Image Warping using Local Homography

A Project Report

submitted by

Pavan C M (13990)

as part of

Advanced Image Processing (E9 246)

undertaken during

January - April 2017

DEPARTMENT OF ELECTRICAL COMMUNICATION ENGINEERING

INDIAN INSTITUTE OF SCIENCE BANGALORE - 560012

April 23, 2017

TABLE OF CONTENTS

CHAPTER 1
Introduction

1.1   Image Warping

Image warping refers to the process of transformation of digital images such that any shapes portrayed in the image may undergo deformation. Mathematically warping involves mapping co-ordinates by a suitable function and resampling. Forward mapping refers to obtaining warped image co-ordinates (x, y) from source image co-ordinates (u, v). If source image is obtained from warped image, it is referred as reverse mapping (provided the transformation is invertible). An example of warped images are shown in figure 1.1.

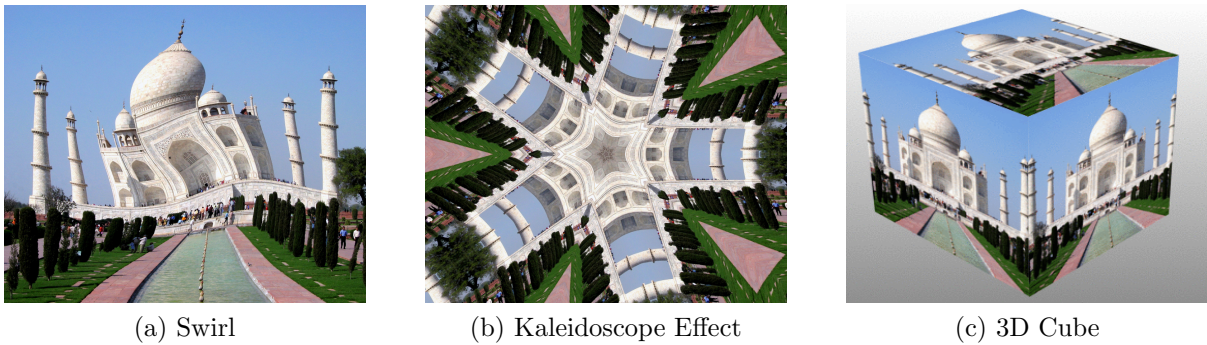| (a) Swirl | (b) Kaleidoscope Effect | (c) 3D Cube |

Figure 1.1: Figures illustrating various warping methods applied on Taj Mahal image

In this project only a subclass of warping techniques known as projective transformations are studied. Projective transformations are represented by a matrix H of size 3×3 and each co-ordinate of source image represented as $x = (x_1, x_2, 1)$ undergoes transformation defined by the expression $x' = \mathbf{H}x$. The matrix $\mathbf{H}$ is known as homography and this transformation is also known as homographic transformation. Homography between two planes is illustrated in figure 1.2.
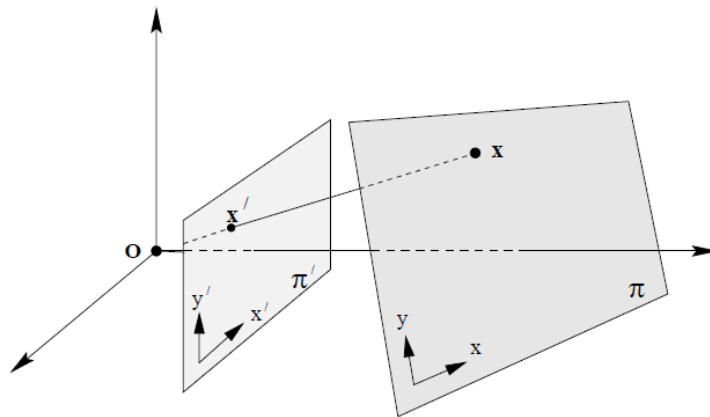

Figure 1.2: Depiction of projective transformation between two planes π' and π. Image taken from [6]

.

$$x' = \mathbf{H}x$$

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} \tag{1.1}$$

Equation 1.1 shows the generic structure of homography. The matrix $\mathbf{H}$ has 9 variables but only 8 degrees of freedom as h_{33} is a scale factor and not computed. Another constraint on equation 1.1 is that the matrix $\mathbf{H}$ has to be invertible.

Projective transforms/warps are applied extensively in areas such as image stitching, video stabilization etc. In this project projective warps in the context of image stitching is discussed. In case of image stitching, the two planes π' and π depicted in figure 1.2 are consecutive images taken from a rotating camera.

1.2 Direct Linear Transformations (DLT)

Basic image stitching pipeline is shown in figure 1.3. Two or more overlapping images which are to be stitched are taken from a rotating camera. Features such as SIFT [1] or SURF [2] are used for detecting as well as for describing keypoints in the input images. Keypoints are matched using feature descriptors with a suitable distance measure (eg. Euclidean distance). Matched features might not be robust and might contain some outlier points, outliers are removed using Random Sample Consensus (RANSAC) [3].
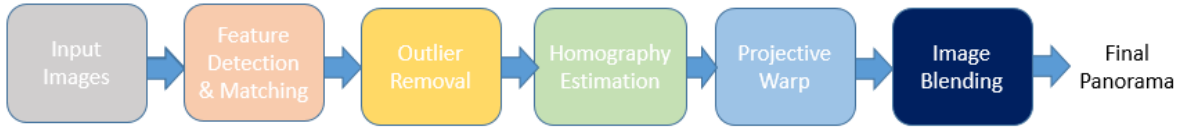

