Introduction

April 23, 2011

The thesis focus on the precision aspect of the chain of 3D reconstruction. The origin of imprecision can lie at any step of the chain. The imprecision caused in a certain step will make useless the precision gained in the previous steps, then be propagated, amplified or mixed with the error in the following steps and finally leads to an imprecise 3D reconstruction. It is difficult to directly improve the precision from final imprecise 3D data. The appropriate approach to obtain a precise 3D model is to study the precision of every component.

We pay more attention to the camera calibration for three reasons. First, it is often the first component in the chain. Second, it is by itself already a complicated system containing many unknown parameters. Third, the intrinsic parameters of camera only need to be calibrated once depending on the camera configuration. In addition, the camera calibration problem is supposed to have been solved since years. But the result is still not satisfying if high-precision is required. The global camera methods can leave residual distortion error as big as one pixel, which can lead to distorted reconstructed scene. We propose two methods in the thesis to correct the distortion with high-precision. With an objective evaluation tool, it is shown that the finally achieved correction precision is about 0.02 pixels.

Precision is also needed in the other image processing tasks like image registration. In contrast to the advance in the invariance of feature detectors, the matching precision has not been studied carefully. We analyze the SIFT method (Scale-invariant feature transform) [10] and evaluate its matching precision. It is shown that by some simple modifications in SIFT scale space, the matching precision can be improved to be about 0.05 pixels in the synthetic tests. A more realistic algorithm is also proposed to increase the matching precision for two real images between which the transformation is locally smooth. A multiple-image denoising method, called "burst denoising", is then proposed to take advantage of precise image registration.

1 Context

This thesis is integrated into the scope of project CALLISTO (Calibration en vision stro par mthodes statistiques) sponsored by ANR (Agence Nationale de la Recherche), whose final aim is to reconstruct 3D scene in high precision. This project relies on the collaboration between different universities and *grandes coles*: CMLA-ENS-Cachan, IMAGINE-ENPC, MAP5-Paris 6 and LCTI-Tlcom Paris. Even though this project mainly concerns about the images taken by consumer digital cameras, the result can be extended to the satellite images. The work thus helped to some extent the MISS (Mathmatiques de l'imageries stroscopique spatiale) project collaborated with CNES (Centre National d'Etudes Spatiales), whose aim is to reconstruct a completely controlled and reliable 3D terrain model from two images taken by the air-bone satellite camera almost simultaneously.

The 3D scene reconstruction can be mainly divided into five components: camera calibration, image rectification, dense image registration, 3D scene reconstruction and 3D merging, meshing and rending (Fig. 1). Some components can be replaced by other techniques to make the chain more adapted for some specific applications.

The camera calibration is the first step in the chain and thus plays a very important role. It consists of camera internal/external parameters calibration and distortion model estimation. Camera internal parameters means the intrinsic parameters of camera, like aspect ration, focal length, principal point and lens distortion parameters. Camera external parameters means camera orientation and position in a fixed world coordinate. And the distortion model describes the geometric deviation of real camera from a pinhole camera.

Image rectification is an auxiliary step for dense image registration. It virtually rotates two cameras about their optic center respectively such that the two cameras planes are co-plane and their x-axis are parallel to the baseline. This generates two images whose corresponding epipolar lines coincide and are parallel to the x-axis of image. A pair of rectified images is helpful for dense stereo matching algorithms. It restricts the search domain for each matching to a line parallel to the x-axis. Due to the redundant degrees of freedom, the solution to rectification is not unique and actually can lead to undesirable distortions or be stuck in a local minimum of the distortion function. This motivated us to propose a robust three-step image rectification in Chapter **??**.

Dense image registration is the thesis subject of Neus Sabater [12], whose aim is to find dense correspondences between two images. This is a difficult problem by considering different imaging conditions when taking images, like different geometric distortion depending on the viewing angle, non-linear lens distortion of camera, changing lighting condition, non-static scene, occlusion, etc. These problems can be to some extent reduced by using a pair

of image with "low B/H" (low baseline/height), which are taken by a satellite almost simultaneously. But this raises a higher demand on the precision on the correspondences. A thorough discussion about how to obtain dense correspondences with emphasis on the control of false alarms, sub-pixel precision and the fattening problem at the contour of image can be found in Neus Sabater's thesis [12].

Once camera is precisely calibrated and a pair of images is accurately and reliably registered, a 3D model of scene can be easily reconstructed on the same order of precision with some classic methods up to a 3D similarity transformation. One pair of images only permits us to reconstruct a partial 3D model. To have a more complete model, it is necessary to take photos with different angles of view around the 3D scene. The pair-wise 3D scene model can be overlapped or completely disjoint. By merging many pair-wise partial 3D models, a dense 3D point cloud can be obtained.

The chain seems complete with all the above components. But the reconstructed 3D scene is just a set of 3D points in the space. For the sake of visualization, it is better to reconstruct also the surface of objects in the scene, in particular for purpose of parts inspection and repairing in industry. Since the precision of 3D points is high, it is hoped that the precision will be kept in the surface reconstruction. This problem is largely discussed in Julie Digne's thesis [2], which treats high precision scanned raw data point sets with up to 35 million points, usually made about 300 different scan sweeps. Even though in her work, the 3D points directly come from the laser scanner instead of from 3D reconstruction of images taken by camera, the principle of problem remains the same: how to reconstruct the surface reliably by keeping the high precision without smoothing and re-sampling ?

Figure 1: The 3D reconstruction chain.

2 Background

The 3D scene reconstruction is not a new subject in computer vision. Before the "computer" and digital camera came into history at the end of year 1970s, the scene reconstruction was already a classic problem in photogrammetry, where it has a different name "stereophotogrammetry". Its aim is always to determine the geometric properties of objects from photographes. At that time, more attention was paid to the methods in optics and metrol-

ogy due to the lack of computational power. The distances and angles are directly measured manually from photographs, objects in scene and cameras separated by a fixed baseline. The precision is not ensured in the measurement, which can lead to inaccurate final result. The lack of imprecision of manual work limited the practical application of early photogrammetry techniques.

