Application for Faculty Position
School of Information Science and Technology, ShanghaiTech University

Zhongwei Tang

Contents

唐忠伟，现任法国国立桥路工程师学校 (巴黎高科集团) 博士后研究员。2003 年本科毕业于复旦大学电子工程系，2004 年前往法国学习，并于 2011 在获得法国高等师范卡尚分校的应用数学博士，后前往美国明尼苏达大学和杜克大学从事博士后研究，主要研究方向为几何先验信息在高精度三维重建 (3D stereo reconstruction)，多图像增强 (multiple images enhancement) 和稀疏表达 (sparse representation) 方面的应用。

近年来从事的项目先后受到过法国国家科研局 (French National Research Agency)，法国国家空间研究中心 (French National Center for Space Studies)，欧洲科研委员会 (European Research Council)，美国国家科学基金 (National Science Foundation) 等的资助。从事的具体工作主要包括高精度的照相机镜头扭曲矫正 (camera lens distortion correction)，高精度的照相机校准 (camera calibration)，多图像去噪 (burst denoting) 和人体动作分类 (unsupervised human actions classification) 等。开发的一部分软件包被交付给法国国家空间研究中心，用于下一代地球观测卫星的研发。

我很希望能够到上海科技大学信息科学与技术学院从事教学科研工作，在此附上我的申请材料，希望各位领导能在百忙之中阅读。如有任何问题，请与我联系。我的联系方式为：

手机电话: +33 (0)6 52 21 57 28
办公室电话: +33 (0)1 47 40 59 43
传真: +33 (0)1 47 40 59 01
电子邮件: tang@cmla.ens-cachan.fr, tangfrch@gmail.com

School of Information Science and Technology
ShanghaiTech University
Building 8
319, Yueyang Road
Shanghai 200031
China

To whom it may concern,

This is the application from Zhongwei Tang Ph.D., for the faculty position of the School of Information Science and Technology (SIST) at ShanghaiTech University. My education background and research experiences in image processing and computer vision will prove to be an effective match for the position.

I am currently a postdoctoral research associate with Dr. Pascal Monasse in IMAGINE at Ecole des Ponts ParisTech, working on the STEREO project funded by French National Research Agency (ANR). I also work closely with Professor Jean-Michel Morel at Ecole Normale Supérieure de Cachan. Before I was a postdoctoral research associate with Professor Guillermo Sapiro in Electrical and Computer Engineering department at Duke University.

I obtained my B.S. degree in electronic engineering from Fudan University in 2003. After one-year work in semiconductor in Shanghai, I was admitted by French National Polytenchique Institute of Toulouse (INP-T) for two-year engineer study. I then spent the next four years at Ecole Normale Supérieure de Cachan, first for the master, then the Ph.D. degree, both in applied mathematics, major in image processing and computer vision.

My researches are at the intersection of multi-view geometry and multi-image enhancement. In the other words, I want to use the geometric prior deduced from images to guide the image enhancement, and use the enhanced images to explore more robust and precise geometric structure. The two problems are complementary to each other and thus benefit each other. The use of geometric prior to stabilize one image enhancement is not new. However, in the case of multiple images, how to interact the multi-image enhancement with multi-view geometry has not been completely solved.

I have been working on the 3D stereo reconstruction, where multiple images taken from different viewpoints are used to reconstruct the 3D rigid scene. As shown in our ANR funded Callisto project, once the camera model and the lens distortion model have been accurately calibrated, the camera is converted into a high precision measurement device. Then the final 3D reconstruction precision will only depend on the precision of matching feature points between images. The software package we have delivered to French National Center for Space Studies (CNES) follows this idea to first rectify images using accurate camera calibration information and then search precise matching points between images along one direction. One part of the software has been modified and integrated into the stereo software of IMAGINE lab, which won the first award of PRoVisG Mars 3D Challenge, organized by Jet Propulsion Laboratory (JPL) of NASA. However, the recent computation shows that the matching precision can be further improved by reducing the noise level of images by accumulating more images taken in a small baseline set up. This is at the core of our ongoing ANR funded STEREO project, which aims at building a complete 3D stereo reconstruction chain, testable and reproducible on Image Processing On Line (IPOL project funded by European Research Council), with an accent on the precision.

Multiple images used in stereo reconstruction are usually taken under control. However, most of images available on the web are taken quite arbitrarily by hand in wide baseline set up and are contaminated by noise, blur and other artifacts. Our first attempt for multi-image denoising has achieved the state-of-art performance for fine image detail restoration. This approach works well in the cases where the scene is quasi-planar or camera motion is almost a rotation. To exploit more complex geometric relationship between

multiple arbitrarily taken images, I plan to adopt the sparse representation framework: each enhanced image patch will be estimated as a sparse linear combination of atoms in an over-complete dictionary. The prior information, like the structured sparsity and the joint sparsity, will be used to select appropriate atoms and thus stabilize the estimation. The selected atoms are similar patches to the underlying image patch up to small approximation error. This is somewhat close to the idea used in our unsupervised human actions classification, which is developed for National Science Foundation (NSF) funded project for early diagnosis of psychiatric disorders. More geometric invariance is planned to be added to make the classification more robust against viewpoint and illumination change.

I hold close collaboration with the team led by Professor Jean-Michel Morel at CMLA of ENS-Cachan, the team led by Dr. Pascal Monasse at IMAGINE of ENPC and the team led by Professor Guillermo Sapiro at Duke University. The team led by Professor Jean-Michel Morel has extensive experience on image denoising and analysis. Its seminal work on Non-Local Mean denoising is considered as the first one to exploit the self-similarity between patches in one image, and is the basis of many other modern denoising algorithms. The team at IMAGINE of ENPC has been working for several years on dense multi-view stereo vision, with the focus on high precision 3D surface reconstruction from images, targeting large-scale data sets taken under uncontrolled conditions. The team led by Professor Guillermo Sapiro is known for its work on sparse representation and statistical machine learning with applications in image and video restoration and classification. Some young and active researchers in China, like Professor Yaxin Peng at Shanghai University, Professor Chunli Shen and Chaomin Shen at East China Normal University and Professor Tieyong Zeng at Hong Kong Baptist University, are also interested in my projects and keep in close contact with me. Finally, technology transfer under license to CNES, Technicolor and DxO Lab is possible in the future.

I apply for the faculty position of SIST at ShanghaiTech University for three reasons. First, it is the target for ShanghaiTech to become a globally recognized top research university under the sponsorship of Shanghai Municipal government and Chinese Academy of Science. The guaranteed high quality and high internationalization level of the university definitely benefits my future career. Second, the new education system adopted by ShanghaiTech goes along with my teaching and research philosophy (please see my research plan and teaching statement in the below). With the interaction and collaboration with highly qualified faculty and students in SIST, I will be able to efficiently deploy my teaching and research plan. Third, ShanghaiTech is located at Zhangjiang high-tech park, the well-known Shanghai Silicon Valley. The geographic advantage makes the university closely connected to the industry. This allows me to interact with the industry and orient my research to real industrial needs.

In summary, I have received rigorous training from image processing and computer vision, holding both theoretic knowledge and practical experience. I am a dynamic and responsible person, highly motivated to teaching and research work. I am also considered by my colleagues as a good and reliable collaborator easy to communicate.

I appreciate that you spend time in reading this letter and the other application materials. I am looking forward to having an opportunity to interview. Please do not hesitate to contact me if you have any question.

Sincerely,

Zhongwei Tang

IMAGINE, Ecole des Ponts ParisTech

Email: tangfrch@gmail.com

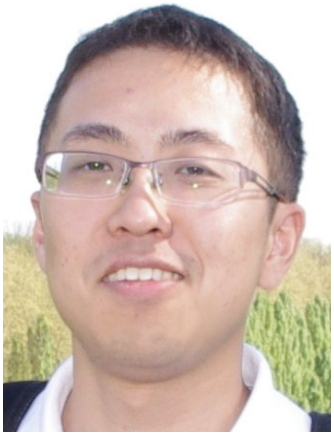
Phone: +33 (0)6 52 21 57 28

Curriculum Vitae

Zhongwei TANG

IMAGINE, Ecole des Ponts ParisTech
6-8, Avenue Blaise Pascal - Cité Descartes
Champs-sur-Marne
77455 Marne-la-Vallée, France

Born on April 11th, 1981, in Pudong, Shanghai
Mobile phone: +33 (0)6 52 21 57 28
Office phone: +33 (0)1 47 40 59 43
Fax: +33 (0)1 47 40 59 01
Email: tang@cmla.ens-cachan.fr, tangfrch@gmail.com
Linkedin: http://www.linkedin.com/pub/zhongwei-tang/53/632/66



Skills

More than 7 years' experience on image processing, multiple view geometry, computer vision, statistical machine learning and pattern recognition

- Image processing: image enhancement, multiple images merging, contrast enhancement, histogram specification, Poisson image editing, image compression
- Multiple view geometry: 3D reconstruction, camera calibration, image rectification, lens distortion correction, image stitching
- Computer vision: feature detection and matching, edge detection, image registration, image segmentation, object tracking, *A Contrario* image analysis
- Image processing chain modeling in digital camera: Gamma correction, white balance, contrast enhancement, RAW image generation, JPEG compression, point spread function, diffraction, image demosaicing, optical lens system, geometric imaging model, anti-aliasing filter, defocus aberration, chromatic aberration, vignetting
- Statistical machine learning and pattern recognition: support vector machine, linear discriminant analysis, Baysian statistics, sparse dictionary learning, clustering, probabilistic graphical model, robust statistics
- Software: co-author of stereo reconstruction software package delivered to French National Center for Space Studies (in C++), co-author of line-based lens distortion correction software (in C++), lens distortion measurement software (in C), pattern-based lens distortion correction software (in C), reflective symmetry detection software (in Matlab), automatic image rectification software (in Matlab)

Professional Experience

- **Postdoctoral Research Associate** 07/2013-present
  IMAGINE, **Ecole des Ponts ParisTech**, France
  - ◇ Work on the project STEREO funded by French National Research Agency (ANR) to study and improve the precision of 3D stereo reconstruction, which helps French National Center for Space Studies (CNES) to design the next generation of Earth observation satellite
  - ◇ Work on a generalized symmetry detection algorithm, including reflective symmetry, mirror symmetry and translation symmetry

- **Postdoctoral Research Associate** 01/2012-06/2013
  Electrical and Computer Engineering department, **Duke University/University of Minnesota**, US
  - ◇ Develop an automatic human actions classification algorithm by using machine learning techniques, with potential applications of video surveillance, abnormal events detection, video indexing, etc.
  - ◇ Develop a reflective symmetry detection algorithm based on image feature points, with potential applications of face recognition, image segmentation, etc.
- **Research Assistant** 07/2011-12/2011
  CMLA, **Ecole Normale Supérieure de Cachan, Cachan, France**
  - ◇ Develop a high-precision pattern-based camera lens distortion correction method
  - ◇ Develop a high-precision image registration algorithm based on image feature points
  - ◇ Develop a parametric high-precision line-based camera lens distortion correction method
  - ◇ Develop an automatic image rectification algorithm
  - ◇ Develop a burst image denoising algorithm based on multiple photos of different viewpoints
  - ◇ Study and implement the 3D reconstruction chain, including distortion correction, camera calibration, image rectification, dense image registration, etc.
- **Tesing Engineer** 09/2003-09/2004
  Semiconductor Manufacturing International Corporation(SMIC), Shanghai, China
  - ◇ Responsible for probe card room and new room member training
  - ◇ Repair, maintain and test probe cards