Figure 1.3: Image Stitching Pipeline

1.2.1 Projective Warp

Let $\mathbf{x} = [x, y]^T$ and $\mathbf{x}' = [x', y']^T$ be the matched points between images I and I' after RANSAC procedure. A projective warp maps $\mathbf{x}$ and $\mathbf{x}'$ using the equation

$$\tilde{\mathbf{x}}' \sim \mathbf{H}\tilde{\mathbf{x}} \tag{1.2}$$

where $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{x}}'$ are in homogeneous co-ordinates and $\mathbf{H} \in R^{2 \times 3}$. Equation 1.2 is valid upto scale, where it is true for any any unknown scalar α, such that $\alpha\tilde{\mathbf{x}}' = \mathbf{H}\tilde{\mathbf{x}}$. DLT is a basic method to estimate H from a set of noisy matched points $\{\mathbf{x_i}, \mathbf{x'_i}\}_{i=1}^{N}$. Equation 1.2 is rewritten as $\mathbf{0_{3 \times 1}} = \tilde{\mathbf{x}}' \times \mathbf{H}\tilde{\mathbf{x}}$ and linearised

$$\mathbf{0_{3 \times 1}} = \begin{pmatrix} \mathbf{0_{3 \times 1}} & -\tilde{\mathbf{x}}^T & y'\tilde{\mathbf{x}}^T \\ \tilde{\mathbf{x}}^T & \mathbf{0_{3 \times 1}} & -x'\tilde{\mathbf{x}}^T \\ -y'\tilde{\mathbf{x}}^T & x'\tilde{\mathbf{x}}^T & \mathbf{0_{3 \times 1}} \end{pmatrix} \mathbf{h}, \quad h = \begin{pmatrix} \mathbf{h_1} \\ \mathbf{h_2} \\ \mathbf{h_3} \end{pmatrix} \tag{1.3}$$

In equation 1.3 only two rows are linearly independent. Let $a_i \in R^{2 \times 9}$ be the first two rows of 1.3 calculated for the ith data $\{\mathbf{x_i}, \mathbf{x_i'}\}$. DLT estimates the nine element matrix $\mathbf{H}$ as

$$\mathbf{h}^* = arg \min_{\mathbf{h}} \sum_{t=1}^{N} \|\mathbf{a_t} h\|^2 = arg \min_{\mathbf{h}} \|\mathbf{A}h\|^2 \tag{1.4}$$

Equation 1.4 is solved with the constraint $\|h\| = 1$, with matrix $A = R^{2N \times 9}$. The solution to equation 1.4 is obtained by Singular Value Decomposition (SVD) and it corresponds the least significant right singular vector of A (There exists a unique null vector since the rank of null space of A is 1).

There are certain limitations associated with the above model. it assumes either the images taken by the rotating camera is parallax free/purely rotational (no translation should be present between two consecutive camera imaging positions) or imaging scene has to be planar [4]. Deviation from the above assumptions results in stitching inconsistencies and it can manifest as ghosting artifact. The above limitation is illustrated in figure 1.4
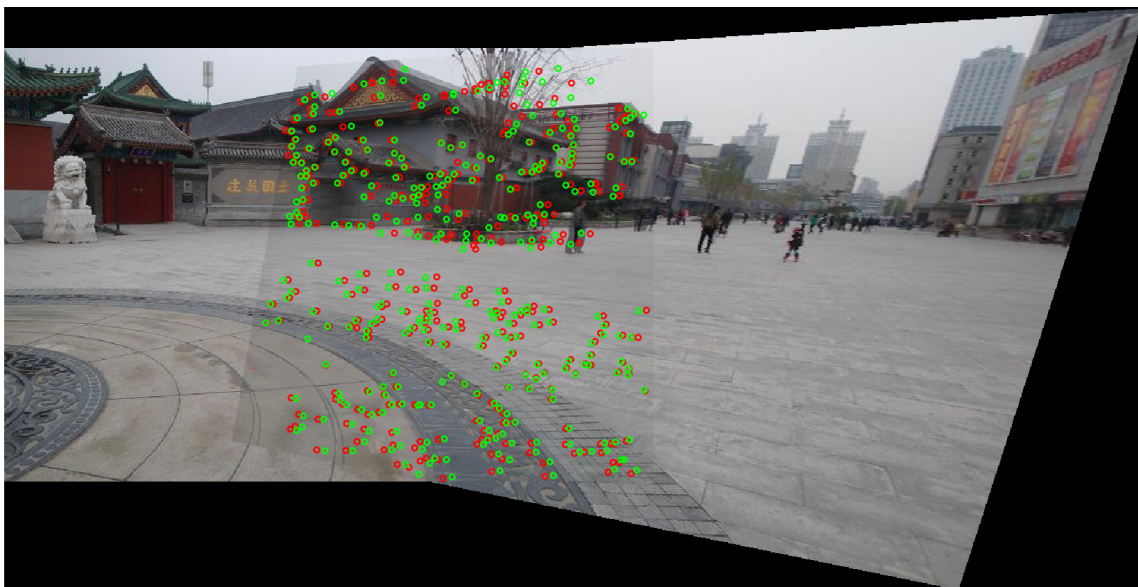

Figure 1.4: Image showing matching points in 2 images. Since the image violated homography conditions, there is a mismatch between the points. Ideally all the points should have been overlapping.