With the advent of the digital camera and high-performance computers, the 3D reconstruction becomes more mature with fewer or no manual measurement. The prompt perhaps comes from the remarkable advance in camera calibration techniques. The famous calibration methods, like Zhang's flexible pattern based method [13], Lavest *et al*'s method [9], Devernay and Faugeras' plumb-line method [4], Hartley's pattern-free method [5], make it possible to calibrate camera and reconstruct 3D scene on site with few human intervention. Nonetheless, 3D reconstruction is still a complicated chain including several components. Any error in a certain component will make the whole system unreliable.

Another possibility to obtain 3D reconstruction is to use 3D laser or LIDAR (LIght Detection And Ranging). Not like the above multi-view geometric method, this kind of methods is active. The position of object in the space is measured by the time delay between transmission of a pulse and detection of the reflected signal. Thus a dense 3D points can be directly obtained by some post-processing like filtering and merging. The traditional idea is that 3D stereo reconstruction is not as precise as the result obtained by 3D laser scanner. A high quality 3D scanner can provide 3D point cloud with the precision about 20 μm. This cannot be achieved by state-of-art image-based 3D stereo reconstruction algorithm. Yet, 3D laser scanner is usually a big, expensive and sophisticated machine, which have to be set up carefully and cannot be easily transported for onsite 3D reconstruction tasks, while a camera, even a reflex camera of good quality, costs almost nothing compared to a scanner machine and can be flexibly transported anywhere to take photos (Fig. 2). In addition, it is not feasible to install a huge laser scanner on satellite to do 3D reconstruction. So it is still of interest to use camera photos to reconstruct 3D scene if the precision can achieve or surpass that of 3D scanner.

3 Challenge

To have a precise final 3D reconstruction from 2D photos, each component should produce a precise result. This gives the first importance to the camera calibration because its imprecision will be inevitably propagated to the other components and usually cannot be corrected later. With a precise camera calibration, it is already sufficient to determine the relative position of two cameras and reconstruct a sparse model of 3D scene from the

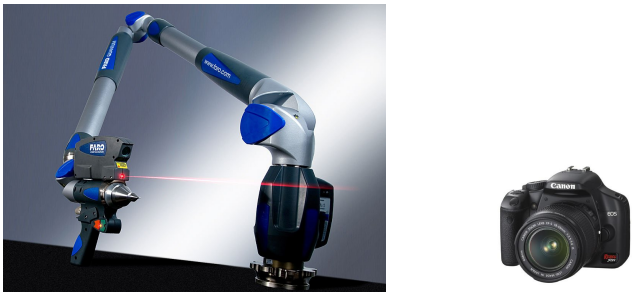


Figure 2: Left: 3D laser scanner. Right: Canon EOS Reflex camera.

correspondences between two images. A typical camera model has the form:

$$C = \mathcal{D}KR[I| - T] \tag{1}$$

with $\mathcal{D}$ non-linear operator of lens distortion, K calibration matrix, R camera orientation matrix and vector T the camera optic center in a fixed world frame. Given a 3D point, it is first transformed into the camera-based frame by the translation T then the rotation R. Then it is projected onto the image plane by K, followed by the non-linear lens distortion represented by $\mathcal{D}$. According to this model, camera calibration consists of two parts: internal parameters calibration ($\mathcal{D}$ and K) and external parameters calibration (R and T). The internal calibration is to retrieve the intrinsic parameters of camera, which should be constant once the camera has fixed its configuration. But in the experiments, it shows that the internal parameters varies from one experiment to the other even if the same camera was used with fixed configuration. This makes the internal calibration unreusable for the other data sets. Similarly, for the external calibration, the camera position and orientation also varies by using two data sets sharing a common camera position. This leads us to think over again the whole camera calibration system. The conjecture about the phenomenon can be two-fold:

1. the distortion model cannot capture the real physical aspect of real distortion

2. the error in external calibration for R and T compensates the error in internal calibration for $\mathcal{D}$ and K. So the whole camera model can be precise in the sense that the observed point is close to the point computed by the camera model. But none of the components ($\mathcal{D}$, K, R, T) is precise.

This is in fact the common drawback of many global camera calibration methods, which perform the internal and external calibration together. Typical global calibration algorithms were tested to verify the problem. In particular, Lavest *et al.* algorithm [9] shows a very small re-projection error

about 0.02 pixels, which confirms the precision of the global camera model. But the residual distortion error can be 10 times bigger. Since Lavest *et al.* algorithm considers also the non-flatness of the pattern and estimate its shape, the factor of 10 can only be explained by the error compensation.

4 Virtual Pinhole Camera

The first aim of camera calibration is to recover a pinhole camera by correcting the distortion. But as explained, due to the error compensation, it is risky to use global calibration methods to correct lens distortion. The error compensation can be avoided by separating the distortion correction from the global calibration. So the camera calibration is decomposed into two steps. The first step is to correct lens distortion to obtain a pinhole camera; the second step is to calibrate the pinhole camera. As for the pinhole camera, the fundamental theorem must be cited [6, 4]:

Theorem 1 *A camera follows the pinhole model if and only if the projection of every line in space onto the camera is a line.*

This theorem can be understood in three different ways. First, the distortion can be corrected by rectifying the distorted lines in image. Second, the distortion correction can be evaluated by measuring the straightness of corrected lines in image. Third, the pinhole camera is not unique because a straight line in image remains to be straight under a 2D homogrpahy. This means the distortion is corrected up to an arbitrary homography. Assume the arbitrary homography is H, then the estimated distortion is $\tilde{\mathcal{D}} = DH$ with D the absolute distortion introduced by camera lens system. By applying the inverse of $\tilde{\mathcal{D}}$ on camera, we obtain a new camera noted by $\tilde{C}$:

$$\tilde{C} = \tilde{\mathcal{D}}^{-1}C = H^{-1}\mathcal{D}^{-1}\mathcal{D}KR[I| \ -T] = H^{-1}KR[I| \ -T]. \qquad (2)$$