Education

- **Ph.D.**, Applied Mathematics (Image Processing and Computer Vision) 10/2007-07/2011
  Centre de Mathématiques et de Leurs Applications (CMLA)
  Ecole Normale Supérieure de Cachan (ENS-Cachan), Cachan, France
- **M.Sc.**, Mathematics, Vision, Learning (ex-DEA MVA) 10/2006-09/2007
  Ecole Normale Supérieure de Cachan, Cachan, France
- **Engineering Degree** (Diplôme d'Ingénieur), Electronic and Signal Processing 10/2004-09/2006
  Institut National Polytechnique de Toulouse (ENSEEIHT-INPT), Toulouse, France
- **B.Sc.**, Electronic Engineering 10/1999-09/2003
  Fudan University, Shanghai, China

Industry Intern Experience

- **ENST** (Ecole Nationale Supérieure des Télécommunications), Paris, France 04/2007-10/2007
  - ◇ Involved in the project DIVINE funded by French National Research Agency (ANR), which aims at developing a system to diffuse image and video flows towards heterogeneous mobile terminals through heterogeneous networks environment
  - ◇ Study on multiple descriptors coding which represents the same source information by a set of independent flows such that the decoder is able to reconstruct the source information with one or some flows lost during the transmission
- **Philips Semiconductor (now NXP)**, Caen, France 03/2006-09/2006
  - ◇ Integrated in the division of "channel decoding" in Philips Semiconductor (now known as NXP

semiconductor) which designs digital TV decoding chips

- ◇ The mobile channel is characterized by the "multi-path" and "time varing" which introduce the effect of inter symbol interference (ISI) (also known as Doppler effect) and inter carrier interference (ICI)
- ◇ Development of mobile channel estimation and equalization algorithms to reduce the above two effects

- **IRIT** (Institut de Recherche en Informatique de Toulouse), Toulouse, France 07/2005-08/2005
  - ◇ The watermark is the information hidden in images, robust again attacks, for ownership or copyright identification
  - ◇ Study and comparison of image watermark algorithms

Awards

- First Prize of "ChunHui" Venture Competition, Guangzhou, 12/2012
- First Prize of Venture Competition organized by the Embassy of the People 's Republic of China in France, 07/2012
- Graduate Summer School travel grants, Park City Mathematics Institute, U.S. 06/2010-07/2010
- International Research Scholarship, ENS-Cachan, France 10/2007-10/2010

Participated Projects

- "Stereo reconstruction at the limits of its precision (STEREO)", directed by Prof. Jean-Michel Morel, sponsored by French National Research Agency (ANR)
  - ◇ Study and improve the precision of 3D stereo reconstruction, which helps French National Center for Space Studies (CNES) to design the next generation Earth observation satellite
  - ◇ Automatize the 3D stereo reconstruction from a couple of images from different viewpoints
- "CDI-Type II: Computational Tools for Behavioral Analysis, Diagnosis, and Intervention of at Risk Children", directed by Prof. Nikolaos Papanikolopoulos and Prof. Guillermo Sapiro, sponsored by National Science Foundation (NSF)
  - ◇ Use video cameras to collect and analyze data regarding human movements
  - ◇ Assist with the early diagnosis of children at risk of developing psychiatric disorders by computer vision and image processing techniques, like tracking, segmentation, pose estimation, etc.
- "Calibration in Stereo Vision by Statistics Methods (Callisto)", directed by Dr. Pascal Monasse, sponsored by French National Research Agency (ANR)
  - ◇ Camera calibration consists in estimating the internal parameters and external positions of camera
  - ◇ Develop a two-step camera calibration approach, more precise than the state-of-art methods
- "Twelve Labours of Image Processing", directed by Prof. Jean-Michel Morel, sponsored by European Research Council (ERC)
  - ◇ "Image Processing On Line (IPOL)" (www.ipol.im): a new concept of publication to support reproducible research and software in image processing and analysis
  - ◇ Help to create and boost this new journal, and review regularly the papers submitted to this journal
- "Automatic Computation of Digital Elevation Models from Satellite Images with Small Baseline", directed by Prof. Jean-Michel Morel, sponsored by French National Center for Space Studies (CNES)
  - ◇ Help to reconstruct a reliable 3D terrain models from two images taken almost simultaneously by air borne cameras or satellite cameras

Journal Reviewer

SIAM Journal on Imaging Sciences, Journal of Mathematical Imaging and Vision, Image Processing On Line, Optics Express, Applied Optics, Journal of the Optical Society of America A, IEEE Geoscience and Remote Sensing Letters, IEEE Transaction on Image Processing

Publications

- Reflective symmetry detection by rectifying randomized correspondences, Z. Tang, M. Tepper and G. Sapiro, *British Machine Vision conference*, 2013
- "Are you imitating me?": Unsupervised sparse modeling for single video group activity analysis, Z. Tang, A. Castrodad, M. Tepper and G. Sapiro, *submitted to International Journal of Computer Vision*, 2012
- High-precision camera distortion measurements by "calibration harp", Z. Tang, R. Grompone von Gioi, P. Monasse and J.M. Morel, *Journal of the Optical Society of America A*, 2012
- High-precision camera distortion correction, Z. Tang, R. Grompone von Gioi, P. Monasse and J.M. Morel, preprint, 2012
- Self-consistency and universality of camera lens distortion models, Z. Tang, R. Grompone von Gioi, P. Monasse and J.M. Morel, *Submitted to IEEE Transaction on Image Processing*, 2012
- Multi image noise estimation and denoising*, A. Buades, Y. Lou, J.M. Morel and Z. Tang, preprint, 2012
- Lens distortion correction with a calibration harp*, R. Grompone von Gioi, P. Monasse, J.M. Morel and Z. Tang, *International Conference on Image Processing*, 2011
- Three-step image rectification*, P. Monasse, J.M. Morel and Z. Tang, *British Machine Vision conference*, 2010
- Towards High-precision Lens Distortion Correction*, R. Grompone von Gioi, P. Monasse, J.M. Morel and Z. Tang, *International Conference on Image Processing*, 2010
- A Note on multi-image denoising*, A. Buades, Y. Lou, J.M. Morel and Z. Tang, *International Workshop on Local and Non-Local Approximation in Image Processing*, 2009

* I am the leading author of the paper, although the authors are ordered in the alphabetic order. Please contact Prof. Jean-Michel Morel (morel@cmla.ens-cachan.fr), the leader of the image processing group of CMLA, ENS-Cachan, for a certificate.

Talks

- Reflective symmetry detection by rectifying randomized correspondences, poster presentation, *British Machine Vision Conference*, Bristol, U.K., September 2013
- Some geometric problems in image denoising, *Siemens Corporate Technology*, Beijing, China, June 2013
- High precision lens distortion measurement and correction, *CGG Vertitas*, Crawley, UK, January 2013
- Sparse modeling in image denoising and human actions classification, *Philips Healthcare*, Paris, France, September 2012
- The precision issue on lens distortion correction, *International Conference on Image Processing*, Brussels, Belgium, September 2011
- Lens distortion correction with a calibration harp, poster presentation, *Gretsi*, Bordeaux, France, September 2011
- Towards High-Precision Lens Distortion Correction, *International Conference on Image Processing*, Hong Kong, China, September 2010

- High-precision Lens distortion correction, *East China Normal University*, Shanghai, China, September 2010
- Three-step image rectification, poster presentation, *British Machine Vision Conference*, Aberystwyth, U.K., August 2010
- Three-step image rectification, poster presentation, *INRIA Visual Recognition and Machine Learning Summer School*, Grenoble, France, July 2010
- Non-parametric lens distortion correction, *Mathématiques de l'Imagerie Stéréoscopique Spatiale (MISS)*, Cachan, France, September 2009
- Essential matrix and three-step image rectification, *Mathématiques de l'Imagerie Stéréoscopique Spatiale (MISS)*, Cachan, France, June 2008
- Fundamental matrix estimation and its application in image rectification, *Centre National d'Etudes Spatiales (CNES)*, Toulouse, France, March 2008

Language Skills

- Chinese native language
- English fluent (living in U.S from December 2011 to June 2013)
- French fluent (living in France from July 2004 to December 2011)

Computer Skills

- Language: C/C++, Matlab, VHDL, Assembly languages Intel 8086, Intel 8051
- Operation system: Windows, Unix/Linux, MacOS
- Software: LaTeX, GIMP, Inkscape, Microsoft Office

References

- Prof. Jean-Michel MOREL
  Professor of Applied Mathematics, CMLA, Ecole Normale Supérieure de Cachan, France
  Email: morel@cmla.ens-cachan.fr Tel: +33 (0)1 47 40 29 87 Fax: +33 (0)1 47 40 59 01
- Dr. Pascal MONASSE
  Researcher of Computer Vision, IMAGINE, Ecole des Ponts, ParisTech, France
  Email: monasse@imagine.enpc.fr Tel: +33 (0)1 64 15 21 76 Fax: +33 (0)1 64 15 21 99
- Prof. Guillermo SAPIRO
  Edmund T. Pratt, Jr. Professor, Department of Electrical and Computer Engineering, Duke University, Durham, U.S.
  Email: guillermo.sapiro@duke.edu Tel: +16126251343 Fax: +16126254583

Research Statement and Plan

My research plan is a natural continuation and extension of my current and previous research projects. I have been working on camera calibration, lens distortion correction, multiple view geometry and image denoising. All the projects involve the interaction between the geometry with multiple images, which is, in my opinion, the key to increase the precision and the robustness. Depending on the problem at hand, the geometric prior can be explicitly used to extract the desired structure or more implicitly incorporated into the model to stabilize the estimation. Along this line, I will present my research plan in three parts, based on my current and previous researches.

1 Stereo vision in high precision

It is commonly admitted that the camera-based 3D stereo reconstruction methods have not yet achieved the precision of 3D laser scanner. A triangulation 3D laser scanner of good quality can produce a point cloud at the precision about 20 μm (at a distance of about 70 cm), better than the state-of-art of stereo reconstruction algorithms based on 2D photos. However, this is only possible on Lambertian objects under very controlled environment with expensive and sophisticated high quality laser scanners. In contrast, today's camera is a passive device, much cheaper and flexible, producing high resolution photos. The general goal is thus to design new algorithms to perform the 3D reconstruction from high quality 2D photos at a comparable precision as 3D laser scanner, converting the camera into a high precision measurement device, suitable for on-site 3D reconstruction tasks, particularly for the tasks like architecture modeling, heritage protection and geology disasters early warning.

