}\)-operator\(^8\) vectorizes a matrix by stacking its columns.
- \(\lceil e_i \rceil^\prime\) denotes the use of skew-symmetric matrices to compute the vector product \(e \times v\)
- \(||v||\) denotes the \(l^2\)-norm of the vector \(v\)

### 3.2. Homography

A homography is a non-singular matrix that describes a linear mapping in homogeneous coordinates \((\mathbb{P}^2)\). If a homography \(H_p\) relates point correspondences \(\{(x_i, y_i), (x'_i, y'_i)\}\), the following bilinear relationships hold

\[
x'_i = \frac{h_{11}x_i + h_{12}y_i + h_{13}}{h_{31}x_i + h_{32}y_i + h_{33}} = x_i', \quad y'_i = \frac{h_{21}x_i + h_{22}y_i + h_{23}}{h_{31}x_i + h_{32}y_i + h_{33}} = y_i'.
\]

where \(h_{ij}\) are entries of the matrix \(H_p\). eq.(1) suggests that only the ratio of the entries of \(H_p\) is significant. As there are only 8 independent ratios amongst the 9 elements of \(H_p\), a homography has 8 degrees of freedom (d.o.f.). This forms the basis for estimating a homography from 4 or more point correspondences using the **Direct Linear Transform algorithm**\(^7\).

#### 3.2.1 Direct Linear Transform (DLT) algorithm

Using **kronecker products** and the **vec operator**\(^8\), we rewrite eq.(1) in the form of the linear system

\[
\begin{bmatrix}
1 & 0 & -x'_i \\
0 & 1 & -y'_i \\
\end{bmatrix}
\begin{bmatrix}
h_{11} \\
h_{12} \\
h_{13} \\
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
0 \\
\end{bmatrix}
\]

\[
A_p h_p = 0. \quad A_p = \begin{bmatrix}
A_{p1}^T & \cdots & A_{pM}^T
\end{bmatrix}^T
\]

In the absence of noise, \(A_p\) is rank-deficient and \(h_p \in \text{nullspace of } A_p\). In the presence of noise \(A_p\) is almost always full rank. In either case, \(h_p\) is estimated as the **right singular vector** corresponding to the **least singular value** of \(A_p^2\). It is easy to observe that \(h_p\) has 8 d.o.f. as \(||h_p|| = 1\).

There is always an uncertainty associated with estimating \(H_p\) using DLT, as the point correspondences used to estimate \(H_p\) are noisy. Typically, noise affects corners in both **views-1 & 2**. But, for the sake of analysis we assume that only corners \(\{(x_i', y_i')\}\) in **view-2** are affected by noise. This assumption is valid only when we seek to find the corner in **view-2** that corresponds to a noisy corner in **view-1**.

#### 3.2.2 Noise model

We assume that the noise \(\epsilon'_i = (\epsilon'_i, \epsilon'_i)^T\) that affects the corners \((x_i', y_i')\) in **view-2** is i.i.d.\(^3\) gaussian with zero mean and covariance \(E[\epsilon'_i \epsilon'_i^T] = \text{diag}(\{\sigma_x^2, \sigma_y^2\})\). This model implies that the error in localizing a corner in the \(x\)-direction is independent of the error in the \(y\)-direction.

Under these assumptions, we examine the effect of estimating a homography from noisy pixel coordinates, using the DLT algorithm.

#### 3.2.3 Perturbation Analysis of the DLT algorithm

In the presence of noise, we estimate the noisy homography \(h_p + \delta h_p\) as the **right singular vector** corresponding to the **least singular value** of the noisy matrix

\[
A_p + N_p = \begin{bmatrix}
1 & 0 & -x'_i + \epsilon'_i \\
0 & 1 & -y'_i + \epsilon'_i \\
0 & 0 & -1
\end{bmatrix} \otimes \Sigma
\]

\[
N_p = \begin{bmatrix}
0 & 0 & -\epsilon'_i \\
0 & 0 & -\epsilon'_i
\end{bmatrix} \otimes \Sigma
\]