H, K being invertible, the decomposition $H^{-1}K = \tilde{K}R'$ is unique by QR decomposition with the constraint that $\tilde{K}$ is an upper-triangle 3×3 matrix and R' is a 3×3 rotation matrix. This new camera becomes $\tilde{C} = \tilde{K}R'R[I| \ -T] = \tilde{K}\tilde{R}[I| \ -T]$. We call it *mathematic (or virtual) camera after distortion correction* because the calibration matrix $\tilde{K}$ and rotation matrix $\tilde{R}$ do not match the physics of the actual camera, but yield a virtual pinhole camera that can be used to the very same purposes. Indeed, consider several positions of the physical camera inducing as many camera models $C_i = DKR_i[I| \ -T_i]$. Applying the correction $\mathcal{D}^{-1}$ to all images obtained from these camera positions yields virtual pinhole cameras $\tilde{C}_i = \tilde{K}\tilde{R}_i[I| \ -T_i]$, which maintains the same relative orientations: $\tilde{R}_i^{-1}\tilde{R}_j = R_i^{-1}R_j$. From these cameras the whole 3D scene can be reconstructed by standard methods, up to a 3D similarity.

5 Image Registration and Denoising

Invariance and precision are two key problems in image registration. Invariance means whether an image matching algorithm can find reliable correspondences under critical geometric or photometric transformations. Precision means whether the matchings between two images are precise. The two problems are crucial for the success of many applications, like super-resolution, image mosaicing, camera calibration, etc. Many efforts have been recently dedicated to obtaining feature detectors more invariant to geometry or photometric transformations, while the matching precision of feature points is always considered enough and has not been carefully studied. In fact, some state-of-art feature detectors combining a robust descriptor give enough invariance for many applications. So it is the time to evaluate and improve the matching precision. We are interested in the matching precision of SIFT [10], a popular scale-invariant feature detector. The matching precision of SIFT method is studied and improved. We show that the matching precision can achieve the precision better than 0.05 pixels if the transformation between two images is locally smooth.

Precise image registration inspires us to invent a new image denoising algorithm, called "burst denoising". Basically, this algorithm aligns a burst of images to a reference image and performs the average operation to reduce the noise level. The average is perhaps the only operation which can preserve the fine details in images. This method is extended to be a mixed algorithm by combining the block denoising when the two images are not related by a rigid transformations.

6 Chapter by Chapter

The thesis is summarized chapter by chapter as follows.

Chapter ??: Camera Model and Projective Geometry

Many algorithms in multi-view geometry are based on the assumption that the camera is ideal pinhole. But in practice, the camera is deviated from a pinhole model by lens distortion. In this chapter, some basic concepts about pinhole camera model and distortion model are introduced. A typical bundle adjustment camera calibration method is also explained in detail to show the problems we shall meet in distortion correction.

Once the distortion is removed and the camera becomes pinhole, the projective geometry is a useful tool to solve different problems in multi-view geometry, like image rectification and mosacing. The geometric constraint becomes more complicated when the number of views increases. Here we concentrate on the simplest two-view geometry because it is difficult for three or more images to share enough stable feature points in practice. The

relation between corresponding points in two views is described by epipolar geometry. In algebraic viewpoint, the epipolar geometry is coded by a 3×3 matrix, called fundamental matrix, F. Fundamental matrix only depends on the relative position of two cameras and the intrinsic parameters of cameras, but not on the 3D scene. Two important observations of epipolar geometry is:

- Given one point x in the left image, its corresponding point in the right image must be on the line called epipolar line, which can be explicitly computed as Fx. This provides a necessary condition to test whether two points correspond to the same 3D point, see Fig. 3.
- Given a set of corresponding points in two images, the 3D scene can be reconstructed up to a 3D projective transformation, as well as the camera position and orientation.

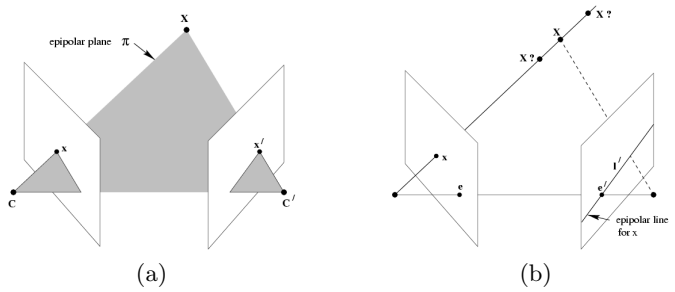


Figure 3: (a) The optic center of two cameras and a 3D point form a epipolar plane which intersects two image planes at the epipolar line. (b) Given one point x in the left image, its correspondence in the right image must be on the corresponding epipolar line.

Chapter ??: Calibration Harp: A Measurement Tool of Lens Distortion Correction Precision

A measurement tool of lens distortion correction precision is introduced in this chapter. Lens distortion is a non-linear deformation which deviates a pinhole camera from central projection. The alignment is the only property preserved in the central projection. So it is reasonable to measure the straightness of the projection of 3D straight lines to evaluate the lens distortion correction precision. To have a precise evaluation in practice, we need some very straight strings of good quality. It is relatively easy to ensure the straightness by tightly stretching the strings and attaching them on a frame, while it is more delicate to choose an appropriate type of string. We tried four types of strings and found that the opaque fishing string is

the best choice for our purpose. An evaluation pattern made up of several parallel tightly stretched opaque fishing strings with a translucent paper as background, called "calibration harp" is thus built (see Fig. 4). The Devernay sub-pixel precision edge detector [1] is used to extract the edge points in image, which are then associated to the line segments detected by LSD (Line Segment Detector) [11]. Finally, the distortion correction is evaluated as the root-mean-square (RMS) distance from the edge points belonging to a same line segment to their corresponding linear regression line.

(a) (b)

Figure 4: (a) The harp made up of opaque fishing strings with a translucent paper as background. (b) A close-up of (a).