1.1 Two-step camera calibration

The 3D stereo reconstruction is a complex chain (see Fig. 1) which requires a maximal precision control at each step to achieve the final high precision. Our project "Callisto" supported by French Research Agency (ANR) focuses on the first step of this chain, i.e. camera calibration, which estimates the internal parameters and external positions of the camera. The preliminary results show that high precision camera calibration is possible if it is performed in two steps separately: camera distortion correction, followed by pinhole camera calibration. This procedure deviates from most of the global camera calibration practice, which calibrates the camera with the distortion together [14]. However, the global camera calibration is a highly non-convex optimization problem and suffers from the error compensation between parameters, reducing the calibration precision [15]. Our two-step calibration avoids the error compensation by first correcting the lens distortion precisely and then linearly calibrate the resulting pinhole camera.

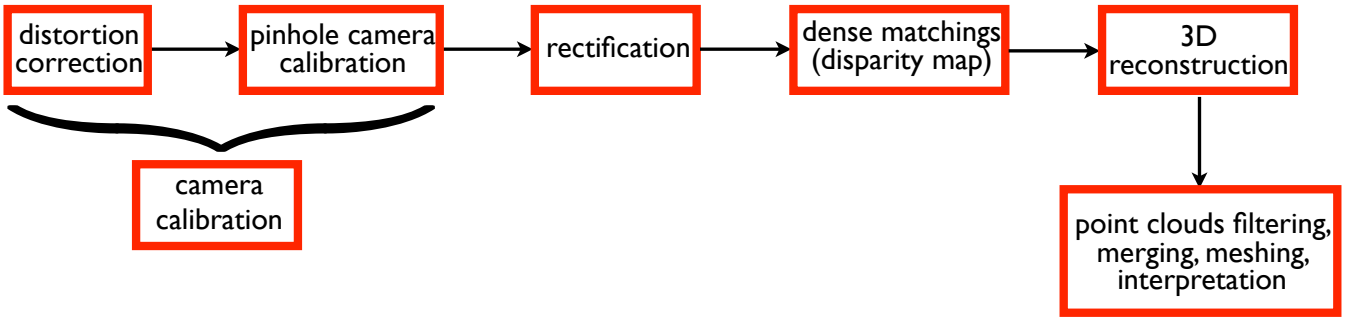


Figure 1: The 3D stereo reconstruction chain.

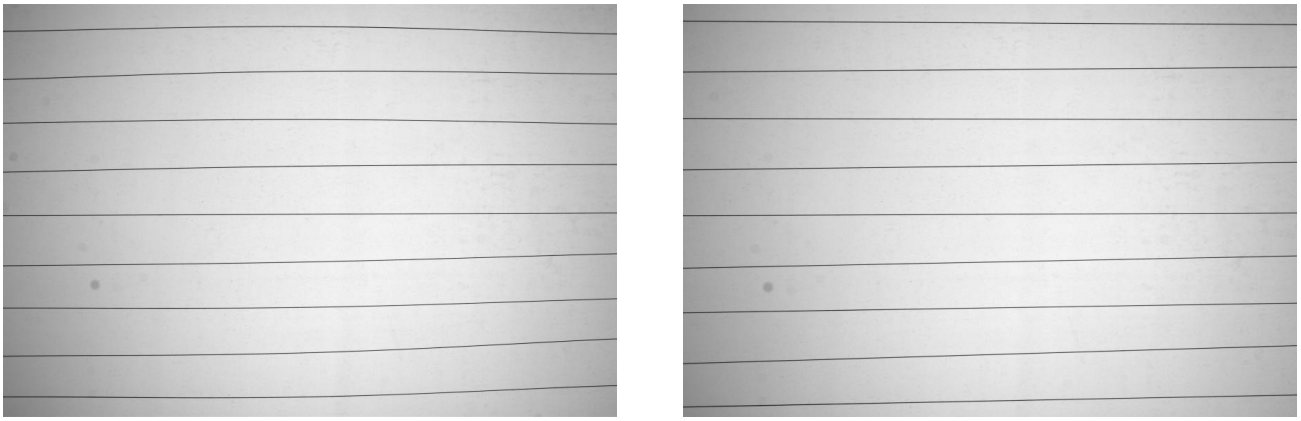
Both the distortion correction and pinhole camera calibration are performed with a pattern. The common practice is to use the same pattern (for example the pattern in Fig. 2a). But our calculations show that even a tiny non-flatness of the pattern (100 μm) can produce an imprecision about 0.3 pixels

in the distortion correction [28]. Due to the difficulty to find a pattern that flat, we built another type of pattern, called "calibration harp" (see Fig. 2b), containing perfectly straight lines, to estimate and correct the distortion at high precision about 1/30 pixels (Fig. 3) [27, 26]. Then the camera calibration is performed with the photos of the calibration pattern (Fig. 2a), corrected by the estimated distortion model.

A caveat is that the distortion depends on the camera focal length and focus, thus on the distance from the camera to the object. So ideally, two patterns should be placed at the same distance from the camera. This is somehow delicate for close range calibration because the viewpoint change in taking the photos of the calibration pattern can possibly change the distance.

(a) (b)

Figure 2: (a) the pattern for camera calibration. (b) the calibration harp for distortion correction.



(a) (b)

Figure 3: (a) distorted image taken by Canon EOS 30D at focal length 18mm. (b) corrected image.

1.2 Small "b/h": a new set up for high precision stereo vision

This problem mentioned above is largely alleviated if the camera focus is far away and the movement of camera (b) during photographing is relatively small with respect to the distance from the camera to the object (h). In this small baseline situation (typically $b/h < 0.1$), all photos of the same object look similar to each other, contain the same distortion and remain sharp at the focus. Except for these advantages, the challenges, in particular the requirement on the high precision, still exist, even become more important. Without entering into technical details, I list in the next the main ideas to achieve the high precision at each step:

- Distortion model estimation: We have shown that, in close range, with the calibration harp, the distortion can be estimated up to 1/30 pixels, even for short focal length lens presenting big distortion. This

precision is much higher than the state of art and many commercial software. In the case of small "b/h" with camera focus at distance, we will need a bigger calibration harp. A smarter way to avoid building the harp is to automatically find the straight lines in the scene and correct the distortion by rectifying them. This is in the framework of RANSAC which requires a robust distortion model estimation by eliminating many "outliers" [9].

- Distortion correction: The estimated distortion model can be used to correct any other image (taken by the same camera under the same condition). It is equivalent to say that the camera does not introduce any distortion, thus it is a pinhole camera which is a linear model in terms of projective geometry [12]. The correction is concerned with the sub-pixel interpolation in the image domain. A high-order interpolation (Fourier interpolation or high-order spline interpolation) is necessary to produce the corrected image at the same quality as the original image. Special attention needs to be paid to not introduce aliasing when the distortion model contains a sub-sampling.

- Pinhole camera calibration: The high precision distortion correction insisted above ensures that the camera becomes very close to a pinhole camera, thus stabilizes its linear calibration in the small "b/h" set up, which is otherwise a sensitive set up. Another trump we have in hand is that ellipse centers on the calibration pattern photos can be detected at about 1/1000 pixels precision, ensuring that the precision gained in distortion correction will not be offset. An analysis of the optimal precision achievable according to the residual distortion and the ellipse detection error will be needed.

- Rectification: Rectification can be considered as a generalized image alignment for two images of any 3D scene. Under pure camera rotation, or with a completely flat scene, the whole image domain can be aligned by a 2D homography. However, in general camera motion and 3D scene, the alignment is only possible in one direction [12]. For calibrated cameras, the rectification can be computed analytically, without losing any precision gained before [10, 20]. The rectification is particularly useful for computing the *disparity map*, which consists of the shifting value between each point in one image and its corresponding point in the other image. On rectified images, the correspondence search domain is reduced from whole image domain to one dimension along the alignment direction.

- Disparity map computation: Disparity map computation is another important step in the 3D reconstruction chain. With calibrated cameras, we can directly compute the 3D position of points from their disparity values. Small "b/h" is a favorite case for computing disparity map in the sense that two images taken almost at the same time with minimal occlusion and illumination variation are easier to compare by *block matching*. However, since the 3D position is proportional to the corresponding disparity value and the value of $\frac{1}{b/h}$, a small error in the disparity value will be amplified by the big factor $\frac{1}{b/h}$. This emphasizes the importance of obtaining a precise disparity map. It is shown in [24] that disparity precision largely depends on the noise level in images. So it is preferred to denoise the images before the disparity computation. In the small "b/h" situation, inspired by our burst denoising algorithm [2, 11], we plan to accumulate multiple images to reduce the noise without introducing any artifact, to finally achieve higher disparity map precision.

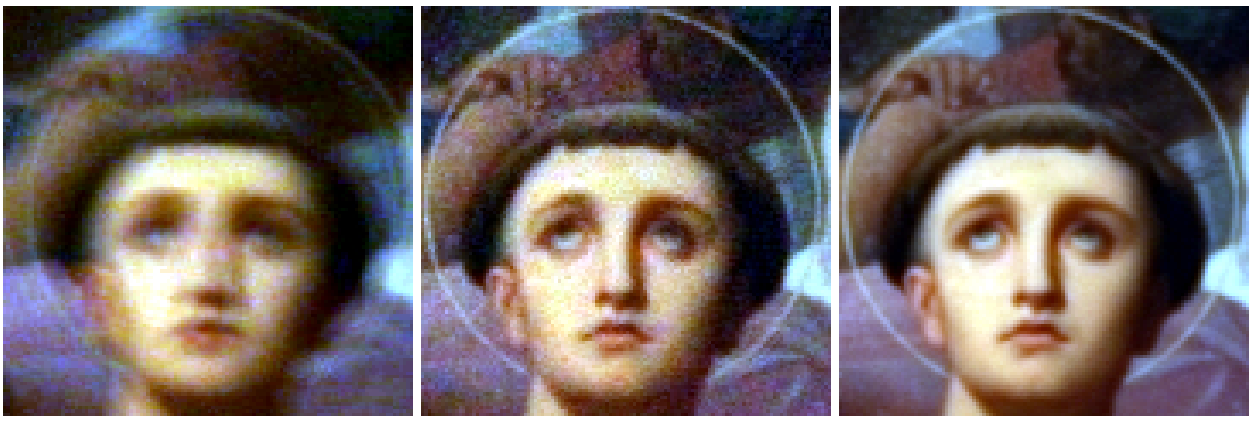
- Point clouds filtering, merging, meshing and interpretation: Following the above procedures, one pair of images can produce an accurate and dense point cloud. With more than two images available, we will have multiple separate point clouds. Point clouds filtering, merging and meshing has been largely discussed in the thesis of Julie Digne [7]. We will thus refer to her thesis to begin the experiments. However, how to interpret the geometric structure from point cloud is still an open problem. One possible way is to adopt the Gestalt theory, which is known to be the single substantial scientific attempt to state the laws of visual perception. Its first mathematic formalization has been recently proposed in [6], based on the Helmholtz principle, which states that no meaningful structure can be perceived in white noise. This formalization has been used to solve many 2D geometric structure detection problems [5, 19, 18]. We plan to extend its application to 3D point clouds.