1.3 Moving DLT or Local Homography

The conditions required for application of homography model is rarely achieved in practice, as there exists some amount of translation between consecutive images. There are two ways by which ghosting introduced by image misaligment can be overcome, one is using suitable deghosting methods in the blending stage of image stitching pipeline shown in figure 1.3 and the other one being usage of alternative models for homography in the alignment stage. Moving DLT or local homography introduced in [5] modifies homography model to reduce stitching inconsistencies when required conditions for homography

model are violated. The idea of local homography is to warp each $\tilde{\mathbf{x}}$ using a location dependent homography

$$\tilde{\mathbf{x}}'_* \sim \mathbf{H}_* \tilde{\mathbf{x}}_* \tag{1.5}$$

where $\mathbf{H}_*$ is estimated from the weighted optimization problem

$$\mathbf{h}_* = arg \min_{\mathbf{h}} \sum_{t=1}^{N} \|w_*^t \mathbf{a_t} h\|^2 \tag{1.6}$$

under the constraint $\|h\| = 1$. The weights $\{w_*^t\}_{t=1}^{N}$ change according to location x_*. An example of weighting function is

$$w_*^t = \exp\left(-\frac{\|x_* - x_t\|^2}{\sigma^2}\right) \tag{1.7}$$

Here σ is a parameter controlling the effect of neighboring pixels and x_t is the co-ordinate of the matched point in the source image. it can be observed in equation 1.7 that pixels near x_t get higher weightage than those very far. Since homography is computed with respect to every pixel x_*, this method is known as local homography.
The problem in equation 1.6 can be written in matrix form as

$$\mathbf{h}_* = arg \min_{\mathbf{h}} \|\mathbf{W}_* \mathbf{A} h\|^2 \tag{1.8}$$

where $\mathbf{W}_* \in \mathbb{R}^{2N \times 2N}$ is the weight matrix composed as

$$\mathbf{W}_* = diag\left(w_*^1 \quad w_*^1 \quad \dots \quad w_*^N \quad w_*^N\right) \tag{1.9}$$

Equation 1.8 is a weighted SVD (WSVD) problem and the solution is the least significant right singular vector of $\mathbf{W}_* \mathbf{A}$. To avoid numerical issues when points are very far, the weight w_* is offset by a parameter $\gamma \in [0, 1]$

$$w_*^t = \max\left(\exp\left(-\frac{\|x_* - x_t\|^2}{\sigma^2}\right), \gamma\right) \tag{1.10}$$

CHAPTER 2
Observations and Results

2.0.1 Implementation Details

Given input images are subjected feature detection and matching using VLFeat library [7]. The mismatches among $\{x_t, x_t'\}_{t=1}^{N}$ are removed using RANSAC [3]. Solving equation 1.8 for all pixel locations x_* is not necessary as neghbouring estimates of $\mathbf{H}_*$ yield practically same estimates. So entire source image I is divided into $C_1 \times C_2$ cells as shown in figure and within each cell all the pixels are subjected to same projective warp. Also weights w_*^t can be computed in parallel as all weights are independent. These drastically reduce the computation time for local homography. The division of source image into cells is shown in figure 2.1.
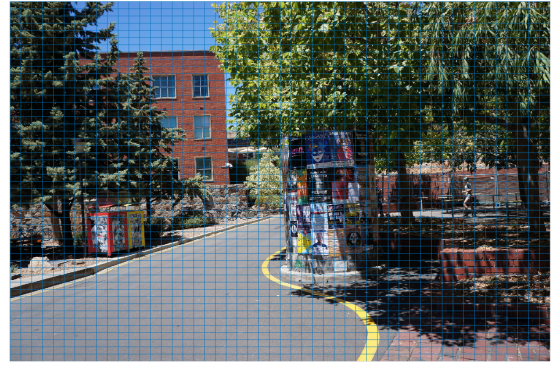
(a) Target Image (b) Source Image with cell division

Figure 2.1: Figure illustrating division of source image into cells. $\gamma = 0.025, \sigma = 8$ were found to give best results for these images

During implementation, the values C_1 and C_2 were chosen from the range $[50, 100]$. The parameters in equation 1.10, γ and σ were found to vary across the dataset and no single value was able to give best results for all images. Hence they were chosen on a case by case basis. σ was varied in the range $[8, 12]$ and γ in $[0.0015, 0.025]$.

2.0.2 Observations

2.0.2.1 Variation of local homography

Histogram of the Frobenius norm of local homography calculated for every cell for a source image is shown 2.2. it can be inferred from the histogram that one particular homography is observed substantially higher (almost 55%) as compared to others. This observation is also illustrated in the corresponding heatmap showed in 2.2. The heatmap representa variation in local homography across source image. Regions with same color are subjected to same homography. As observed in the heatmap, majority of the regions undergo same projective transformation and the region of variation in local homography is observed around the co-ordinates of matched points. The above observation generalizes to all images in the dataset.
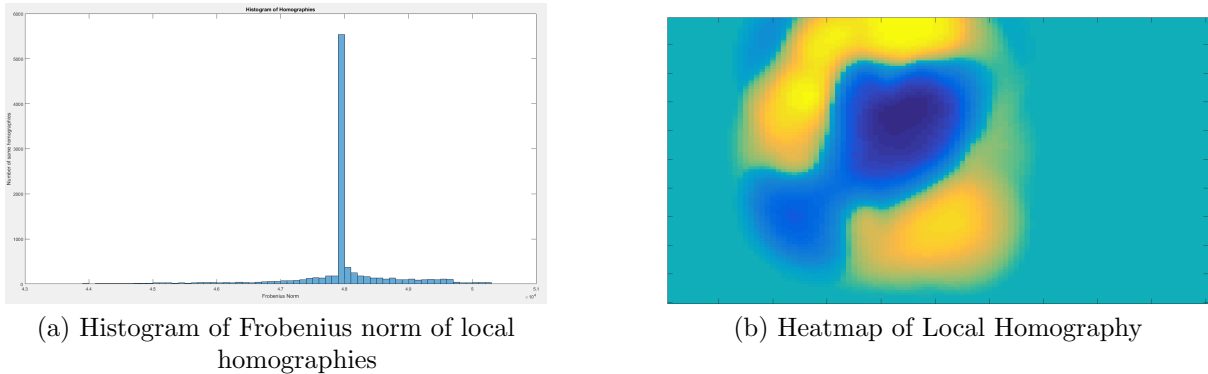
(a) Histogram of Frobenius norm of local homographies