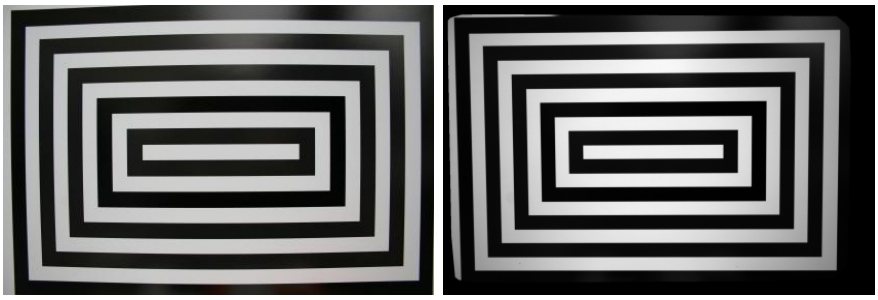
Chapter ??: Non-parametric lens distortion correction

This chapter presents a first attempt to correct the distortion in high precision. By high precision, we mean that the residual error between the camera and its numerical model obtained by calibration should be far smaller than the pixel size. At first sight, this problem seemed to have been solved adequately by recent global calibration methods. The celebrated Lavest *et al.* method [9] measures the non-flatness of a pattern and yields a remarkably small re-projection error of about 0.02 pixels, which outperforms the precision of other methods. For the goals of computer vision, this precision would be more than sufficient. Yet, this paper describes a seriously discrepant accuracy measurement contradicting this hasty conclusion. According to the measurement tool of distortion correction precision developed in Chapter **??**, the only objective and correct criterion is straightness of corrected lines.

Following this tool, the accuracy criterion used herewith directly measures the straightness of corrected lines. We shall see that this straightness criterion gives a RMSE as big as 0.2 pixel, which contradicts the 0.02 pixel reprojection accuracy. This significant discrepancy means that, in the global optimization process, errors in the external and internal camera parameter are being compensated by opposite errors in the distortion model. Thus, an inaccurate distortion model can pass undetected. Such facts raise a solid

objection to global calibration methods, which estimate simultaneously the lens distortion and the camera parameters. This chapter reconsiders the whole calibration chain and examines an alternative way to guarantee a high accuracy. A useful tool toward this goal will be proposed and carefully tested. It is a direct non-parametric, non-iterative, and model-free distortion correction method. By non-parametric and model-free, we mean that the distortion model allows for any diffeomorphism.

This non-parametric method requires a flat and textured pattern. By using the dense matchings between the pattern and its photo, a distortion field can be obtained by triangulation and local affine interpolation. The obtained precision compares favorably to the distortion given by state of the art global calibration and reaches a RMSE of 0.08 pixels (see Fig. 5 for a correction example). The non-flatness of the pattern is a limitation of this method and can introduce a systematic error in the distortion correction. Nonetheless, we also show that this accuracy can still be improved in the next two Chapters.



(a) (b)

Figure 5: A correction example. (a) distorted image. (b) corrected image.

Chapter ??: Self-Consistency and Universality of Camera Lens Distortion Models

This Chapter is a preparation for the next chapter. Due to the difficulty to control the flatness of a pattern, we will go back to the parametric method in the next chapter to obtain a higher correction precision. For any parametric method, an appropriate distortion model is necessary. Even though there exists many distortion models in literature, it is not clear which one is more appropriate than the others. In addition, the role of distorted points and undistorted points seems to be interchangeable in literature, which makes the distortion model more ambiguous.

In this chapter, the concepts of "self-consistency" and "universality" are introduced to evaluate the validity and precision of camera lens distortion models. Self-consistency is evaluated by the residual error when distortion

generated with a certain model is corrected (using the model in reverse way) by the best parameters for the same model. Analogously, universality is measured by the residual error when a model is used to correct distortions generated by a family of other models.

Five classic camera lens distortion models are reviewed and compared for their degree of self-consistency and universality. The realistic synthetic experiments show that the polynomial model is self-consistent and more universal than the other models. The polynomial model, with order from 8 to 19, permits to approximate any other four models, and the inverse of any other four models including itself, at the precision about 1/100 pixels. This high order is more than compensated by its linearity and its translation invariance, which makes it independent of the distortion center. A real experiment shows that the polynomial model of degree 6 can approximate a real distortion field between a textured pattern and its photo at 1/100 pixel precision (see Fig. 6). So the polynomial model will be chosen in the next chapter as distortion model to improve the correction precision.

Chapter ??: High Precision Camera Calibration with a Harp

This chapter is a continuation of the last chapter to improve the precision of lens distortion correction. Even though non-parametric pattern based method do not depend on the a priori choice of a distortion model with a fixed number of parameters, to achieve a high precision, they require a very flat non deformable plate with highly accurate patterns printed on it. It is shown that $100\mu m$ flatness error can introduce 0.3 pixels error in distortion estimation. Either to fabricate such pattern or to estimate its shape is very difficult in practice. Perhaps the thing we can easily ensure the quality is the straightness of strings. That is the reason why we resort to the plumb-line method to correct the distortion.

Based on the well-known fact that a camera follows the pinhole model if and only if the projection of every line in space onto the camera is a line, it is sufficient to correct the distorted lines to obtain the pinhole camera. The "calibration harp" built in Chapter **??** to evaluate the precision of distortion correction can be directly used here. But this time, it is used as both a tool of distortion correction and a validation tool. It seems that the hardware problem of plumb-line method is solved. But we still need a good distortion model to integrate into the plumb-line method to treat different types of realistic lens distortion. According to the test of the self-consistency and universality of different models, the polynomial model seems more adapted to correct real distortion. In addition, it is invariant to the translation to the distortion center. So the distortion center can be fixed anywhere without being estimated. This is a big advantage compared to other models.