For small noise levels\(^4\), the effect of estimating a homography from noisy correspondences is best explained by the **theory of first-order perturbation analysis**\(^\[4,6\]**. Using this theory, we express the perturbation in the homography \(\delta h_p\), and its associated uncertainty \(\Sigma h_p\), in the compact form

\[
\delta h_p = A_p (W \otimes I_2) \Sigma \quad \text{where } \Sigma = \text{diag}(\{\sigma_x^2, \sigma_y^2\})
\]

\[
A_p = E(\delta h_p \delta h_p^T) = A_p (W \otimes I_2) \Sigma (W \otimes I_2) A_p^T
\]

\[
\Sigma = A_p \Sigma A_p^T
\]

where \(A_p^T\) is the **pseudoinverse** of the matrix \(A_p\). \(W = \text{diag}(\{w_1, \ldots, w_M\})\) is the **solutions to the Total Least Squares problem** of eq.(3)

\(^2 h_p\) is the solution to the **Total Least Squares problem** of eq.(3)

\(^3\)Independent and identically distributed

\(^4\)Quantifying “small” is difficult since it depends on the image size. A noise variance of 4 pixels in a 1024 x 768 image is << noise variance of 4 pixels in a 352 x 288 image.
\[ \Xi = [\epsilon'_{11} \epsilon'_{12} \cdots \epsilon'_{1M} \epsilon'_{2M}]^T. \] The matrix \( W \) is comprised of the homogenous scale factors \( \{\alpha_i\} \), while \( \Xi \) is comprised of the noise affecting correspondences in view-2.

**Appendix-1** provides details on deriving the expressions for \( \delta_{hp} \) and \( \Lambda_{hp} \). Smaller entries in \( \Lambda_{hp} \) correspond to a smaller uncertainty in estimating \( h_p \), and vice-versa.

The proposed structure of \( \Lambda_{hp} \) (eq.(7)) has geometric appeal as explained in Sec-3.2.4, and is easily extended to normalized DLT[7] in Sec-3.2.5. In addition, \( \Lambda_{hp} \) can be propagated to projective algebraic terms such as epipoles, parallaxes (Sec-3.3.2), and epipoles. Further, the structure of \( \Lambda_{hp} \) in eq.(7) allows us to implicitly estimate the unknown noise variances, as described in 3.2.6.

Another advantage of the proposed structure of \( \Lambda_{hp} \) is that it can be readily expressed in alternative forms, such as the structure of the covariance matrix proposed by Hartley [7] and Kanatani [10].

\[ \Lambda_{hp} = \left( [(W^{-1} \otimes I_2)A_p]_p \right)^T [E[I\Xi^T]]^{-1} [(W^{-1} \otimes I_2)A_p]^T \] (8)

### 3.2.4 Structure of the covariance matrix \( \Lambda_{hp} \)

- \( \Lambda_{hp} \) is singular, \( \text{rank}(\Lambda_{hp}) \leq 8 \), and \( h_p = \text{null}(\Lambda_{hp}) \). The redundancy \( (||h_p|| = 1) \) in estimating a homography using DLT, forces \( \Lambda_{hp} \) to be singular. As a result, there is no proper density function associated with the perturbation \( \delta_{hp} \).

- The uncertainty in \( h_p \) increases with noise variance. But for large variances the expression for \( \Lambda_{hp} \) is inaccurate, as first-order perturbation analysis of the matrix \( A_p + N_p \) is no longer sufficient.

- The uncertainty in \( h_p \) increases with the homogenous scale factors \( \{\alpha_i\} \). By choosing point correspondences that are clustered around the fixed-point\(^b\) of the homography, we can reduce the uncertainty in \( h_p \). But, we run the risk of incorrectly estimating a projective transform as being nearly affine. Alternatively, we could arrive at a better estimate of the homography by choosing point correspondences that are spread over a larger area. But this comes at the expense of increased uncertainty (larger \( \alpha_i \)'s). This tradeoff is reminiscent of the bias-variance tradeoff in standard estimators.

- The uncertainty in \( h_p \) depends on the orientation of the observed planar facet. The least uncertainty in \( h_p \) results when the camera in view-2 is looking straight at the planar facet (fronto-parallel situation). The reason is that the least error in localizing corners in view-2 results, when camera-2 is in a fronto-parallel situation. Based on this observation, we choose the planar facet that most directly faces camera-

2 (reduces uncertainty) and has the largest area (reduces bias), to estimate the epipolar geometry.