(b) Heatmap of Local Homography

Figure 2.2: Local homography variation

2.0.2.2 Qualitative Comparison

Image stitching with global homography (baseline) and local homography were qualitatively compared and the results for 2 images are shown in figure 2.3. As evident from the figure, local homography removes blurring and ghosting happening due to improper image alignment. Global homography performs poorly as the views of the images do not differ purely by rotation. Note that in the image stitching of overlapped regions, simple averaging is performed. If exposure differences are present simple averaging results in visible seams whose removal requires advanced blending methods. In this project only the effect varying projective warps is studied.

2.0.2.3 Quantitative Comparison

To quantify the alignment accuracy of an estimated warping function $f : \mathbb{R}^2 \to \mathbb{R}^2$, root mean squared error of f on a the set of matched points $\{x_t, x_t'\}_{t=1}^N$ is computed.

$$RMSE(f) = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\|f(x_t) - f(x_t')\|^2} \tag{2.1}$$

This measure is obtained under the assumption that the matched correspondences are reasonably accurate. This measure only gives information on how well matched points are aligned in the overlapping region. This does not consider distortions appearing in non-overlapping regions, therefore this does not always represent the perceptual quality of stitched image. RMSE decreases for local homography in general for optimal σ and γ values (although in some cases it can increase, the increased amount is generally small) as compared to that of baseline.

2.0.2.4 Effect of σ

The effect of standard deviation parameter σ used in the weight function is shown in figure 2.4. In certain images with sigmificant textures in the overlapping region or near the overlapping region smaller σ values doesn't result in smooth transition of local homography giving rise to visible discontinuities as observed in figure 2.4. Higher σ values perform better in such cases.
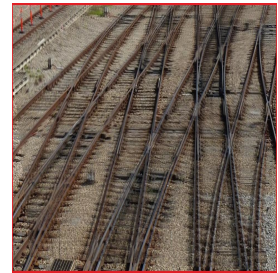
(a) Global homography (RMSE = 5.5
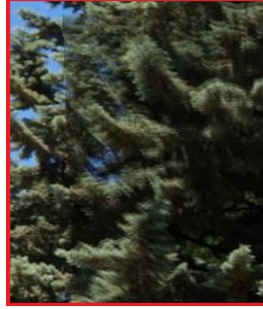
(d) Local homography ($\gamma = 0.025, \sigma = 8$)
(RMSE = 6.09)
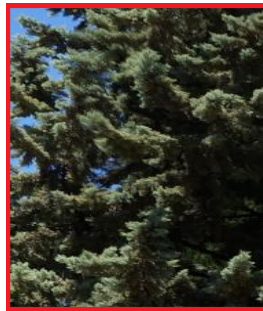
(g) Global homography (RMSE = 15.62)

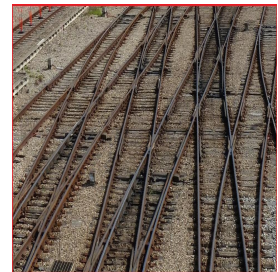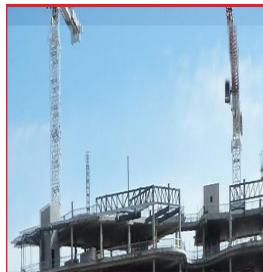(j) Local homography ($\gamma = 0.0015, \sigma = 12$)
(RMSE = 13.38)

Figure 2.3: Global and Local homography comparison

(a) $\sigma = 8, \gamma = 0.025$ (RMSE = 12.02)

(d) $\sigma = 12, \gamma = 0.025$ (RMSE = 13.1662)