Photos of different orientations are taken to estimate the best coefficients of polynomial model to correct the distortion (see Fig. 7). Real experiments

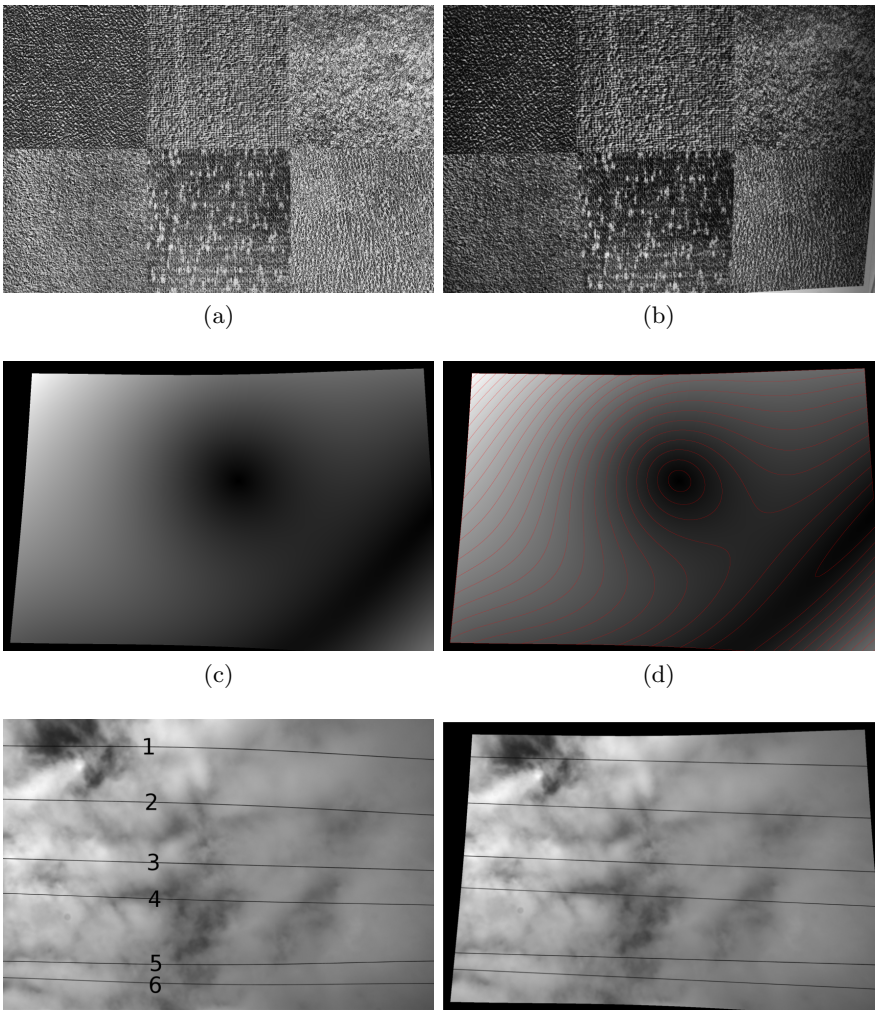


Figure 6: (a) digital pattern. (b) a photo of the digital pattern. (c) the distortion field constructed by the estimated parameters of the polynomial model. (d) level lines of (c) with quantization step of 20. (e) distorted images of tightly stretched lines. (f) corrected image by polynomial model.

show that no artificial bias is created on the corrected strings and higher precision is attained compared to non-parametric pattern based method. Both the harp of sewing strings and the harp of opaque fishing strings are tested in the experiments. With the harp of sewing strings, the correction precision is better than the non-parametric pattern based method and no global artificial bias is observed. With the harp of opaque fishing strings, the residual oscillation due to braid pattern of sewing strings is largely reduced (see Fig. 8). We do gain a factor about 2 over sewing strings harp and achieve the average correction precision about 0.02 pixels This precision is much better than the result given the global camera calibration, which is not stable and varies with the parameters used in the distortion model.

Chapter ??: Three-Step Image Rectification

The epipolar geometry enters into a special situation when the two camera planes are co-plane and parallel with the line connecting the optic center of two cameras. In this case, the corresponding epipolar lines also coincide and align with the x-axis of the image. This means that one point has the same y-coordinate as its corresponding point. This special geometry can be achieved by rotating two camera without changing their optic center. This is equivalent to apply a homography on each image respectively. This process is called image rectification in two-view geometry (Fig. 9 shows a pair of images before and after rectification). A pair of stereo-rectified images is helpful for dense stereo matching algorithms. It restricts the search domain for each match to a line parallel to the x-axis. Due to the redundant degrees of freedom, the solution to stereo-rectification is not unique and actually can lead to undesirable projective distortions or be stuck in a local minimum of the distortion function. Many rectification methods reduce the distortion by different explicit measure. But it is not clear which measure is the most appropriate. We propose a rectification method by three steps of camera rotation. In each step, the distortion is explicitly reduced by minimizing the rotation angle. For un-calibrated cameras, this method can be formulated as an efficient minimization algorithm by optimizing only one natural parameter, the focal length. This is in contrast with many methods which optimize between 3 and 6 parameters.

Chapter ??: Matching Precision of SIFT

Image features detection and matching is a fundamental step in many computer vision tasks. Many methods have been proposed in recent years, with the aim to extract image features fully invariant to any geometric and photometric transformation. Even though the state-of-art has not achieved the full invariance, many methods, like SIFT [10], Harris-affine [8] and Hessian-affine [7] combining a robust and distinctive descriptor, give sufficient in-