Unfortunately, the DLT algorithm outlined in Sec-3.2.1 cannot be used to estimate the homography \( H_p \), as the ordering of the right singular vectors of \( A_p \) is sensitive to small perturbations in the pixel coordinates. Hartley[7] points out that the sensitivity is due to the large condition number of the matrix \( A_p^T A_p \), and observes that an anisotropic normalization of the pixel coordinates alleviates the problem.

Instead of estimating a homography from noisy pixel coordinates, we estimate the homography in normalized coordinates, using the Normalized DLT algorithm[7]. To be specific, the normalized homography \( h_n^p + \delta_n \) is estimated as the right singular vector corresponding to the least singular value of the normalized matrix

\[ A_n^p + N_n^p = \begin{bmatrix} 1 & 0 & 0 & -e_{n1}^p & -e_{n2}^p & -e_{n3}^p & -e_{n4}^p & -e_{n5}^p \end{bmatrix} \begin{bmatrix} x_n^p & y_n^p \end{bmatrix} \] (9)

\[ x_n^p = \frac{N_1}{N_1(3,3)} \cdot \begin{bmatrix} x_{11}^p + e_{n1}^p & y_{11}^p + e_{n2}^p \end{bmatrix} = \frac{N_2}{N_2(3,3)} \begin{bmatrix} x_{11}^p + e_{n1}^p & y_{11}^p + e_{n2}^p \end{bmatrix} \] (10)

\[ N_1 = \text{ichol} \left( \sum_{i=1}^{M} N_i \right), \quad N_2 = \text{ichol} \left( \sum_{i=1}^{M} \frac{N_i^2}{N_i} \right) \] (11)

The inverse cholesky factors\(^c\) \( N_1 & N_2 \) normalize pixel coordinates in views-1 & 2 respectively. For small noise levels, the dependence of \( N_2 \) on the noise terms in eq.(11) is so negligible that we ignore it. The homography \( H_p \) in pixel coordinates is related to the homography \( H_n^p \) in normalized coordinates as

\[ H_p = \frac{N_2}{N_1} H_n^p \] (12)

### 3.2.5 Perturbation Analysis of Normalized DLT

Along the lines of Sec-3.2.3, we can express the perturbation in the normalized homography \( (\delta_{hp}) \) as

\[ \delta_{hp} = A_n^p \otimes (W^n \otimes I_2) \Xi^n \] (13)

\[ W^n = \text{diag}([H_n^{(1)}; H_n^{(2)}; \ldots; H_n^{(5)}]), \quad \Xi^n = [\epsilon'_{11} \epsilon'_{12} \cdots \epsilon'_{1M} \epsilon'_{2M}]^T \]

The corresponding uncertainty is specified as

\[ \Lambda_{hp} = A_n^p \otimes [W^n \Xi^n] \text{diag}(\sigma_1^2, \sigma_2^2) N_{25} \Lambda_n^p \] (14)

where \( N_{25} = \begin{bmatrix} N_{25}(1,1) & N_{25}(1,2) \n N_{25}(2,1) & N_{25}(2,2) \end{bmatrix} \)

**Appendix-2** provides details on deriving \( \Lambda_{hp} \). Notice that the structure of \( \Lambda_{hp} \) in eq.(14) is similar to the structure of \( \Lambda_{hp} \) in eq.(7), with the exception of the term \( N_{25} \).

---

\(^b\) Using the property \( (AB)^T = B^T A^T \), we can rewrite eq.(6) as eq.(6)

\(^c\) Remarks that \( N_{25} \) is comprised of the noise affecting correspondences in view-2.
However, eq.(14) only describes the uncertainty of the normalized homography $\hat{h}_p^N$. We need to propagate $\Lambda_{hp}$ through eq.(12), to determine $\Lambda_{hp}$, in pixel coordinates. The process begins by propagating noise through $h$ in eq.(12), and finally through the normalization $\hat{h}$ in (13). The uncertainty in $h$ has the following structure

$$\Lambda_{hp} = \frac{1}{|h|} \begin{bmatrix} I_0 - hh^T & \Lambda_{p} & I_0 - hh^T \end{bmatrix} \left( N^{-1}_2 \otimes N_1^1 \right) \left| h \right|$$

where $h = (N^{-1}_2 \otimes N_1^1) \hat{h}_p^N$, and $\Lambda_{hp} = (N^{-1}_2 \otimes N_1^1) \Lambda_{hp}$.