Figure 2.4: variation of performance with σ

2.0.2.5 Effect of γ

The effect of offset value γ is shown in figure 2.5. Typically γ values affect the textured regions which are far from the overlap region(region consisting matched keypoints), as the weight function will be equal to γ value in that part of the image. Very large or very small γ values might result in discontinuities in the stitched image due to sudden change in weights across the region. γ values between the lower and higher extremes in images generally produce better results.
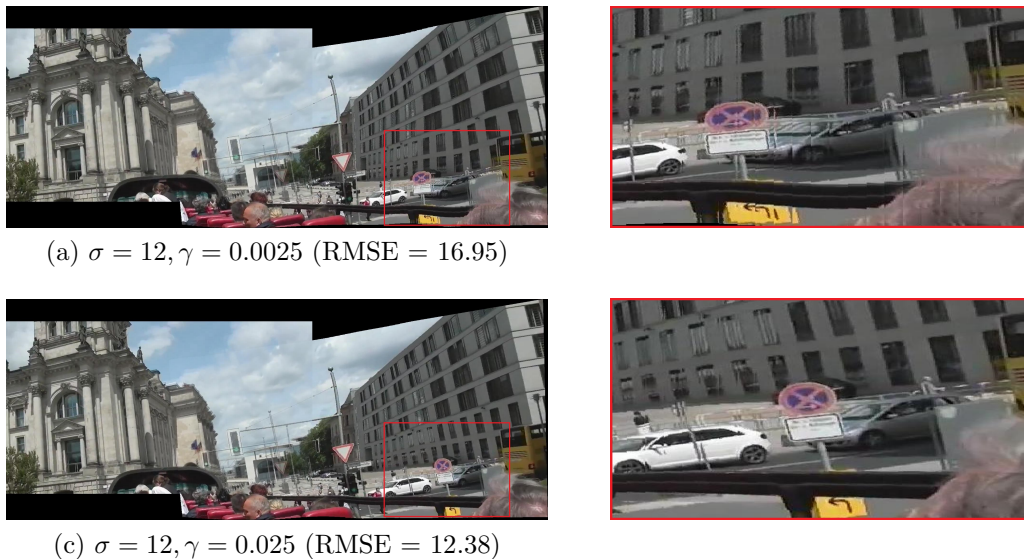

(a) $\sigma = 12, \gamma = 0.0025$ (RMSE = 16.95)

(c) $\sigma = 12, \gamma = 0.025$ (RMSE = 12.38)

Figure 2.5: variation of performance with γ

2.0.2.6 Variation with C_1 and C_2

Increase in C_1 and C_2 increases computational time for calculating local homographies. Many a times high values of C_1 and C_2 is often wasteful as the neighbouring regions yield practically same homography estimates as observed in figure 2.2. Figure 2.6 shows effect of variation with different C_1 and C_2 values. Smaller C_1 and C_2 values results in ghosting artifacts as homography is applied to larger regions, producing effects similar to those observed in global homography.
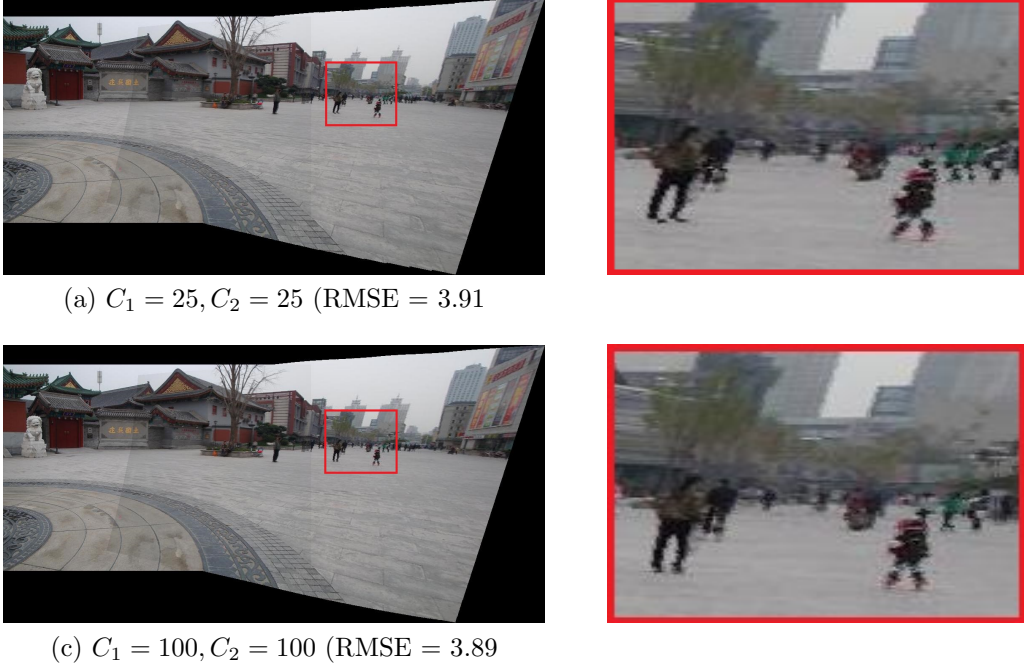

(a) $C_1 = 25, C_2 = 25$ (RMSE = 3.91

(c) $C_1 = 100, C_2 = 100$ (RMSE = 3.89

Figure 2.6: variation with C_1 and C_2

2.1 Limitations

- The extrapolation of homographic transformation in the non-overlapping areas produce unnatural scaling effects resulting in perspective distortions. As shown in equation 2.2, there exists a non linear relationship beween x and x' (similarly y and y'). Since H is calculated from corresponding points and extended to non-overlapping regions, resulting in unnarural scaling. The amount of distortion can be reduced by projecting image on a cylindrical plane as shown in figure 2.7.

$$x' = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}} \quad y' = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}} \tag{2.2}$$

(a) Rectilinear Projection   (b) Cylindrical Projection

Figure 2.7: Effect of compositing surface

- The performance of local homography is sensitive to parameters γ and σ. Careful selection of these parameters is vital for obtaining best results.