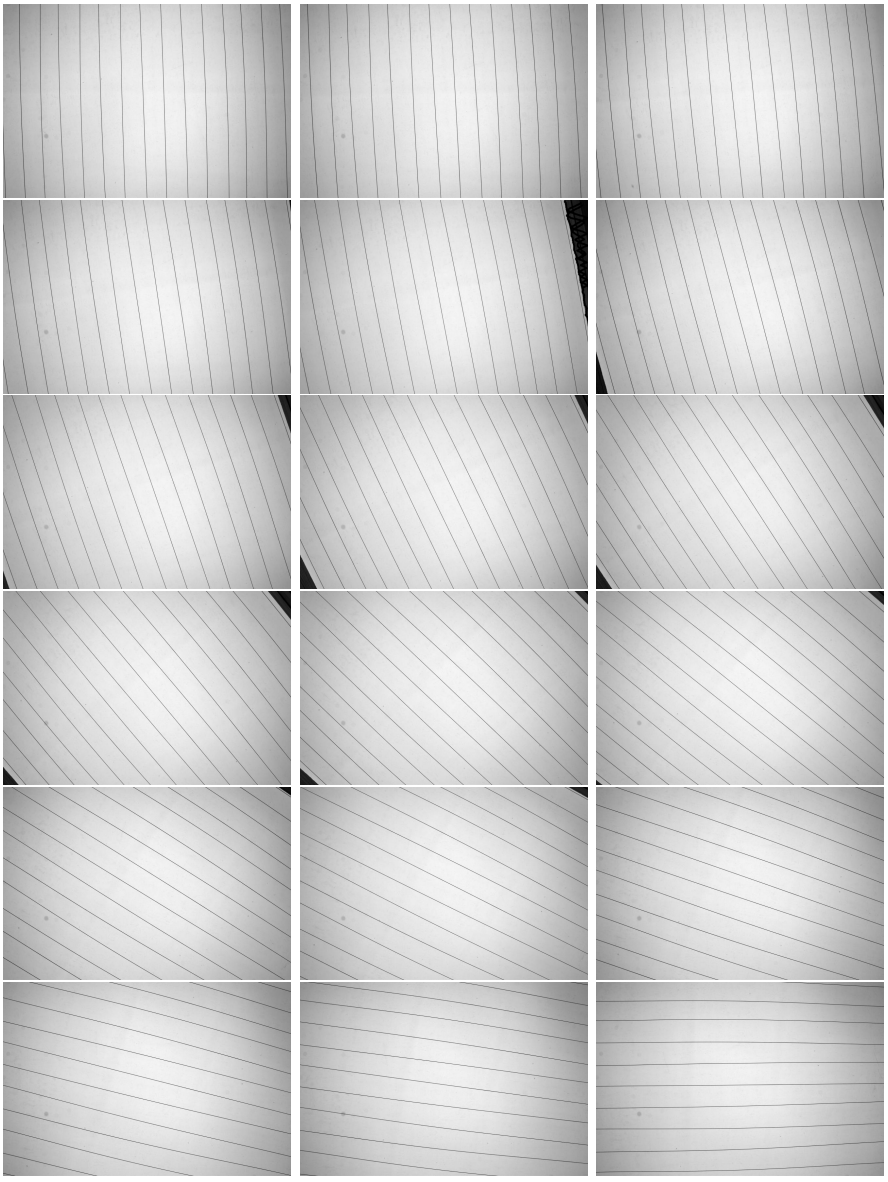
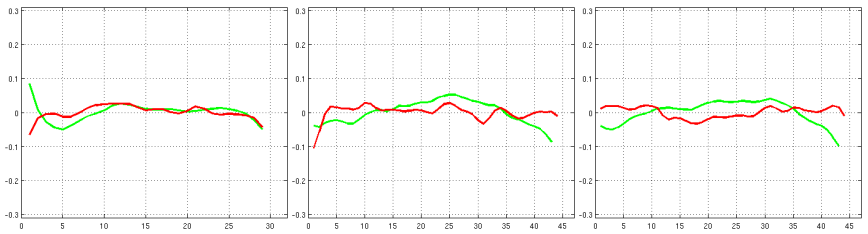
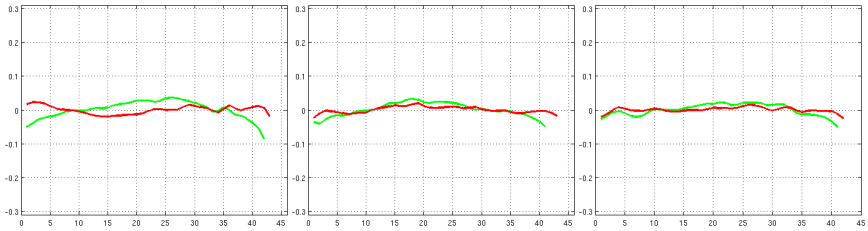


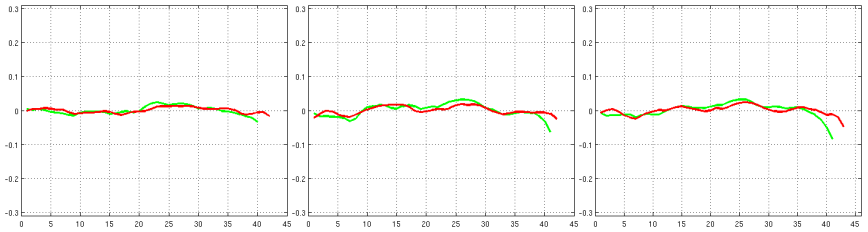
Figure 7: Distorted fishing strings taken by the camera fixed on a tripod with different orientations.



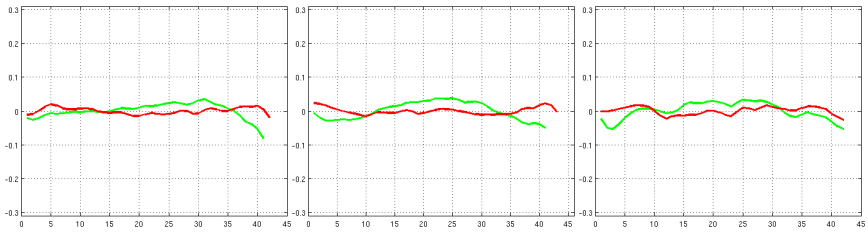
(a) 0.027/0.020 (b) 0.034/0.022 (c) 0.034/0.016



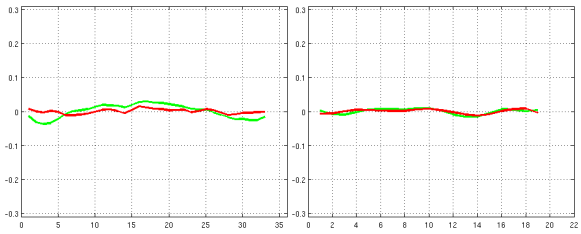
(d) 0.027/0.012 (e) 0.021/0.010 (f) 0.018/0.008