Substituting for $\Lambda_{hp}$ from eq.(14), we find that

$$\Lambda_{hp} = \Lambda_{hp}^\Lambda \Lambda_{hp}^\gamma$$

Alternatively, we can rewrite eq.(16) in terms of $\sigma^2_x$ and $\sigma^2_y$ as

$$\Lambda_{hp} = \sigma^2_x \Lambda_{hp}^x + \sigma^2_y \Lambda_{hp}^y$$

In Sec.3.2.6, we use this structure of $\Lambda_{hp}$ to find $\sigma^2_x$ and $\sigma^2_y$. Comments about the structure of the uncertainty $\Lambda_{hp}$ from Sec.3.2.4 are still valid.

We conclude the study of uncertainty in estimating homographies using DLT, by noting that the estimated homography $\hat{h}_p$ is a random vector with mean $\hat{h}_p$ true homography defined in eq.(12) covariance $\Lambda_{hp}$ defined in eq.(16).

Note: There is no proper density function associated with $\hat{h}_p$, as $\Lambda_{hp}$ is singular.

3.2.6 Numerically estimating $\Lambda_{hp}$

Provided we know the noise variances ($\sigma^2_x$, $\sigma^2_y$), we can find an estimate $\hat{\Lambda}_{hp}$ of the true covariance $\Lambda_{hp}$, by replacing the true quantities ($A_p^N, H_p^N$) in eq.(17), with observed or estimated quantities ($A_p^N, H_p^N$).

In the absence of noise statistics, we estimate $\sigma^2_x, \sigma^2_y$ from the statistics of the algebraic transfer error in view-2. The algebraic transfer error for the $i^{th}$ point correspondence is defined as

$$r_i^T = \begin{bmatrix} x_i^T + e_i^T y_i \end{bmatrix} - \begin{bmatrix} H_p^N(3,:) \end{bmatrix} \begin{bmatrix} x_i \end{bmatrix}$$

where $r_i$ is a random vector with 2 sources of uncertainty 1. noisy point correspondences in view-2

2. error in point transfer using the noisy homography $\hat{H}_p$

Assuming the transfer errors to be gaussian distributed,

$$r_i^T \sim N \left( H_p^N(3,:) \begin{bmatrix} x_i \end{bmatrix} \right)$$

$$\hat{H}_p = \begin{bmatrix} H_p^N(3,:) \end{bmatrix} \begin{bmatrix} x_i \end{bmatrix}$$

Please refer to [7] for details on deriving $\Lambda_{hp}$. It can be shown that the matrix $R = [r_1 \ldots r_M]$ formed by stacking the algebraic transfer errors for the $M$ point correspondences, has the following statistics

$$E[RRT] = M \left[ \sum_{i=1}^M r_i r_i^T \right] = M \left[ \begin{bmatrix} \sigma^2_x & 0 \\ 0 & \sigma^2_y \end{bmatrix} \right] + \sum_{i=1}^M E[h_p^N]$$

The left-hand-side of eq.(21) can be estimated directly from the data. We find estimates $\hat{\sigma}^2_x, \hat{\sigma}^2_y$ of the true noise variances $\sigma^2_x, \sigma^2_y$, by solving the linear system of eq.(21), after these substitutions:

$$\hat{J}_p = \begin{bmatrix} \hat{J}_p^x & \hat{J}_p^y \end{bmatrix}$$

3.2.7 Refining the homography $\hat{H}_p$

We now examine how knowledge of $\hat{\Lambda}_{hp}$ could be used to refine $\hat{H}_p$. Suppose, $\mathcal{I}$ is the set of noisy correspondences in the two views, and $\mathcal{I}_hp$ is the subset with homography $\hat{H}_p$. We refine $\hat{H}_p$ by using correspondences in $\mathcal{I} \setminus \mathcal{I}_hp^9$ that are consistent with $\hat{H}_p$, while discarding inconsistent correspondences in $\mathcal{I}_hp$. We define the $i^{th}$ point correspondence as consistent only when the Mahalanobis norm of its algebraic transfer error given by $d_{alg}[i]$ is $< 5.9915$.