- The isotropic nature of Gaussian weighting function results in 'wavy' effects in final stitched image, observed when γ values are very small.
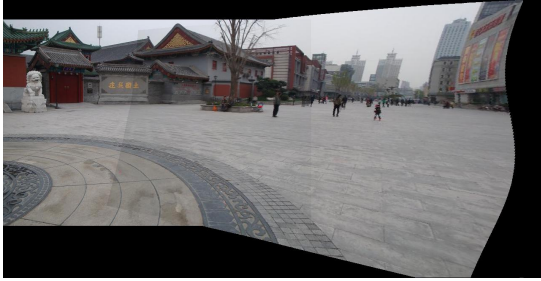

Figure 2.8: 'Wavy' effects

CHAPTER 3
Shape Preserving Projective Warp

Local homography performs well in the overlapping regions, but introduces projective distortions in non-overlapping parts resulting in unnatural stretching of shapes and enlarging of sizes. Authors of [8] propose combination of projective and similarity warps to overcome distortion. Similarity warp is a subclass of projective warp, which is only composed of translation, scaling and rotation as shown in equation 3.1 and thus introduces no shape distortion nor non-linear scaling. Also similarity warp is invariant to ratio of lengths and angle, making a better candidate than homography. Similarity warp can be interpreted as a combination of panning, zooming and camera in plane rotation.

$$S = \begin{pmatrix} s\cos\theta & -s\sin\theta & t_x \\ s\sin\theta & s\cos\theta & t_x \\ 0 & 0 & 1 \end{pmatrix} \tag{3.1}$$

3.1 Proposed Warp

Let f be the proposed warping function $f : \mathbb{R}^2 \to \mathbb{R}^2$ that maps $\mathbf{x} = [x, y]^T$ and $\mathbf{x}' = [x', y']^T$. $\mathbb{R}^2$ is divided into two half-spaces R_H and R_L (They correspond to overlapping and non-overlapping regions respectively). For $(x, y) \in R_H$, local homography $\mathbf{H}_*$ is used. For other half space R_L, the warping function continuously extrapolates $\mathbf{H}_*$ to become a similarity transformation S. Let (x, y) be the co-ordinates obtained from rotation of (u, v) by an angle θ.

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} \tag{3.2}$$

After the change of co-ordinates, the homography equation 1.1 becomes

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} \sim \begin{pmatrix} h_{11}\cos\theta + h_{12}\sin\theta & -h_{11}\sin\theta + h_{12}\cos\theta & h_{13} \\ h_{21}\cos\theta + h_{22}\sin\theta & -h_{21}\sin\theta + h_{22}\cos\theta & h_{23} \\ h_{31}\cos\theta + h_{32}\sin\theta & -h_{31}\sin\theta + h_{32}\cos\theta & h_{33} \end{pmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \tag{3.3}$$

Equation 3.3 is valid upto a scale factor. Hence we can choose $h_{33} = 1$. Choosing $\theta = \tan^{-1}\frac{h_{32}}{h_{31}}$, equation 3.3 can be simplified to

$$\begin{aligned} x' &= \frac{h'_{11}u + h'_{12} + h'_{13}}{1 - cu} \\ y' &= \frac{h'_{21}u + h'_{22} + h'_{23}}{1 - cu} \end{aligned} \tag{3.4}$$

$$\text{Where } \begin{pmatrix} h'_{11} & h'_{12} \\ h'_{21} & h'_{22} \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$
$$(h'_{13}, h'_{33}) = (h_{13}, h_{33}) \tag{3.5}$$
$$c = \sqrt{h_{31}^2 + h_{32}^2}$$

3.1.0.7 Linearity property

The denominator term in equation 3.4 is dependent only on single parameter u as opposed to equation 2.2 which depends on two parameters. Under certain conditions (constant u) equation 3.4 results in a linear transformation which preserves ratio of lengths and producing no shape distortion.

Let $R_H = \{(u,v)|u \leq u_1\}$ and $R_L = \{(u,v)|u > u_1\}$ where u_1 is the line partitioning $\mathbb{R}^2$. For $(u,v) \in R_H$, homography $\mathbf{H}_*$ is applied and for $(u,v) \in R_L$, similarity S is applied. For continuity it is required to satisfy $S(u_1,v) = \mathbf{H}_*(u_1,v)$. This warp is only C^0 continuous, as it is not differentiable at u_1, which results in bending of the image at u_1 .
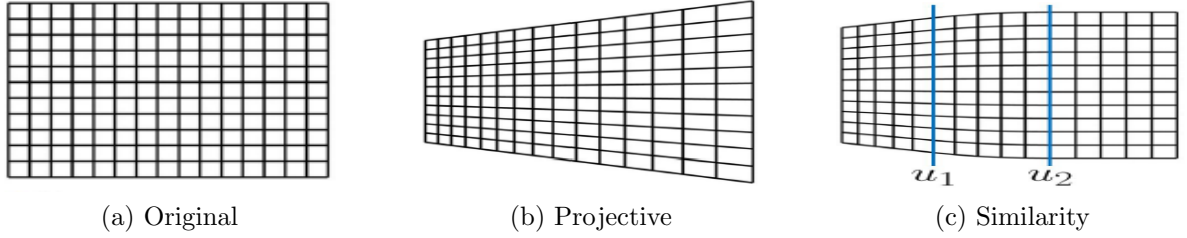

(a) Original   (b) Projective   (c) Similarity

Figure 3.1: Comparison of various warps. Image taken from [8]