(g) 0.012/0.008 (h) 0.020/0.012 (i) 0.021/0.013

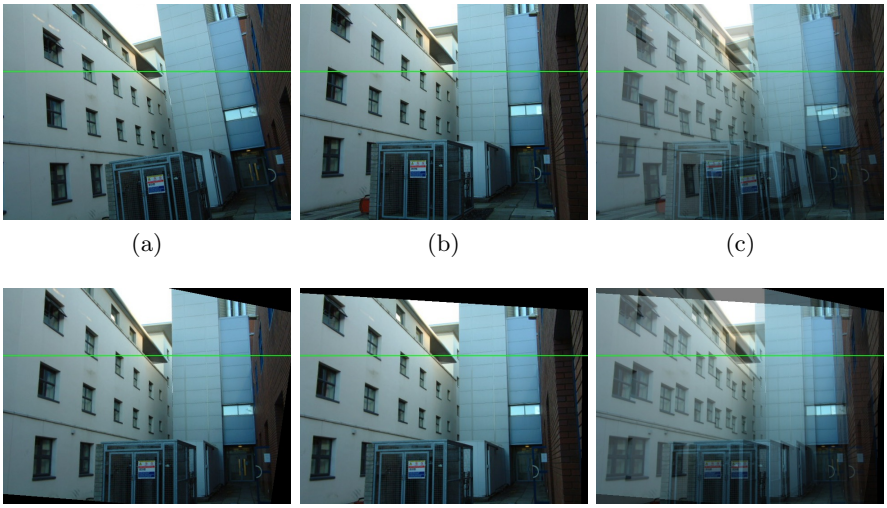


(j) 0.023/0.009 (k) 0.026/0.010 (l) 0.025/0.012



(m) 0.020/0.007 (n) 0.008/0.006

Figure 8: Correction performance of the proposed plumb-line method with a harp made up of fishing strings. The distance (in pixels) from the edge point to the corresponding linear regression line is plotted. The x-axis is the index of edge points. The range of y-axis is from -0.3 pixels to 0.3 pixels. The straightness error (in pixels) measured as root mean square distance from the edge points to their linear regression line is just below each figure. Note that each figure contains two curves because there are two sides for one string. The camera focal length is fixed 55 mm and the distance between camera and object is about 100 cm.



(a) (b) (c)

(d) (e) (f)

Figure 9: Rectification example. (a) and (b): two original images. (c) the blend of two original images. (d) and (e): two rectified images. (f) the blend of two rectified images. A horizontal line is added to images to check the rectification.

variance for many practical applications. In contrast to the advance in the invariance of feature detectors, the matching precision has not been paid enough attention even though the repeatability and stability are extensively studied. Matching precision is evaluated on a pair of images and reflects to some extent the *average relative localization precision* between two images. It depends on the localization precision of feature detector, the scale change between two images, the descriptor construction and matching protocol. In this chapter, we focus on the SIFT method and measures its matching precision by average residual error under different geometric transformations. For scale invariant feature detector, the matching precision decreases with scale of features. This drawback can be avoided by canceling the sub-sampling in SIFT scale space. This first schema improves the matching precision only when there is no scale change between two images. An iterative schema is thus proposed to treat the scale change. For real images, a local filtering technique is used to improve the matching precision if the transformation between two image is locally smooth.

The applications of precise SIFT matchings can be envisaged in three directions:

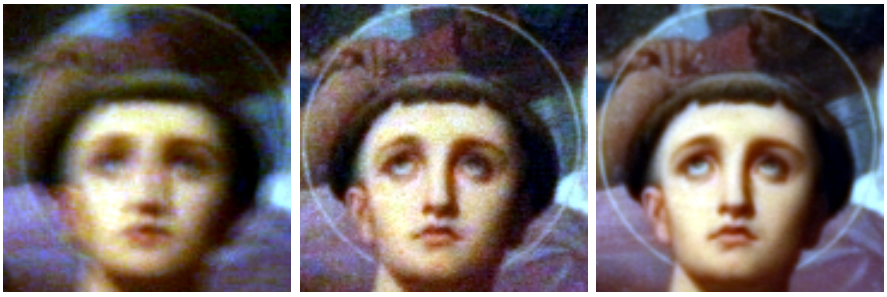
- panorama: the aim is to perform a photo montage seamlessly from several photos. These photos are taken at large distance from the object and the overlapping between two adjacent photos is important.

- super-resolution: the aim is to obtain an image in higher resolution from a burst of low resolution images of a flat object.
- global camera calibration: the aim is to calibrate camera model parameters by several photos of a flat pattern.

Chapter ??: Burst Denoising

Denoising is one of the most important image enhancement techniques. Even though denoising algorithms have been largely developed since years, most of them concentrate on the single image denoising. However, with the increase of memory size and data storage speed, today it is possible for cameras to take a burst of images in a few seconds. This opens a new possibility to do image denoising, in particular with dim light. Perhaps all of us have the frustrated experience to take photos in museum under low light conditions, where the flash of camera and tripod are forbidden. In such situation, taking photographs with a hand-held camera is problematic. If the camera is set to a long exposure time, the photograph gets motion blur. If it is taken with short exposure, the image is noisy. This dilemma can be solved by taking a burst of images, each with short-exposure time, as shown in Fig. 10. But then, as classical in video processing, an accurate registration technique is required to align the images. Denote by $u(x)$ the ideal non noisy image color at a pixel x. Such an image can be obtained from a still scene by a camera in a fixed position with a long exposure time. The observed value for a short exposure time τ is a random Poisson variable with mean $\tau u(x)$ and the standard variation proportional to $\tau u(x)$. Thus the SNR increases with the exposure time proportionally to τ. The core idea of the burst denoising method is a slight extension of the same law. The only assumption is that the various values at a cross-registered pixel obtained by a burst are i.i.d. Thus, averaging the registered images amounts to averaging several realizations of these random variables. An easy calculation shows that this increases the SNR by a factor proportional to $\sqrt{n}$ where n is the number of shots in the burst. (We call SNR of a given pixel the ratio of its temporal standard deviation to its temporal mean). Fig. 10 summarizes the possibilities offered by an image burst. A long exposure image is exposed to motion blur. The short exposure image is noisy, but sharp. Finally, the image obtained by averaging the images of the burst after registration is both sharp and noiseless. In this real example the burst taken in a gallery had 16 images. The noise should therefore be divided 4.