$$d_{alg}[i] = r_i^T \left( \begin{bmatrix} \hat{\sigma}^2_x & 0 \\ 0 & \hat{\sigma}^2_y \end{bmatrix} + \hat{J}_p \hat{\Lambda}_{hp} \hat{J}_p^T \right) r_i$$

where $\hat{\sigma}^2_x, \hat{\sigma}^2_y$ are the estimated noise variances, and $\hat{J}_p$ is defined in eq.(20). The number 5.9915 specifies the Probability(misclassifying an inconsistent correspondence) $< 5\%[7]$. This follows from the fact that $d_{alg}[i]$ is a $\chi^2$ distributed random variable with 2 d.o.f. The proposed thresholding scheme is based on the thresholding scheme for sampson distances, utilized in estimating homographies using RANSAC[7].

The success of the proposed inpainting scheme depends on the ability to reliably estimate the epipolar geometry from noisy correspondences. So we examine the effect of noise on estimating the epipolar line constraint.

---

7 From eq.(12) we find that a perturbation $\delta h_p$ in $h_p$, induces the perturbation $\delta h = (N^{-1}_2 \otimes N_1^1) \delta h_p$, in $h$. The corresponding uncertainty is $\Lambda h = E[\delta h \delta h^T] = (N^{-1}_2 \otimes N_1^1) \Lambda h_p (N^{-1}_2 \otimes N_1^1)$

8 The gaussian assumption breaks down in the presence of outliers.

9 The set difference $A \setminus B$ is defined as $A \setminus B = \{x : x \in A \land x \notin B\}$
3.3. Epipolar Geometry

Given two views of a static scene, the epipolar line constraint \[7\] describes a generic relationship between corresponding points \(^{10}\) in the two views. Suppose \(x_j = [x_j, y_j, 1]^T\) and \(x_j' = [x_j', y_j', 1]^T\) are corresponding points \(^{10}\) in views-1 & 2 respectively. They satisfy the epipolar line constraint
\[
\begin{align*}
x_j^T F x_j &= 0, \\
x_j^T F^T x_j' &= 0
\end{align*}
\] (23)
where \(F\) is a singular matrix depending on the relative position and orientation of the cameras in the two views. The matrix \(F\) is referred to as the fundamental matrix. The epipolar line constraint derives its name from the fact that it describes a point-line incidence relation \(^{11}\). The entities \(l_{1j}, l_{2j}\) are referred to as conjugate epipolar lines.

If \(x_j\) is a missing pixel, its corresponding point in view-2 must satisfy the epipolar line constraint of eq.\((23)\). We exploit this knowledge to inpaint the missing pixel \(x_j\).

3.3.1 Estimating the Fundamental matrix

Given a planar facade that induces a homography \(H_p\) between the two views and the epipole \(^{12}\) \(e_j\) in view-2, the fundamental matrix \(F\) is specified as
\[
F = [e_j]_x H_p
\] (24)
The epipole \(e_j\) is the intersection of all epipolar lines in view-2. As mentioned before, if \(x_j, x_j'\) are corresponding points in the two views, \(x_j'\) lies on the line \(l_{2j} = F x_j\). It can be shown\(^{7}\) that the point \(\tilde{x}_j' \sim H_p x_j\) \(^{13}\), also lies on the line \(l_{2j}\). This means that the line joining \(x_j'\) and \(\tilde{x}_j\) is a fragment of the epipolar line \(l_{2j}\). Given two such fragments we can compute \(\tilde{e}_j\), and the full epipolar geometry follows from eq.\((24)\). The vector joining \(\tilde{x}_j\) and \(x_j'\) is referred to as the parallax vector. Consequently, this approach to estimating \(F\) is referred to as the plane+parallax method.