2 Multi-image restoration

With the popularization of smart phone and shared network, today it is very easy to take photos and share with the others. A direct consequence is that we have access to a large number of photos taken before one same scene. These photos are usually taken from different viewpoints, at different times and under varying illumination. As the photos collected on Picasa or Flickr show, they are often made by hand with compact cameras without much effort to improve the quality. A natural question to ask is whether we can restore an image of good quality from a series of low-quality images. This is somehow similar to the problem in Section 1, where we claim that clean images are the key to compute an accurate disparity map and eventual lead to a high-precision 3D reconstruction. Another realistic application is a super-resolution system which produces high-resolution images for HDTV from low-resolution images for standard TV.

2.1 Burst denoising

Burst denoising [2] is our first attempt to multi-image enhancement, motivated from the frustrating experience of taking photographs in a 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. With the more recent digital cameras, this dilemma can be solved by taking a burst of images, each with short-exposure time, as shown in Fig. 4. 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 variation proportional to $\tau u(x)$. Thus the SNR (Signal-to-Noise Ratio) 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. (Independent and Identically Distributed). 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. 4 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.



(a) (b) (c)

Figure 4: (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. 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.

2.2 Sparse multi-image restoration

While burst denoising benefits from the redundant information in multiple images, it takes the risk that the registration is not enough precise so that some pixels are slightly misaligned. In fact, the ideal image registration is only possible when camera motion is a pure rotation around the optic center or the 3D scene is flat. This seldom happens in practice. So burst denoising usually applies on a dominant plane in the image and some more reserved denoising methods [1] apply on the other parts of the image. This limits the application of burst denoising.

Another approach more robust against the registration error is to estimate an image as a sparse linear combination of atoms in a dictionary. The estimation in a fixed dictionary gives already impressive single image denoising result [4], which can be again improved by using a dictionary learned from images [8, 17]. The core idea of dictionary learning is to construct a dictionary which is more adapted to the image, thus promotes a sparse representation for the image and a non-sparse representation for the noise. Since the over-completeness of dictionary requires a lot of training data, the learning is usually performed on small image patches due to their low dimension, relatively simple structure and self-similarity [22]. Multiple images are thus appropriate for dictionary learning, with a large number of image patches whose redundancy and self-similarity are naturally obtained by the overlapping between images.

In the framework of sparse modeling, image restoration consists in estimating image patches which are similar to the original image patches and have a sparse representation in the learned dictionary. Due to the over-completeness of dictionary, the degree of freedom to select such patches is huge, which makes the estimation unstable. So more prior information should be imposed to stabilize the estimation. Two concepts are appropriate to stabilize multiple image restoration. One concept is the joint sparsity and the other one is the structured sparsity. Joint sparsity explores the fact that a set of image patches share the same atoms in a dictionary. In the other words, they can be represented by the same set of atoms. This helps to stabilize the estimation. In the case of multiple image restoration, the set of images have some overlapping areas, which makes the application of the joint sparsity appropriate. Besides the joint sparsity, the concept of structured sparsity is also appropriate both for dictionary learning and sparse representation. For dictionary learning, instead of learning a "chaotic" dictionary from natural image patches, we add more constraints to make the learned atoms more structured, explicitly expressing some geometry like the patch orientation. For sparse representation, the selected atoms should not be arbitrary neither. Instead, the atoms should present similar structure as the image patches they want to represent.

3 Transformation invariant classification

Classification can be considered as the complement to image processing in the sense that image processing produces an enhanced image, while classification determines which class an image belongs to. The general approach employed in classification is to first extract some features from each class of training images and then learn a classifier from the features. A new image can be classified according to the classifier response to its features.

Both the features and the classifier are important for classification. On the one hand, we need distinctive and robust features which are invariant to geometric transformation and illumination change. On the other hand, we need to find a good classifier which correctly builds separable models for each class of images from their features. Important advances have been made in both aspects these years. For invariant features, the seminal work of SIFT first achieved the fully similarity invariance and robustness to illumination change [16]. ASIFT (Affine SIFT) [21] simulates two missing parameters of camera left by SIFT and achieves fully affine invariance, which is sufficient to explain any local planar deformation. For classifiers, besides famous SVM and its kernel variants [3, 13], sparse models also show the state-of-art performance on the action and image classification [29].

So it is natural to combine the advances of two aspects to boost the performance of classification on

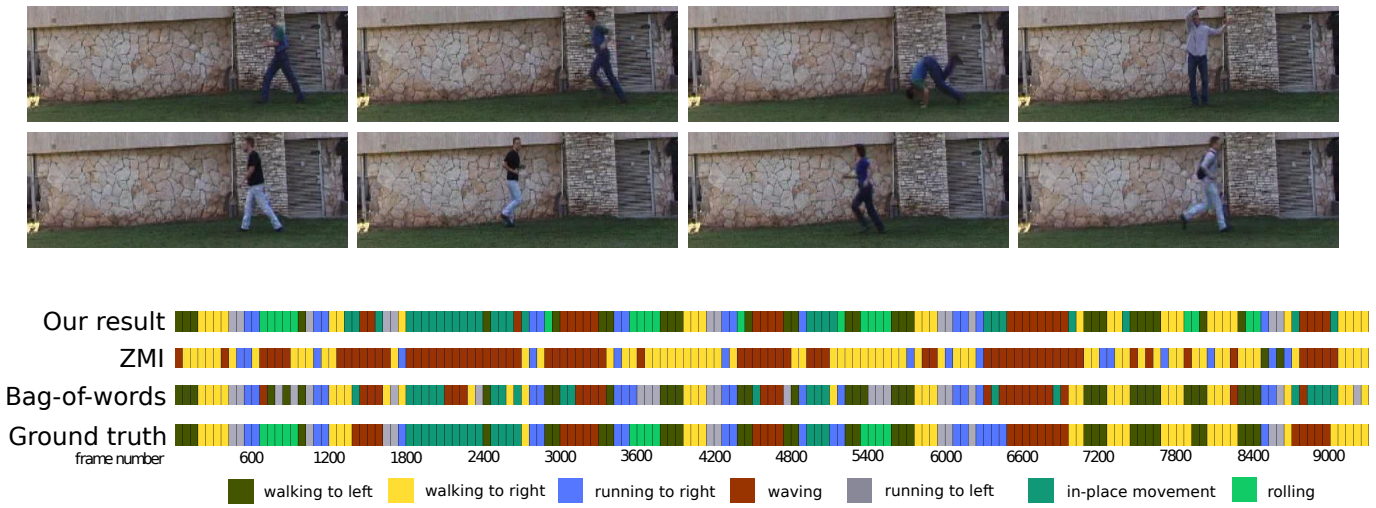


Figure 5: Analysis of the Outdoor video on 320 1-second intervals. The ground truth is illustrated by seven colors and captions. The clustering results only show the action grouping, without action identity information. The nonparametric spectral clustering algorithm detects seven groups of action for our method, four for Zelnik-Manor and Irani's method (ZMI) [30], and four for bag-of-words method. The rand index [23] for the clustering is 92.6%, 62.0%, and 88.9% for our method, ZMI, and bag-of-words method, respectively.

the invariance. One possible path is to first learn a dictionary from each class of images, then simulate the affine transformation on the atoms in the dictionary to obtain an extended dictionary. A global dictionary can be constructed by simply concatenating each sub-dictionary. The classification decision on a new image is then determined by which sub-dictionary contributes the most in its sparse representation. This similar approach has been employed by our unsupervised human actions classification and achieved state-of-art performance, even without incorporating the feature invariance (Fig. 5) [25]. We also plan to extend this work by building features describing actions in a more invariant way against video camera viewpoint and intrinsic action variation.

4 Summary

All of the above projects require the expertise of image processing, stereo vision, multi-view geometry and machine learning. The experimentation skills, including the photograph practice and the proficient coding skill are also indispensable to make the projects successful. Since the projects are closely related to the work I did with Dr. Pascal Monasse (IMAGINE, Ecole des Ponts ParisTech, France), Prof. Jean-Michel Morel (CMLA, ENS-Cachan, France) and Prof. Guillermo Sapiro (ECE, Duke University, U.S.), the regular exchanges and collaborations are under plan. The results will also interest the satellite conception team led by Dr. Bernard Rougé and Dr. Gwendoline Blanchet at French National Center for Space Studies (CNES). The transfer of technique and knowledge to CNES will help to design the next generation of Earth observation satellite.

The School of Information Science and Technology at ShanghaiTech University is very new and dynamic, showing high potential to become globally recognized top research university, in particular with its innovative advisory education system and its high quality faculty team. I am happy to see that there are many distinguished professors doing the research in satellite design, image processing and computer vision, which coincides exactly with my research direction. I hope that my research will also interest them and finally lead to collaboration to boost the research level of the university.

References

[1] A. Buades, B. Coll, and J.M. Morel. A non local algorithm for image denoising. *IEEE Conference on Computer Vision and Pattern Recognition*, 2:60–65, 2005.

[2] A. Buades, Y. Lou, J.M Morel, and Z. Tang. A note on multi-image denoising. *Local and Non-Local Approximation in Image Processing*, pages 1–15, 2009.

[3] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[4] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007.

[5] A. Desolneux, L. Moisan, and J.-M. Morel. A grouping principle and four applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(4):508–513, 2003.

[6] A. Desolneux, L. Moisan, and J.-M. Morel. *From Gestalt Theory to Image Analysis: A Probabilistic Approach*. Springer-Verlag, collection "Interdisciplinary Applied Mathematics", vol.34, 2008.

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

[8] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006.

[9] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[10] A. Fusiello, E. Trucco, and A. Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12:16–22, 2000.

[11] G. Haro, A. Buades, and J.-M. Morel. Photographing paintings by image fusion. *SIAM Journal Imaging Science*, 5(3):1055–1087, 2012.

[12] R.I Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge Univ. Press, 2000.

[13] T. Hofmann, B. Scholkopf, and A.J. Smola. Kernel methods in machine learning. *Annals of Statistics*, 36(3):1171–1220, 2008.

[14] J. Lavest, M. Viala, and M. Dhome. Do we really need accurate calibration pattern to achieve a reliable camera calibration? *European Conference on Computer Vision*, 1:158–174, 1998.

[15] H. Li and R. Hartley. A non-iterative method for correcting lens distortion from nine point correspondences. *OmniVision*, 2005.

[16] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[17] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Transactions on Image Processing*, 17(1):53–69, 2008.

[18] L. Moisan, P. Moulon, and P. Monasse. Automatic homographic registration of a pair of images, with a contrario elimination of outliers. *Image Processing On Line*, 2012.

[19] L. Moisan and B. Stival. A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. *International Journal of Computer Vision*, 57(3):201–218, 2004.

[20] P. Monasse, J.M. Morel, and Z. Tang. Three-step image rectification. *British Machine Vision Conference*, 2010.

[21] J.M. Morel and G.Yu. ASIFT: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2(2):438–469, 2009.

[22] Bruno A. Olshausen and David J. Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision Research*, 37(23):3311–3325, 1997.

[23] W. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66:846–850, 1971.

[24] N. Sabater, J.M. Morel, and A. Almansa. How accurate can block matches be in stereo vision ? *SIAM Journal of Imaging Sciences*, 4(1):472–500, 2011.