3.1.0.8 C^1 extrapolation

To overcome distortions produced in C^0, a genralized warp in C^1 is proposed. Let R_L be divided into two regions $R_T = \{(u,v)|u_1 < u < u_2\}$ and $R_S = \{(u,v)|u_2 \leq u\}$, where u_1 and u_2 are parameters of warping. The warping function f is given by

$$f(u,v) = \begin{cases} \mathbf{H}_*(u,v) & \text{if } (u,v) \in R_H \\ T(u,v) & \text{if } (u,v) \in R_T \\ S(u,v) & \text{if } (u,v) \in R_S \end{cases} \tag{3.6}$$

Here T acts as a buffer when homography is gradually changed to similarity. For simplicity T and S are parametrized as

$$S(u,v) = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \qquad T(u,v) = \begin{pmatrix} f_x(u) \\ f_y(u) \end{pmatrix} v + \begin{pmatrix} g_x(u) \\ g_y(u) \end{pmatrix} \tag{3.7}$$

The parameters of S, α, β, t_x, t_y are obtained by invoking continuity and differentiability property of f at u_1 and u_2. Functions $f_x(u), f_y(u), g_x(u), g_y(u)$ are assumed to be quadratic whose parameters are obtained in a similar fashion as that of S. The goal of this warping method is to mintain perspective of each image with the warping function approaching similarity transformation. In order to achieve this every image I_i is associated with a cost E_i as a function of u_1 and u_2 which measures the deviation of warp from the nearest similarity in Frobenius norm

$$E_i(u_1, u_2) = \min_{a_i, b_i} \int\int_{(x,y)\in\Omega_i} \left\| J_i(x,y,u_1,u_2) - \begin{pmatrix} a_i & -b_i \\ b_i & a_i \end{pmatrix} \right\|_F^2 dx \, dy \tag{3.8}$$

12

Where Ω_i is domain of I_i, $J_i(x, y, u_1, u_2)$ is the Jacobian matrix of f evaluated at (x, y) which will be a function of u_1 and u_2. The optimal value of u_1 amd u_2 is obtained by minimizing E_i. In case of multiple images total energy is minimized.

3.2 Results

Local homography and shape preserving warp are compared in figure 3.2 Shape preserving warp removes the wavy distortion introduced by local homography resulting in more natural looking panorama. Note that RMSE doesn't change as compared to local homography, as the warp corresponding to overlap region is not changing in shape preserving warp.

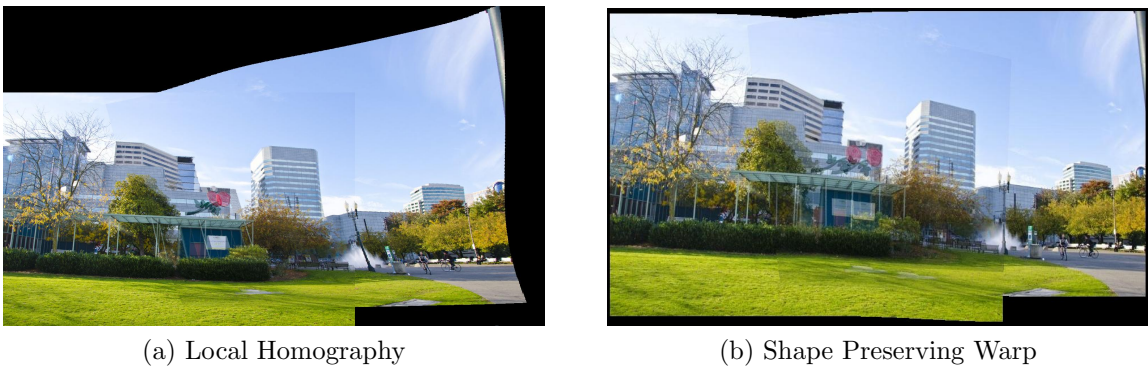(a) Local Homography

(b) Shape Preserving Warp

Figure 3.2: Comparison of local homography and shape preserving warp

In certain images there appears to be a tradeoff between in performance in overlaping and non-overlapping regions. Extensive similarity warping results in blurring of the image in overlapping regions as observed in figure 3.3.

(a) Local Homography

(b) Shape Preserving Warp
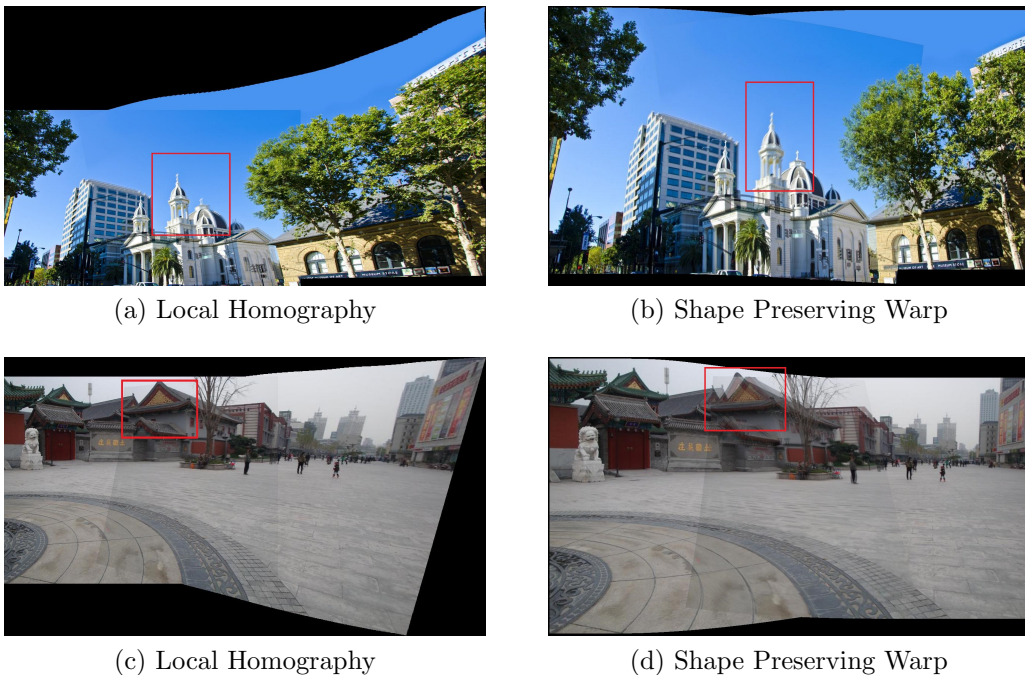
(c) Local Homography