To reliably estimate the fundamental matrix using the plane+parallax method, we need to identify at least two parallax vectors with little-or-no uncertainty in their orientation. This is a challenging task given noisy point correspondences in the two views. For short parallax vectors \((x_j' \& x_j')\) are close), even small perturbations in the end-points significantly alters the orientation of the parallax vector. This suggests that the length of a parallax vector could be used to determine its reliability. But, it is likely that grossly mismatched correspondences are mistaken for reliable parallax vectors. So, we assume that grossly mismatched correspondences have been removed prior to inpainting. The sources of error affecting the orientation of a parallax vector are
1. noise in the correspondence \(x_j'\)
2. noise in \(\tilde{x}_j\) due to noise in the estimated homography \(\hat{H}_p\)

We now examine how uncertainty in estimating \(H_p\) can be combined with the length of a parallax vector, to identify the two most reliable parallax vectors.

3.3.2 Reliability of parallax vectors

Given a noisy homography \(\widehat{H}_p\), an estimate of its uncertainty \(\tilde{\Lambda}_{y_j}\), noise variances \(\sigma_{x_j'}^2, \sigma_{y_j'}^2\), and a noisy point correspondence \(\{ (x_j, y_j), (x_j' + e_{x_j'}, y_j' + e_{y_j'})\} \) is inconsistent\(^{14}\) with the homography, we can construct an epipolar line segment by joining \(\left(\frac{H_p(1,:)}{H_p(3,:)} x_j, \frac{H_p(2,:)}{H_p(3,:)} y_j\right)\) and \((x_j' + e_{x_j'}, y_j' + e_{y_j'})\). The length of this line segment is
\[
d_j = \frac{\| \tilde{e}_j \|}{\| \hat{H}_p(1,:) \tilde{e}_j + \hat{H}_p(2,:) \tilde{e}_j \|}^2
\] (25)
It is easy to observe that \(d_j\) is a random variable. Using the noise model of Sec.3.2.2 and the covariance structure \((\Lambda_{y_j})\) of \(y_j\) defined in eqs.\((19)\) and \((20)\), we can prove that
\[
e \{ d_j \} = \tilde{d}_j + \text{tr}
\[
\tilde{d}_j = \left[ \begin{array}{c} x_j' \\ y_j' \end{array} \right] - \frac{1}{\hat{H}_p(3,:) x_j} \hat{H}_p(2,:) y_j, \Lambda_{y_j} \hat{H}_p(3,:) x_j
\] (26)
where \(\tilde{d}_j\) is the true length of the epipolar line segment. We define the reliability of a parallax vector as the ratio \(d_j^{-1}\left(\sigma_{x_j'}^2 + \sigma_{y_j'}^2 + \text{tr}(\Lambda_{y_j})\right)\). The ratio is minimum when the noise is minimum and the length of the line segment is maximum. Given 2 or more correspondences that are consistent with the homography \(\widehat{H}_p\), we find the two parallax vectors with the highest reliability in the sense of
\[
d_j^{-1}\left(\sigma_{x_j'}^2 + \sigma_{y_j'}^2 + \text{tr}(\tilde{\Lambda}_{y_j})\right)
\] (27)
where \(d_j\) is the observed length of the epipolar line segment, defined in eq.\((25)\). The chosen parallax vectors are intersected to find the epipole \(\tilde{e}_j\) in view-2. The full epipolar geometry follows from eq.\((24)\).

4. Putting it all together

Fig.\((1)\) summarizes the process of inpainting in the proposed approach. The user first selects images corresponding to views-1 \& 2. We then proceed to identify putative correspondences in the two views (colored dots in Figs\(1(a),1(d)\)).
Next, the user outlines a hypothetical plane \( \mathcal{R} \) & the regions to be inpainted \( \{\Omega_k\} \), in view-1. The hypothetical plane outlined in blue in Fig.1(a), is any region that includes correspondences from a planar facet. Using the set of correspondences \( \in \mathcal{R} \), we estimate a homography \( \tilde{H}_p \), its associated uncertainty \( \tilde{\Lambda}_{hp} \), and noise variances \( \tilde{\sigma}_1^2, \tilde{\sigma}_2^2 \). Using the strategy of Sec-3.2.7, we refine the homography \( \tilde{H}_p \) until the ratio of consistent correspondences is a minimum. The result of the refinement is shown in Figs.1(a),1(d). The cyan points in Figs.1(a),1(d) are consistent correspondences, while the yellow & green points are the inconsistent correspondences \( \in \mathcal{R} \& \notin \mathcal{R} \) respectively. We use the refined homography \( \tilde{H}_p \) to determine the two most reliable parallax vectors, as outlined in Sec-3.3.2. The best set of parallax vectors labeled 1 and 2 in Fig.1(d) are intersected to find the epipole \( \xi' \) in view-2. The full epipolar geometry follows from eq.(24).