[25] Z. Tang, A. Castrodad, M. Tepper, and G. Sapiro. Unsupervised sparse modeling for grouping human actions in video sequences. *Preprint, ECE, Duke University*, 2012.

[26] Z. Tang, R. Grompone von Gioi, P. Monasse, and J.M. Morel. High-precision camera distortion measurements by "calibration harp". *Journal of the Optical Society of America A*, 29(10):2134–2143, 2011.

[27] 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*, pages 617–620, 2011.

[28] Rafael Grompone von Gioi, Jérémie Jakubowicz, Jean-Michel Morel, and Gregory Randall. LSD: A fast line segment detector with a false detection control. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(4):22–732, 2010.

[29] J Wright, Y. Ma, J Mairal, G. Sapiro, T. Huang, and S. Yan. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, 8(6):1031–1044, 2010.

[30] L. Zelnik-Manor and M. Irani. Statistical analysis of dynamic actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1530–1535, 2006.

Teaching Statement

My teaching experiences date back to the years when I did my Ph.D. study in Centre des Mathématiques et de Leurs Applications (CMLA) at Ecole Normale Supérieure de Cachan (ENS-Cachan). I was responsible for teaching two computer based courses about image processing at the graduate level, totally 30∼35 students per course and 96 hours per academic year. I also supervised some master student projects on image processing and analysis algorithms. These experiences allowed me to develop and improve my teaching skills. Due to the lack of teaching materials, I also spent much time in developing the exercises and the corresponding computer programs.

My first teaching principle is to make the students understand what I teach. Image processing is an interdisciplinary subject, closely related to mathematics and engineering. But many students in my class only had either mathematical background or engineering background. This happened to me when I taught the course "partial differential equation in image processing". Engineering students usually grabbed the main idea but had difficulty in understanding the underlying partial differential equation, while mathematic students can understand every mathematic detail but often lost the big picture. In this situation, I first encouraged the students to discuss and exchange their ideas. When they were discussing, I walked around in the class to note down what are the missing points they still do not understand. Then I began to write down the equations and also show one image on the big screen. Each time the equations evolve or the parameters change, the image changes correspondingly by computer program I wrote beforehand. This allows both the mathematic students and engineering students to understand how images are "processed". I also emphasized the missing points by using more exemplar images or even some toy examples to make sure that the students understand. During the course, the students are always encouraged to interrupt me to ask any question, which were answered through the interactive illustration if possible. The typical questions are collected for the course preparation of the next year.

My second teaching principle is to make the course interesting. The interest and curiosity are two prerequisites to learn. In the class, I usually did not ask students to do many mathematic exercises because this does not tell image processing from other mathematics courses. Instead, I asked them to implement some basic image processing algorithms. They were all very interested in making their own program work. The programming is perhaps an awkward exercise for the students who have no or little programming experience. To solve this problem, I usually make two or three students to work together as a group, at least one of them having engineering background with basic programming skills and the others probably having a more mathematical profile. In order to make the students concentrate on the algorithmic aspect, I designed very well commented programs which only lack some key steps. I went to each group to make sure that at least one student understood and could teach his teammates. Another very helpful tool I use is the Image processing On Line (IPOL), which has the ambition to make all the image processing algorithms accessible and testable on line. Students could directly have a first impression about what image processing can do by testing the algorithm on their own images. This is very helpful to do a general presentation at the beginning of the course to attract the attention of students.

I also understand that my future position will require me to lead my own research group and guide students at undergraduate and graduate level. This is not exactly the same as teaching a course. By definition, research is an iterative process of search and re-search, and there is no standard solution to problems. So it is crucial to train the students to think independently. When I guided master students projects, I never told students the answer or the idea directly. Instead students were asked to first understand what is the problem, then study the references to see how the others solve the problem and what is the state-of-art solution. All these must be written as a technical report, as the training for paper writing. The next phase, which is also more difficult, is to make the students have their own ideas. I usually asked the students to reproduce the state-of-art algorithm and do extensive tests. Then I sat down with the students and discussed the results together. In the discussion, I would inspire the students to find the weak points

of the algorithm and propose their own solutions. Once new results (either worse or better) were available, the discussion would continue until we converged to a mature solution. During the process, the students were also required to update the technical report in detail, which will be a good draft for the final paper.

I believe that teaching and research are complementary to each other. A qualified professor should be good at both aspects. This goes along with the new education system adopted by ShanghaiTech University, where professors are required to do researches and teaching as well. And each undergraduate is assigned to a supervisor from a relevant field to guide their studies and researches. So ShanghaiTech University is a perfect place for me to practice my teaching philosophy. I hope I can have more opportunities to teach in the future to further improve my teaching skills and stay close to the students as a helpful advisor.

High-precision camera distortion measurements with a "calibration harp"

Zhongwei Tang,[1,*] Rafael Grompone von Gioi,[2] Pascal Monasse,[3] and Jean-Michel Morel[2]

[1]*Electrical and Computer Engineering, University of Minnesota, Minneapolis, Minnesota 55455, USA*
[2]*Centre de Mathématiques et de Leurs Applications, Ecole Normale Supérieure de Cachan, Cachan, 94230, France*
[3]*IMAGINE, LIGM, Université Paris-Est/Ecole des Ponts ParisTech, Champs-sur-Marne, 77455, France*
**Corresponding author: tang@cmla.ens-cachan.fr*

Received April 5, 2012; revised June 26, 2012; accepted August 6, 2012;
posted August 24, 2012 (Doc. ID 165884); published September 18, 2012

This paper addresses the high-precision measurement of the distortion of a digital camera from photographs. Traditionally, this distortion is measured from photographs of a flat pattern that contains aligned elements. Nevertheless, it is nearly impossible to fabricate a very flat pattern and to validate its flatness. This fact limits the attainable measurable precisions. In contrast, it is much easier to obtain physically very precise straight lines by tightly stretching good quality strings on a frame. Taking literally "plumb-line methods," we built a "calibration harp" instead of the classic flat patterns to obtain a high-precision measurement tool, demonstrably reaching 2/100 pixel precisions. The harp is complemented with the algorithms computing automatically from harp photographs two different and complementary lens distortion measurements. The precision of the method is evaluated on images corrected by state-of-the-art distortion correction algorithms, and by popular software. Three applications are shown: first an objective and reliable measurement of the result of any distortion correction. Second, the harp permits us to control state-of-the art global camera calibration algorithms: it permits us to select the right distortion model, thus avoiding internal compensation errors inherent to these methods. Third, the method replaces manual procedures in other distortion correction methods, makes them fully automatic, and increases their reliability and precision.

OCIS codes: 100.3008, 150.1488.

1. INTRODUCTION

The precision of three-dimensional (3D) stereovision applications is intimately related to the precision of the camera calibration, and especially of the camera distortion correction. An imprecise distortion model produces residual distortion that will be directly back-projected to the reconstructed 3D scene. Such imprecision can be a serious hindrance in remote sensing applications such as the early warnings of geology disasters, or in the construction of topographic maps from stereographic pairs of aerial photographs. The fast growing resolution of digital cameras and of their optical quality is transforming them into (potential) high-precision measurement tools. Thus, it becomes important to measure the calibration precision with ever higher precision.

A first step toward high-precision distortion corrections is to perform precise distortion measurements. This basic tool can then be used to evaluate the precision of a correction method, or can become part of the correction method itself.

Camera and lens distortion measurement methods usually require a flat pattern containing aligned elements. The pattern is photographed using the target lens, and the distortion is measured by how much the observed elements deviate from the straight alignment on the pattern. For example, the classic DxO-labs' software, a good representative of camera maker practice (http://www.dxo.com/) uses a pattern with a grid of aligned dots. Distortion is measured by the positional errors associated with the maximal deviation in a row; see Fig. 1. Similar methods are proposed by the standard mobile imaging architecture (SMIA), European Broadcasting Union (EBU) (http://www.ebu.ch/), image engineering (IE) (http://www.image-engineering.de/), and International Imaging Industry Association (I3A) (http://www.i3a.org/) standards. These measurements are generally manual and require a perfectly flat pattern.

Every lens distortion correction method includes, implicitly, a lens distortion measurement. These methods can be roughly classified into four groups:

- reprojection error minimization methods,
- pattern matching methods,
- enlarged epipolar geometry-based methods, and
- plumb-line methods.

Reprojection methods usually rely on a planar pattern containing simple geometric shapes. In these methods, the lens distortion is estimated together with the camera internal and external parameters [1–5], by minimizing the reprojection error between the observed control points on the pattern and the reprojected control points simulated by the pattern model and camera model. The distortion is measured in terms of reprojection error once all the parameters have been estimated. Both the internal parameter errors and the external parameter errors contribute to the reprojection error. Unfortunately, these errors can compensate each other. Thus, a small reprojection error may be observed while the internal and external parameters are not well estimated. This compensation effect reduces the precision of the estimation of the lens distortion parameters [6]. It hinders accurate distortion estimation from the reprojection error. Notice that this high precision is not

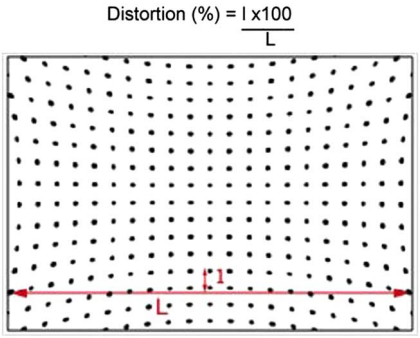


Fig. 1. (Color online) DxO lens distortion measurement standard. Estimation of distortion from an image of a dot chart.

always required. In some applications, the distortion estimated in these bundle adjustment-based reprojection methods, which makes the 3D geometry consistent with the pin-hole model, is precise enough for a human observer.

A quite different kind of method proceeds by matching a photograph of a flat pattern to its digital model. These methods estimate the distortion field by interpolating a continuous distortion field from a set of matching points. Several variants exist depending on the kind of pattern, matching, and interpolation technique. A common and crucial assumption for these methods is that the pattern is flat. In practice, however, it is difficult to produce a very flat pattern, and the consequences of a tiny flatness flaw are considerable. For example, it is reported in [7] that a flatness error of about 100 μm for a 40 cm broad pattern can lead to an error of about 0.3 pixels in the distortion field computation for a Canon EOS 30D camera of focal length 18 mm with the distance between the camera and the object about 30 cm. The only physical method to assess a pattern flatness at a high precision is interferometry, but it requires the pattern to be a mirror, which is not adequate for photography. Furthermore, camera calibration requires large patterns, which are therefore flexible. Deformations of the order of 100 μm or more can be caused by temperature changes, and by a mere position change of the pattern, which deforms under its own weight.

Recently, more attention has been paid to pattern-free methods (or self-calibration methods) where the distortion estimation is obtained without using any specific pattern. The distortion is estimated from the correspondences between two or several images in the absence of any camera information. The main tool is the so-called enlarged epipolar constraint, which incorporates lens distortion into the epipolar geometry. Some iterative [8,9] or noniterative methods [10–20] are used to estimate the distortion and correct it by minimizing an algebraic error. The estimated distortion can be used as the initialization in bundle adjustment methods to improve the camera calibration precision [21].