(d) Shape Preserving Warp

Figure 3.3: Blurring in overlapped regions

Cylindrical projection also reduces projective distortion in non-overlaping regions as stated in the previous section. Figure 3.4 compares cylindrical projection and shape preserving warp. It can be inferred from the figure that cylindrical projection does not reduce projective distortion always. However in certain cases cylindrical projection and shape preserving warp performance is similar.
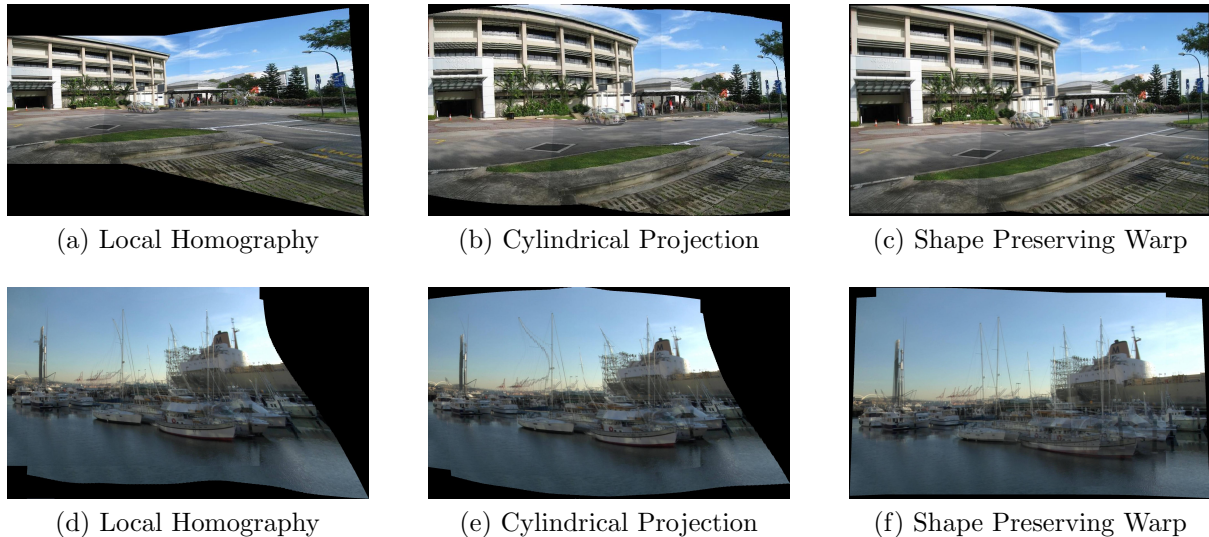

| (a) Local Homography | (b) Cylindrical Projection | (c) Shape Preserving Warp |

| (d) Local Homography | (e) Cylindrical Projection | (f) Shape Preserving Warp |

Figure 3.4: Comparison of various warps

3.3 Conclusion

In this project local homography based image warping methods as applicable to image stitching was studied. Local homography performed better than global homography, particularly in overlapping regions. However local homography suffered from perspective distortions in non-overlapping parts. To overcome this similarity warp was proposed in non-overlapping parts to preserve shape and size of objects observed in the image. Also extensive similarity warps in certain images lead to blurring in overlapping parts.

In majority of the images shape preserving warp method performed well resulting in natural looking panoramas. The distortions introduced by this method is reasonably low and such distortions can be overcome by employing suitable blending algorithms.

REFERENCES

[1] Lowe, David G. "Distinctive image features from scale-invariant keypoints." International journal of computer vision 60.2 (2004): 91-110.

[2] Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. "Surf: Speeded up robust features." Computer visionECCV 2006 (2006): 404-417.

[3] Fischler, Martin A., and Robert C. Bolles. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography." Communications of the ACM 24.6 (1981): 381-395.

[4] Szeliski, Richard. "Image alignment and stitching: A tutorial." Foundations and Trends in Computer Graphics and Vision 2.1 (2006): 1-104.

[5] Zaragoza, Julio, et al. "As-projective-as-possible image stitching with moving DLT." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013.

[6] Hartley, Richard, and Andrew Zisserman. Multiple view geometry in computer vision. Cambridge university press, 2003.

[7] Vedaldi, Andrea, and Brian Fulkerson. "VLFeat: An open and portable library of computer vision algorithms." Proceedings of the 18th ACM international conference on Multimedia. ACM, 2010.

[8] Chang, Che-Han, Yoichi Sato, and Yung-Yu Chuang. "Shape-preserving half-projective warps for image stitching." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014.