If there is more than one planar facet in the scene, the best inpainting is achieved for the planar facet that most directly faces camera-2, as described in Sec-3.2.4.

4.1. Inpainting using the “epipolar line constraint”

From Sec-3.3, we know that if \( x_m \) is a missing pixel in view-1, its corresponding point in view-2 satisfies the epipolar line constraint of eq.(23). Suppose \( l_{m1} = (\mu u (F))_{x \in \mathcal{R}} x_m \) and \( l_{m2} = F x_m \) are conjugate epipolar lines in views-1 & 2, corresponding to \( x_m \). In principle, we could inpaint \( x_m \) by searching for the best matching patch along the epipolar line \( l_{m2} \), in view-2. But this is time consuming, computationally intensive, and not always correct. A simpler alternative to inferring missing pixels is globally stretching/shrinking \( l_{m2} \) until fragments of \( l_{m1} \& l_{m2} \) are aligned.

We repeat the steps outlined below for each pair of conjugate epipolar lines \( \{l_{m1}, l_{m2}\} \), in each region \( \Omega_k \).

1. Identify landmarks on \( l_{m1} \) & \( l_{m2} \) as points with the largest gradient magnitude along the epipolar line ( \( \Rightarrow \) edge runs across the epipolar line)
2. Match the neighborhood around each landmark in view-1 with every landmark in view-2, using LP-NCC
3. Find at least 2 uniquely matching landmarks on \( l_{m1} \& l_{m2} \)
4. Rotate \( l_{m1} \& l_{m2} \) around the respective epipoles
5. Find the 1-d affine transform \( A_m \) that aligns \( l_{m1} \& l_{m2} \), using the coordinates of the matching landmarks
6. Transform the points \( (x_m, y_m) \in \{l_{m1} \cap \Omega_k\} \) using \( A_m \), to find points on \( l_{m2} \) with the missing intensities

Any error in identifying landmarks introduces errors in the 1-d affine transform and hence the inpainting. Failure to identify landmarks results in the inability to inpaint a specific conjugate line pair.

Contrary to our assumption, conjugate epipolar lines can be aligned using a locally affine transform. We are currently investigating the use of elastic registration[13] techniques to locally align a pair of conjugate epipolar lines.

5. Results

The effectiveness of the proposed approach is best demonstrated in Fig.1, where we attempt to inpaint parts of a pillar like structure. It is difficult to accurately estimate the epipolar geometry in this example, as we have a limited number of correspondences, and short parallax vectors. A quick look at Fig.1(d) reveals that the proposed inpainting scheme has succeeded in inpainting the pillar like structure. The quality of the inpainting can be judged by the accuracy with which corners on the roof, and vertical structures within the missing region have been reconstructed.

We provide two more examples of the proposed inpainting, wherein traditional image inpainting methods would have failed. In the first example ( Figs.2(a)-2(d) ) we attempt to recover the missing detail in the bookshelf, while in the second example ( Figs.2(e)-2(h) ) we inpaint a lamp & the window panes behind it. A quick look at Fig.2(a),2(h), reveals that we have successfully recovered the missing detail in the bookshelf of Fig.2(a), and inpainted the lamp & the missing window panes in Fig.2(h). Notice that we succeed in recovering the arch like structure on the window panes, in Fig.2(h). But, a closer examination of Figs.2(d),2(h) reveals that the some of the objects are scaled incorrectly. This is a drawback of assuming that a pair of conjugate epipolar lines can be aligned globally using a 1-d affine transform.

6. Conclusions

We extend the scope of image inpainting by using the epipolar line constraint to infer the missing pixels from an exemplar image. We overcome the difficulty in accurately estimating the epipolar geometry from noisy point correspondences, by using the proposed structure of uncertainty in a homography and the proposed measure of the reliability of a parallax vector.