The so called "plumb-line" methods, which correct the distortion by rectifying distorted images of 3D straight lines, date back to the 1970s (see Brown's seminal paper in 1971 [22]). Since then, this idea has been applied to many distortion models: the radial model [23–25], the field of view (FOV) model [26], or the rational function model [27]. These methods minimize the straightness error of the corrected lines. According to the fundamental theorem to be introduced in Section 2, the plumb-line methods minimize an error directly related to the distortion, without suffering from the above-mentioned drawback, namely a numerical error compensation. On the other hand, in spite of their name, current digital plumb-line methods usually involve flat patterns with alignments on them and not the plumb lines that were originally proposed in analog photogrammetry.

Taken literally, these methods should use photographs of 3D straight lines. When a high precision is required, this setup becomes much easier to build than a flat pattern. The main purpose of this article is to show that very accurate distortion evaluation and correction can be obtained with a basic plumb-line tool, which we called "calibration harp." The calibration harp is nothing but a frame supporting tightly stretched strings. Nevertheless, the photographs of a calibration harp require a new numerical treatment to exploit them. The strings will have to be detected at high subpixel accuracy and their distortion converted into an invariant measurement.

Distortion measurements can be used to evaluate the distortion of a camera, but also its residual distortion after correction. Two aspects of the measurement should be clarified here. In this paper we discuss distortion measurements that apply to the camera conceived as a whole: It is impossible to tell which part the relative position and deformation of the CCD, and the lens distortion itself, play in the global camera distortion. The distortion measurement is therefore not a pure optical lens distortion measurement, but the distortion measurement of the full acquisition system of camera + lens in a given state. Different lenses on different cameras can be compared only when the camera calibration matrix is known. On the other hand, the residual error due to different correction algorithms can be compared objectively after applying an appropriate normalization on the corrected images.

This paper is organized as follows: The fundamental theorem characterizing undistorted cameras is introduced in Section 2. Section 3 uncovers the simple fabrication secrets of calibration harps. The image processing algorithms needed for an automatic measurement are presented in Section 4, and Section 5 introduces the two most relevant measures. Section 6 demonstrates two applications, to the measurement of residual distortion after applying a calibration method, and to the transformation of existing manual distortion correction methods into automatic and far more precise ones. Finally, Section 7 concludes the paper.

2. FROM STRAIGHT LINES TO STRAIGHT LINES

In multiple-view geometry, the pinhole camera is the ideal model that all techniques tend to approximate at best by calibrating the real cameras. This model corresponds to the ideal geometric perspective projection. The next theorem characterizes perspective projections by the fact that they preserve alignments. Instead of restating the simplified version in [28], we prefer to make it more precise. The proof of the theorem can be found in [6].

Theorem 1. *Let* **T** *be a continuous map from* $\mathcal{P}^3$ *to* $\mathcal{P}^2$ *[from 3D projective space to two-dimensional (2D) projective plane]. If there is a point* **C** *such that*

(a) *the images of any three points belonging to a line in* $\mathcal{P}^3$ *not containing* **C** *are aligned points in* $\mathcal{P}^2$,

(b) *the images of any two points belonging to a line in* $\mathcal{P}^3$ *containing* **C** *are the same point in* $\mathcal{P}^2$,

(c) *there are at least four points belonging to a plane not containing* **C**, *such that any group of three of them are nonaligned, and their images are also nomaligned, then* **T** *is a pinhole camera with center* **C**.

This theorem provides us with a fundamental tool to verify that a camera follows the pinhole model. Nevertheless, rectifying straight lines does not define a unique distortion correction: two corrections can differ by any 2D homography that preserves all alignments. More concretely, assume that the real camera model is $\mathbf{P} = \mathcal{D}$ with **C** the coordinate of camera optic center in a given 3D world frame, **R** the camera 3×3 orientation matrix, **K** the camera 3×3 calibration matrix, and $\mathcal{D}$ the camera lens nonlinear distortion. The estimated distortion can be written as $\mathcal{D}\mathbf{H}^{-1}$ with $\mathbf{H}^{-1}$ the unknown homography introduced in the distortion correction, and can be different from one correction algorithm to another. By applying the inverse of the estimated distortion, the recovered pinhole camera is $\hat{\mathbf{P}} = \mathbf{H}$. The homography **H** can enlarge or reduce the straightness error, which makes the comparison of different correction algorithms unfair. This effect can be compensated by two strategies.

To arrive at a universal measurement, a first strategy that we will consider is to normalize the homography:

1. Select four points $\mathbf{P}_{i=1,\ldots,4}$ in the distorted image in general position (not three of them aligned). For example, they can be the four corners of the distorted image.
2. Find their corresponding points $\mathbf{P}'_i$ in the corrected image, according to the correction model: $\mathbf{P}'_i = \mathbf{H}\mathcal{D}^{-1}\mathbf{P}_i$. Note that **H** is different from one correction algorithm to another.
3. Compute the normalization homography $\hat{\mathbf{H}}$ that maps $\mathbf{P}'_i$ to $\mathbf{P}_i$: $\mathbf{P}_i = \hat{\mathbf{H}}\mathbf{P}'_i$. Note that $\hat{\mathbf{H}}$ is different from one correction algorithm to another.
4. Apply the normalization homography $\hat{\mathbf{H}}$ on the corrected image.

With this normalization, the final correction model is $\hat{\mathbf{H}}\mathbf{H}\mathcal{D}^{-1}$.

A second possible strategy would be to fix specific parameters in the correction model. For example, since the zoom factor in the distortion correction is mainly determined by the order-1 parameters in the correction model, it is sufficient to set all the order-1 parameters to be 1 to obtain a unique distortion measurement. Unfortunately, this will not be possible for some nonparametric distortion correction methods. The first strategy therefore is more general.

3. BUILDING A CALIBRATION HARP

Theorem 1 suggests to take a set of physical straight lines, as a calibration pattern. However, a common practice actually contradicts the plumb-line basic idea: line patterns are printed and pasted over a flat plate. There are many sources of imprecision in this setup: the printer quality is not perfect, the paper thickness is not perfectly uniform, the pasting process can add bubbles or a nonuniform glue layer, and the supporting surface is not perfectly flat either. Notwithstanding these defects, if only a pixel precision is required, this setup is quite sufficient. Nonetheless, when high subpixel precision is involved, the flatness errors cannot be neglected. For current camera precision, a flatness error of 100 μm (the thickness of current writing paper) for a 40 cm pattern can lead to errors in the observed image coordinates of about 0.3 pixels [7]. High precision aims at final 3D reconstructions far more precise than this. So the measuring tool error should be also far smaller, if only possible.

The obvious advantage of "real" plumb lines on 2D patterns with straight lines on them is that it is much easier to ensure a very precise physical straightness for lines than a very precise physical flatness for a physical pattern plate. Yet, the precision of the resulting measurement or correction depends on the straightness of the physical lines. In [22], to achieve a high precision, plumb lines were made of very fine white thread and stabilized by immersion of the plumb bobs in oil containers. Illumination was provided by a pair of vertically mounted fluorescent fixtures. A dead black background was provided for the plumb-line array to highlight the contrast. The points on the lines were measured with a microscope. The measuring process required from 5 to 6 h for generating 324 points. The digital procedure proposed here will be automatic and faster.

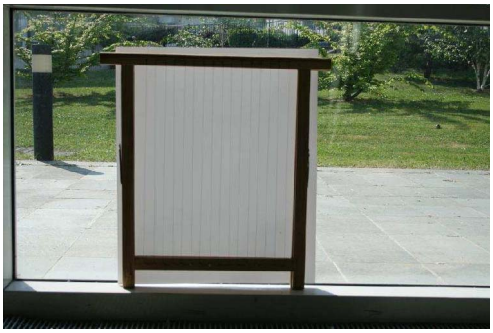
For the applications where the precision is not a crucial point, straight lines present in the natural scene can be used. In [27,29], the straight lines are obtained by photographing the architectural scenes and points on the lines are detected by Canny edge detector.

In order to keep the high precision and simplify the fabrication procedure, we built a simple calibration pattern by tightly stretching strings on a frame as shown Fig. 2. The pattern looks like the musical instrument, hence its name. The setup warrants the physical straightness of the lines. Its construction does not require any experimental skill, but only good-quality strings. Indeed twisted strings show local width oscillations; other strings do not have a round section, and a little torsion also results in width variations that can have a large spatial period. Rigid strings may have a remanent curvature. Finally, a (tiny) gravity effect can be avoided by using well-stretched vertical lines.

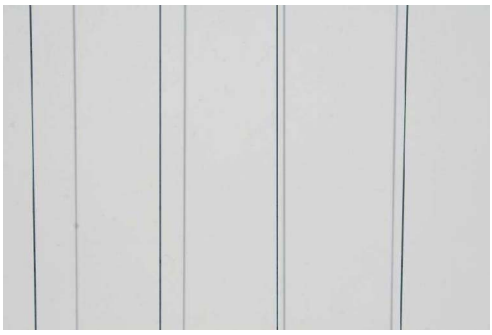
In our experiments, three different strings were tested: a sewing string, a smooth tennis racket string, and an opaque fishing string, all shown in Fig. 3. Sewing strings have a braid pattern and their thickness oscillates. Tennis racket strings are rigid and require a very strong tension to become straight. Fishing strings are both smooth and flexible, and can therefore be easily tightened and become very straight. The transparent ones, however, behave like a cylindrical lens, producing multiple complex edges. Opaque fishing strings end up being the best choice to build a calibration harp. Figure 5 shows an evaluation of the obtained straightness. We took photos of the three types of strings, and corrected their distortion. The green curves show the signed distance from edge points of a corrected line to its regression line. The red curve shows the high-frequency component of the corresponding distorted line. The high frequency is computed as follows:

- Transform the edge points extracted from the distorted line into the intrinsic coordinate system, which is determined by the direction of the regression line computed from these points. In this coordinate system, the x-coordinate is the distance between the edge points and a reference edge point along the regression line and the y-coordinate is the signed distance from the edge points to the regression line. This produces a one-dimensional signal (see Fig. 4).
- Apply a big Gaussian convolution of standard deviation $\sigma = 40$ pixels on this one-dimensional signal to keep only the low frequency component.

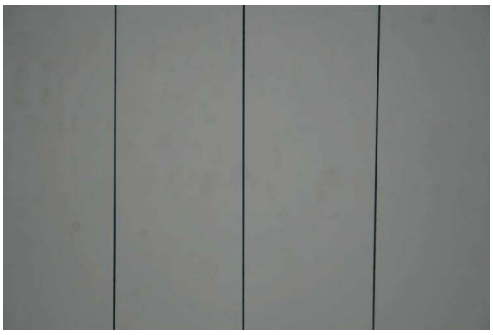
(a) The harp with a uniform opaque object as background



(b) The harp with a translucent paper as background



(c) A close-up of the harp with a uniform opaque object as background



(d) A close-up of the harp with a translucent paper as background

Fig. 2. (Color online) "Calibration harp." Shadows can be observed in (a) and (c), while there is no shadow in (b) or (d).