Even though the denoising power of burst denoising is eventually hemmed by the low growth of the square root, dividing the noise by the mentioned factors and getting an artifact free image is in no way a negligible ambition. Indeed, even the best state of the art denoising methods can create slightly annoying artifacts. If a fine non-periodic texture is present in an image, it is



(a) (b) (c)

Figure 10: (a) one long-exposure image (time = 0.4 sec, ISO=100). (b) one of 16 short-exposure images (time = 1/40 sec, ISO = 1600). (c) the average after registration. All images have been color-balanced to show the same contrast. The long exposure image is blurry due to camera motion. The middle short-exposure image is noisy, and the third one is some 5.6 times less noisy, being the result of averaging 32 short-exposure images.

virtually indistinguishable from noise, and actually contains a flat spectrum part which has the same Fourier spectrum as the white noise. Such fine textures can be distinguished from noise only if several samples of the same texture are present in other frames and can be accurately registered.

Yet, this method rises serious technical objections. The main technical objection is: how to register globally the images of a burst? Fortunately, there are several situations where the series of snapshots are indeed related to each other by a homography, and we shall explore these situations first. The homography assumption is actually valid if one of the assumptions is satisfied:

- the only motion of the camera is an arbitrary rotation around its optic center;
- the photographed objects share the same plane in the 3D scene;
- the whole scene is far away from the camera.

In those cases, image registration is equivalent to computing the underlying image homography. But this registration should be sub-pixel accurate. To this aim we will use the precise SIFT in Chapter **??** and a generalization of ORSA (Optimized Random Sampling Algorithm, [3]) to register all the images together. Yet, in general, the images of 3D scene are not related by a homography, but by an epipolar geometry. Even if the camera is well-calibrated, 3D point-to-point correspondence is impossible to obtain without knowing the depth of the 3D scene. Therefore, we should not expect that a simple homography will work everywhere in the image, but only on

a significant part. On this part, we shall say that we have a dominant homography. At each pixel that is well-registered, the registered samples are i.i.d. samples of the same underlying Poisson model. As a result, a signal dependent noise model will be accurately estimated for each colour channel. This model simply is a curve of image intensity versus the standard deviation of the noise.

Averaging does not work at the mis-registered pixels, and block matching methods are at risk on the fine image structures. Thus they will be combined. The simple combination used here will be a convex combination of them, the weight function being based on the noise curve and on the observed standard deviation of the values for the accumulation at a certain pixel. If this standard deviation is compatible with the noise model, the denoised value will be the mean of the samples. Otherwise, the standard deviation test will imply that the registration at this point is inaccurate and a conservative denoising will be applied.

7 Main Contributions

The main contributions of this thesis are listed as follows:

- a concept of virtual pinhole camera
- a tool to evaluate the precision of distortion correction with a calibration harp
- a non-parametric pattern based distortion correction method
- a concept of "self-consistency" and "universality" of distortion model
- a distortion correction method with a calibration harp
- a robust three-step image rectification
- an evaluation and improvement of SIFT matching precision
- a burst denoising algorithm

And the thesis leads to the following publications and reports:

- R. Grompone von Gioi, P. Monasse, J.M. Morel and Z. Tang. Lens distortion correction with a calibration harp. *IEEE International Conference on Image Processing*, 2011
- R. Grompone von Gioi, P. Monasse, J.M. Morel and Z. Tang. Self-consistency and universality of camera lens distortion models. Submitted, 2011

- R. Grompone von Gioi, P. Monasse, J.M. Morel and Z. Tang. Correction de distorsion optique avec une harpe de calibration. Submitted, 2011
- A. Buades, Y. Lou, J.M. Morel and Z. Tang. Multi image noise estimation and denoising. Preprint, 2010
- P. Monasse, J.M. Morel and Z. Tang. Three-step image rectification. *British Machine Vision conference*, 2010
- R. Grompone von Gioi, P. Monasse, J.M. Morel and Z. Tang. Towards High-precision Lens Distortion Correction. *IEEE International Conference on Image Processing*, 2010
- A. Buades, Y. Lou, J.M. Morel and Z. Tang. A Note on multi-image denoising. *International Workshop on Local and Non-Local Approximation in Image Processing*, 2009

References

[1] F. Devernay. A non-maxima suppression method for edge detection with sub-pixel accuracy. Technical Report 2724, INRIA rapport de recherche, 1995.

[2] J. Digne. *Inverse Geometry: From the raw point cloud to the 3D surface - Theory and Algorithms.* PhD thesis, École Normale Supérieur de Cachan, 2010.

[3] Martin A. Fischler and Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. Of the ACM*, 24:381395, 1981.

[4] Olivier Faugeras Frédéric Devernay. Straight lines have to be straight. *Mach. Vision Appli.*, 13:14–24, 2001.

[5] R. Hartley and S. B. Kang. Parameter-free radial distortion correction with center of distortion estimation. *IEEE PAMI*, 13091321, 2007.

[6] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, ISBN: 0521623049, 2000.

[7] Cordelia Schmid Krystian Mikolajczyk. An affine invariant interest point detector. *ECCV*, pages 128–142, 2002.

[8] Cordelia Schmid Krystian Mikolajczyk. Scale and affine invariant interest point detectors. *IJCV*, 60(1):63–86, 2004.

[9] Dhome M. Lavest J., Viala M. Do we really need accurate calibration pattern to achieve a reliable camera calibration. *ECCV*, 1:158–174, 1998.

[10] David G Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91110, 2OO4.

[11] J.-M. Morel G. Randall R. Grompone von Gioi, J. Jakubowicz. Lsd: A fast line segment detector with a false detection control. *IEEE Trans. on PAMI*, 99, 2008.

[12] N. Sabater. *Reliability and Accuracy in Stereovision. Application to Aerial and Satellite High Resolution Images*. PhD thesis, École Normale Supérieur de Cachan, 2009.

[13] Z. Zhang. A flexible new technique for camera calibration. *ICCV*, pages 663–673, September 1999.