The quality of the inpainting can be improved further, by

1. using elastic matching to align conjugate epipolar lines
2. non-linear refinement[7] of the estimated homography
3. using additional views to fully inpaint the occluding objects.

Appendix-1. Perturbation in DLT homography

Using first-order perturbation analysis[6] we can express the perturbation in the DLT homography \( \delta_{hp} \) in terms of the unperturbed singular values & right singular vectors \( (\sigma_j, v_j)_{j=1}^9 \) of \( A_p \).

\[
\delta_{hp} = -\sum_{j=1}^9 \frac{\|v_j\|_F^2}{\sigma_j} [A_p^T N_p h_p + N_p^T A_p h_p] \quad (A1)
\]

We simplify eq.(A1) by noting that \( A_p h_p = 0 \), and

\[
\delta_{hp} = \sum_{j=1}^9 \frac{\|v_j\|_F^2}{\sigma_j} [T_{ij} T_{jk} N_p h_p] \quad (A2)
\]

where \( T_{ij} \) is the pseudo inverse[6] of the matrix \( A_p \). Using eq.(4) we can express the term \( N_p h_p \) as
Appendix-2. Uncertainty in Normalized DLT

From eq. (13), the perturbation in the normalized homography is \( \delta h_p = A_p^T (W \otimes I_2) \Xi \). The corresponding uncertainty \( \Lambda_{hp} \) is

\[
\Lambda_{hp} = E \left\{ \delta h_p \delta h_p^T \right\} = A_p^T (W \otimes I_2) E (\Xi \Xi^T) (W \otimes I_2) A_p^{T T}
\]

(A8)

To model the statistics of the noise vector \( \Xi^n \), we need to examine the effect of normalizing the noise terms \( \{ \epsilon_{nm}^n \} \). From eq. (10),

\[
\begin{bmatrix}
\epsilon_{11}^n \\
\epsilon_{12}^n \\
\epsilon_{13}^n \\
\epsilon_{21}^n \\
\epsilon_{22}^n \\
\epsilon_{23}^n \\
\epsilon_{31}^n \\
\epsilon_{32}^n \\
\epsilon_{33}^n \\
\end{bmatrix} =
\begin{bmatrix}
N_1(1,1) & N_2(1,2) & N_3(1,3) \\
N_1(2,1) & N_2(2,2) & N_3(2,3) \\
N_1(3,1) & N_2(3,2) & N_3(3,3) \\
\end{bmatrix}
\begin{bmatrix}
\epsilon_{11}^m \\
\epsilon_{12}^m \\
\epsilon_{13}^m \\
\epsilon_{21}^m \\
\epsilon_{22}^m \\
\epsilon_{23}^m \\
\epsilon_{31}^m \\
\epsilon_{32}^m \\
\epsilon_{33}^m \\
\end{bmatrix}
\]

(A9)

To study the effect of normalization on the noise terms \( \{ \epsilon_{nm}^n \} \), we need to examine the effect of normalizing the noise terms \( \{ \epsilon_{nm}^n \} \). From eq. (10),

\[
\begin{bmatrix}
\epsilon_{11}^n \\
\epsilon_{12}^n \\
\epsilon_{13}^n \\
\epsilon_{21}^n \\
\epsilon_{22}^n \\
\epsilon_{23}^n \\
\epsilon_{31}^n \\
\epsilon_{32}^n \\
\epsilon_{33}^n \\
\end{bmatrix} =
\begin{bmatrix}
N_1(1,1) & N_2(1,2) & N_3(1,3) \\
N_1(2,1) & N_2(2,2) & N_3(2,3) \\
N_1(3,1) & N_2(3,2) & N_3(3,3) \\
\end{bmatrix}
\begin{bmatrix}
\epsilon_{11}^m \\
\epsilon_{12}^m \\
\epsilon_{13}^m \\
\epsilon_{21}^m \\
\epsilon_{22}^m \\
\epsilon_{23}^m \\
\epsilon_{31}^m \\
\epsilon_{32}^m \\
\epsilon_{33}^m \\
\end{bmatrix}
\]

(A10)

Using eqs. (A7, A10) in eq. (A8) and simplifying, we obtain the expression (eq. (14)) for \( \Lambda_{hp} \) in compact form.

References