- The difference between this convolved signal and the original signal in the intrinsic coordinate system is considered as the high frequency. The red curves in Fig. 5 show the high frequency (due to the border effect of Gaussian convolution, there is a sharp increase of magnitude at the border).

To ensure the precision of the edge detection in the string images, a uniform background whose color contrasts well with the string color must be preferred. Using an opaque background is not a good idea because this requires a direct lighting and the strings project shadows on the background [Figs. 2(a) and 2(c)]. The sky itself is hardly usable: a large open space is needed to avoid buildings and trees entering the camera field of view, and clouds render it inhomogeneous; see Fig. 6. The simplest solution is to place a translucent homogeneous paper or plastic sheet behind the harp and to use back illumination, preferably natural light to make it more uniform [see Figs. 2(b) and 2(d)].

The acquisition aspects are also important for producing high-quality measurements: lens blur, motion blur, aliasing, and noise, must be as reduced as physically possible. To that aim, a tripod and timer were used to reduce camera motions, but also to avoid out-of-focus strings while taking photos at different orientations. Of course, changing focus changes the distortion. Thus each distortion calibration must be done for a fixed focus, and is associated with this focus.

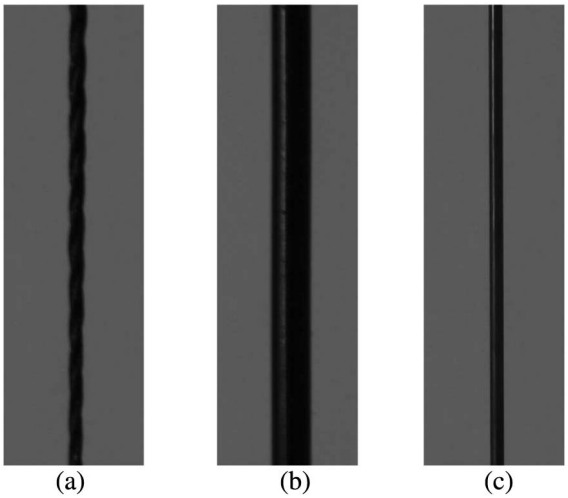


Fig. 3. Three types of strings. (a) sewing line, (b) tennis racket line, (c) opaque fishing line.

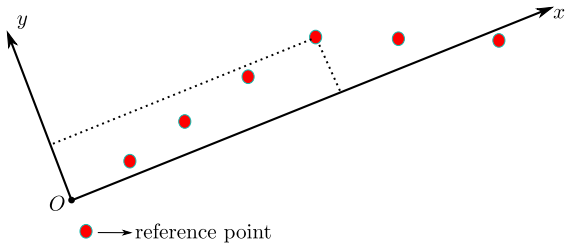
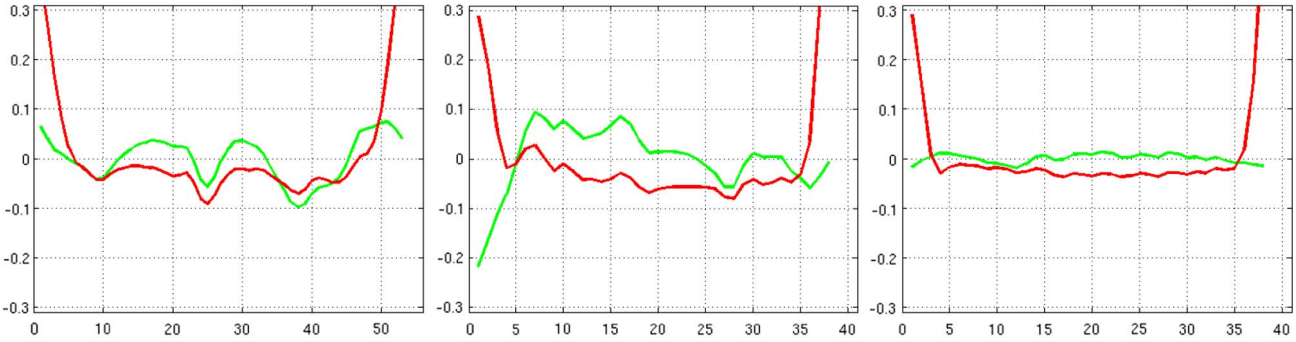


Fig. 4. (Color online) Intrinsic coordinate system of the edge points extracted on the distorted line. The red points are the distorted edge points. The x direction is determined by the direction of the regression line. The x-coordinate is the distance to the reference point along the regression line, and the y-coordinate is the signed distance from the edge points to the regression line.



(a) The sewing string (b) The tennis racket string (c) The opaque fishing string

Fig. 5. (Color online) The small oscillation of the corrected lines is related to the quality of the strings. The green (upper) curves show the signed distance (in pixels) from the edge points of a corrected line to its regression line. The red (lower) curves show the high frequency of the corresponding distorted line. The corrected line inherits the oscillation from the corresponding distorted line. (a) Sewing string, (b) tennis racket string, (c) opaque fishing line. The x-axis is the index of edge points. The range of the y-axis is from –0.3 pixels to 0.3 pixels. The almost superimposing high-frequency oscillation means that the high frequency of the distorted strings is not changed by the distortion correction. In such a case, the straightness error includes the high frequency of the distorted strings and does not really reflect the correction performance. So it is better to use a string that contains the smallest high-frequency oscillation. Among the three types of strings, the opaque fishing string shows the smallest such oscillations. The larger oscillation of the sewing string is due to a variation of the thickness related to its twisted structure, while the tennis racket string is simply too rigid to be stretched, even though this is not apparent in Fig. 3(b).

4. STRAIGHT EDGES EXTRACTION

In this section, we describe the procedure to extract accurately and smooth the aligned edge points, which will be used to measure the distortion.

Devernay's algorithm [30] is the classic subpixel accurate edge detector. The implementation of Devernay's detector is very simple since it is derived from the well-known nonmaxima suppression method [31,32]. On good quality images (SNR larger than 100), Devernay's detector can attain a precision of about 0.05 pixels.

Straightness measurements require the detection of groups of edge points that belong to the same physical straight line, and the rejection of points that do not belong to any line. To this aim, line segments are detected on the image using the line segment detector (LSD) algorithm [33,34]. When applied to photographs of the calibration harp, the detection essentially corresponds to the strings. In case of a strong distortion, one string edge could be cut into several line segments.

LSD works by grouping connected pixels into line support regions; see Fig. 7—for more details we refer to [33,34]. These regions are then approximated by a rectangle and validated. The line support region links a line segment to its support pixels. Thus, Devernay's edge points that belong to the same line support region can be grouped as aligned; points belonging to none are ignored.

For photos of strings, almost every pixel along each side of one string is detected as an edge point at subpixel precision. So there are about 1000 edge points detected for a line of length of about 1000 pixels. This large number of edge points opens the possibility to further reduce the detection and aliasing noise left by subsampling the edge points.

The subsampling step must be done carefully. First the edge points are resampled to warrant a uniform sampling step along the edge; this will facilitate the following steps. The resampling uses a step of $d = L/N$ where L is the length of a line and N is the number of extracted edge points on the line. The interpolation of an edge point (x', y') between two adjacent edge points (x_1, y_1) and (x_2, y_2) is expressed by

$$x' = \frac{b}{a+b}x_1 + \frac{a}{a+b}x_2, y' = \frac{b}{a+b}y_1 + \frac{a}{a+b}y_2,$$

where a and b are the distances between the points; see Fig. 8. Then, a Gaussian blur with $\sigma = 0.8 \times \sqrt{t^2 - 1}$ is needed before a subsampling of factor t to avoid aliasing [35]. We have two one-dimensional signals (x-coordinate and y-coordinate of

(a) Photo of the harp taken against the sky (b) Photo of the harp taken against a translucent paper using a tripod

Fig. 6. (Color online) The quality of photos depends on the harp, its background, and the stability of camera for taking photos.

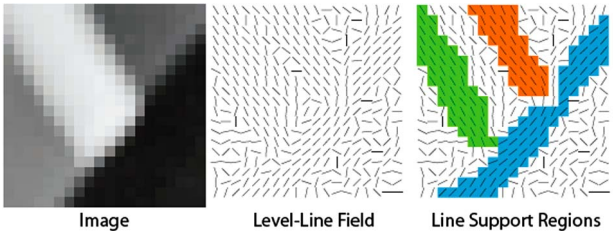


Fig. 7. (Color online) The LSD algorithm computes the level-line field of the image. The level-line field defines at each pixel the direction of the level line passing by this pixel. The image is then partitioned into connected groups that share roughly the same level-line direction. They are called line support regions. Only the validated regions are detected as line segments. Devernay's edge points belonging to the same validated line support region are considered as the edge points of the corresponding line segment.

edge points) along the length of the line. The Gaussian convolution is performed on both one-dimensional signals, parameterized by the length along the edge. Finally, the subsampling of integer factor t keeps one edge point out of t.

5. DISTORTION MEASUREMENTS

This section examines two natural distortion measurements that are somewhat complementary.

A. Root-Mean-Square Distance

According to Theorem 1, the most direct measure should be the straightness error, defined as the root-mean-square (RMS) distance from a set of distorted edge points that correspond to the same physical line, to their global linear regression line; see Fig. 9.

Given N edge points $(x_1, y_1), \ldots, (x_N, y_N)$ of a distorted line, the regression line

$$\alpha x + \beta y - \gamma = 0 \tag{1}$$

is computed by

$$\alpha = \sin\theta, \qquad \beta = \cos\theta, \qquad \gamma = A_x \sin\theta + A_y \cos\theta,$$

where

$$A_x = \frac{1}{N}\sum_{i=1}^{N} x_i, \qquad A_y = \frac{1}{N}\sum_{i=1}^{N} y_i,$$

$$V_{xx} = \frac{1}{N}\sum_{i=1}^{N} (x_i - A_x)^2, \qquad V_{xy} = \frac{1}{N}\sum_{i=1}^{N} (x_i - A_x)(y_i - A_y),$$

$$V_{yy} = \frac{1}{N}\sum_{i=1}^{N} (y_i - A_y)^2, \qquad \tan 2\theta = -\frac{2V_{xy}}{V_{xx} - V_{yy}}.$$

Since (α, β) is a unit vector, the signed distance from point (x_i, y_i) to the line is given by

$$S_i = \alpha x_i + \beta y_i - \gamma.$$

Given L lines, with N_l points in line l, the total sum of squared signed distance is given by

$$S = \sum_{l=1}^{L}\sum_{i=1}^{N_l} |S_{li}|^2 = \sum_{l=1}^{L}\sum_{i=1}^{N_l} (\alpha_l x_{li} + \beta_l y_{li} - \gamma_l)^2. \tag{2}$$

Thus, the RMS straightness error is defined as

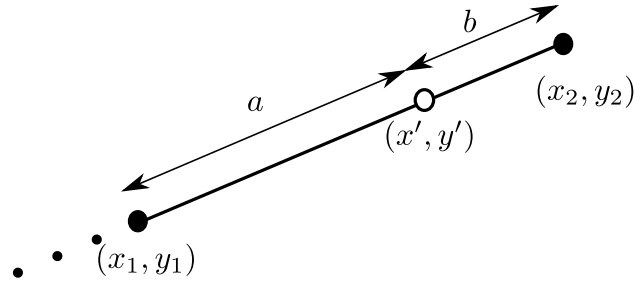


Fig. 8. Line resampling. The black dots $(x_1, y_1), (x_2, y_2), \ldots$ are the edge points extracted by Devernay's detector. They are irregularly sampled along the line. The resampling (in white dots) is made along the line with the uniform length step d. Linear interpolation is used to compute the resampled points.

$$d = \sqrt{\frac{S}{N_T}}, \tag{3}$$

where $N_T = \Sigma_{l=1}^{L} N_l$ is the total number of points.

B. Maximal Error

An alternative measure is the average maximal error defined by

$$d_{\max} = \sqrt{\frac{\sum_{l=1}^{L} |\max_i S_{li} - \min_i S_{li}|^2}{L}}. \tag{4}$$

In the classic camera maker practice, the maximal error is defined by

$$\max_l |\max_i S_{li} - \min_i S_{li}|,$$

which would become instable with the calibration harp, some of the strings being potentially distorted by blur or wrong detection.

This measure is traditionally used in manual settings; for example, see Fig. 1. While traditionally the measures are made relative to the line joining the extremities of the distorted edge (see Fig. 10), here we use the signed distance to the regression line to make it more comparable to the previous measure. The use of a signed distance and the difference between the maximal and minimal values is needed to handle correctly the fact that there are values on both sides of the regression line; see Fig. 10.

6. APPLICATIONS

In this section, real photographs of the calibration harp will be used to evaluate the residual camera distortion when this distortion has been corrected with three state-of-the-art correction methods or two popular software. In addition, we can feed any plumb-line method with the precise edge points detected on the harp images to improve the correction precision. The efficient Alvarez *et al.* algorithm [23] thus becomes an

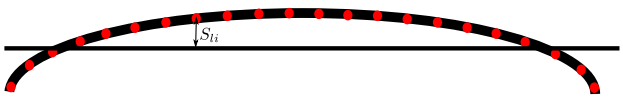


Fig. 9. (Color online) Distance from a set of points to their global linear regression line.

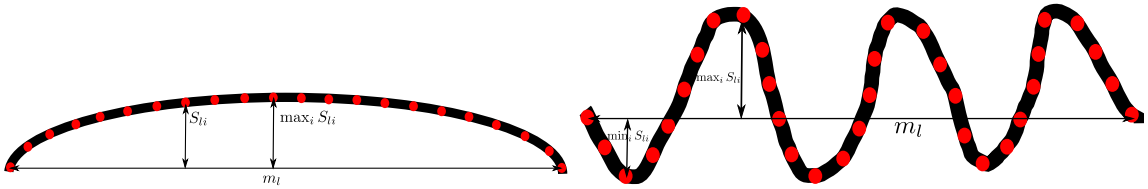


Fig. 10. (Color online) Left: traditional distortion measure: the maximal distance to the line defined by the extremities of the edge. Right: the regression line crosses the distorted line; the difference between the maximal and minimal signed distance to the line measures the full width of the distorted line.

automatic parametric distortion correction method. Our lens distortion measurement algorithm can be tested on the online demo version available at http://bit.ly/lens-distortion.

A. Precision

It is reported in [30] that Devernay edge points have a precision better than 0.05 pixels under the zero-noise condition. As proposed in Section 4, the precision of Devernay edge points can be further improved by applying Gaussian convolution of standard deviation $0.8 \times \sqrt{t^2 - 1}$, followed by a subsampling of factor t. The only parameter to adjust here is the factor t, which corresponds to the assumed regularity of the lens distortion. We are only interested in realistic lens distortion, which makes a straight line globally convex or concave. Thus local edge oscillations due to noise can be harmlessly removed. In the experiments, the value of $t = 30$ was chosen, which is enough to remove the local oscillation while keeping the global distortion.

B. Measuring the Residual Error after Distortion

As a first main application, the "calibration harp" permits us to evaluate the performance of any distortion correction algorithm by measuring its residual distortion in corrected images. The procedure is as follows:

1. A series of photos of the "calibration harp" are taken at different orientations.
2. These photos are processed by a camera distortion correction algorithm.
3. The corrected images are normalized by a homography as described in Section 2.
4. The residual distortion is measured by the proposed method.

Three distortion correction algorithms and two software were tested. With the exception of the classic Lavest *et al.* [4] calibration method, all the others are designed to only correct the lens distortion without estimating the other camera parameters:

- The Lavest *et al.* method [4]: probably the most advanced pattern-based global camera calibration method, which estimates and corrects for the pattern nonflatness, using a bundle adjustment technique. Various distortion parameter configurations are allowed in this method: two radial parameters and two tangential parameters for a partial distortion model, two radial parameters for a partial radial distortion model, five radial parameters for a complete radial distortion model, and five radial parameters and two tangential parameters for a full distortion model.
- A nonparametric lens distortion correction method requiring a textured flat pattern [7]. The pattern is obtained by printing a textured image and pasting it on an aluminum plate, which is thick and solid.
- The DxO-Optics-Pro software: a program for professional photographers automatically correcting lens distortion (even from fisheyes), color fringing and vignetting, noise, and blur. This software reads the EXIF of each image to know exactly which camera, lens, and settings have been used. It therefore uses a fixed lens distortion estimation for each supported camera model.
- PTLens: Photoshop plug-in that corrects lens pincushion/barrel distortion, vignetting, and chromatic aberration.

The distorted photographs to be corrected are shown in Fig. 11, and Table 1 shows the residual distortion measurements obtained by the calibration harp, after applying the corrections specified by the various methods.

The Lavest *et al.* method depends on the parameter configuration of the distortion model integrated in the global calibration process. Since the global calibration process only minimizes the reprojection error and does not control the distortion correction, it can happen that the error in internal parameters compensates the error in external parameters. In consequence, the minimized reprojection error is small, but neither the estimated distortion parameters nor the other parameters are correct. In fact this is the common drawback of global camera calibration methods based on bundle

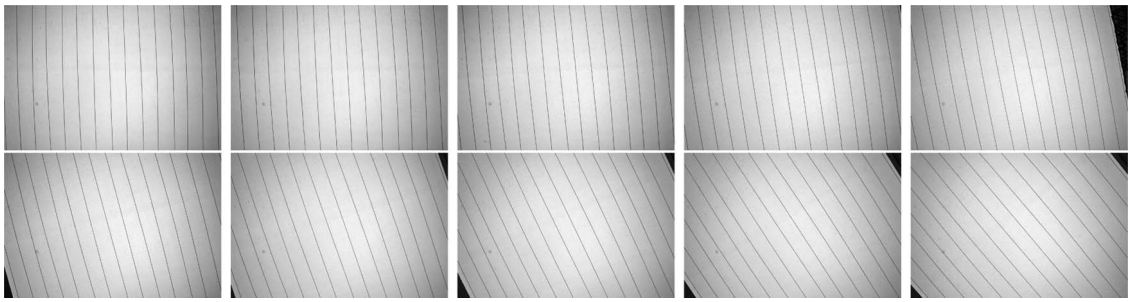


Fig. 11. Distorted photos of the "calibration harp."

Table 1. Distortion Correction Performance of Three Algorithms, Measured by RMS Distance d and Maximal Distance $d_{\max}$

Method	d (pixels)	$d_{\max}$
Original distortion	2.21	6.70
Lavest (two radial and two tangential parameters)	0.07	0.30
Lavest (two radial parameters)	0.07	0.29
Lavest (full distortion parameters)	0.60	3.00
Lavest (full radial distortion parameters)	0.59	2.90
Textured pattern	0.04	0.16
DxO Optics Pro	0.32	0.99
PTLens	0.46	1.51

adjustment. The textured pattern-based method requires a perfectly flat pattern. Even though it is not very feasible to fabricate a perfectly flat pattern, a pattern made of a thick and solid aluminium plate gives a good flatness condition and thus a precise distortion correction. DxO Optics Pro includes many precalibrated distortion models depending on the camera type and parameters setting. But these distortion models are only calibrated on several fixed focused distances and obtain by interpolation the distortion models focused on the other distances. Once the camera parameters are extracted from the EXIF of each image, DxO Optics Pro asks the user to manually input the focused distance before performing the correction. This makes the distortion correction result less precise; considering this, the results are surprisingly good. PTLens works in the similar manner as DxO Optics Pro except that it does not ask users to provide the focused distance information. It is not clear how PTLens recovers this information, which is not available in EXIF. Probably PTLens applies a fixed correction for each focal length, independently of the focus. This coarse approximation may explain why its correction precision is not as good as DxO Optics Pro's.

We also note that $d_{\max}$ is always larger than d. This is not surprising, since $d_{\max}$ is the largest displacement with respect to the linear regression line of the edge points.

C. Strengthening Plumb-Line Distortion Correction Methods

Any plumb-line distortion correction method requires as input the edge points on distorted lines, which are themselves projections of 3D straight lines. The distortion can then be corrected by aligning the edge points belonging to the same line. To see this, let us introduce the widely used radial distortion correction model:

$$x_u - x_c = f(r_d)(x_d - x_c), \tag{5}$$

$$y_u - y_c = f(r_d)(y_d - y_c), \tag{6}$$

with (x_u, y_u) the corrected point, (x_d, y_d) the distorted point, (x_c, y_c) the distortion center, and $r_d = \sqrt{(x_d - x_c)^2 + (y_d - y_c)^2}$ the distorted radius. Usually $f(r_d)$ is a polynomial of r_d and can be written as

$$f(r_d) = k_0 + k_1 r + k_2 r^2 + \cdots + k_N r^N \tag{7}$$

with $k_0, k_1, k_2, \ldots, k_N$ the distortion parameters. Assume we have L lines and there are N_l points on line l, $l = 1, 2, \ldots, L$. A natural way to correct the distortion is to minimize the sum of squared distances from the corrected points to the corresponding regression line:

$$D = \frac{1}{L}\sum_{l=1}^{L}\frac{1}{N_l}\sum_{i=1}^{N_l} S_{li}^2 = \frac{1}{L}\sum_{l=1}^{L}\frac{1}{N_l}\sum_{i=1}^{N_l}\frac{(\alpha_l x_{u_{li}} + \beta_l y_{u_{li}} + \gamma_l)^2}{\alpha_l^2 + \beta_l^2} \tag{8}$$

with the linear regression line $l{:}\alpha_l x + \beta_l y + \gamma_l = 0$ computed from the corrected points $(x_{u_{li}}, y_{u_{li}})$, $i = 1, \ldots, N_l$. The only unknown parameters in D are $k_0, k_1, \ldots, k_N$ and (x_c, y_c). With an appropriate initialization of these parameters, D can be efficiently minimized by nonlinear optimization algorithms, such as the Levenberg–Marquardt algorithm.

Instead of minimizing D, Alvarez *et al.* [23] proposed to minimize the measurement:

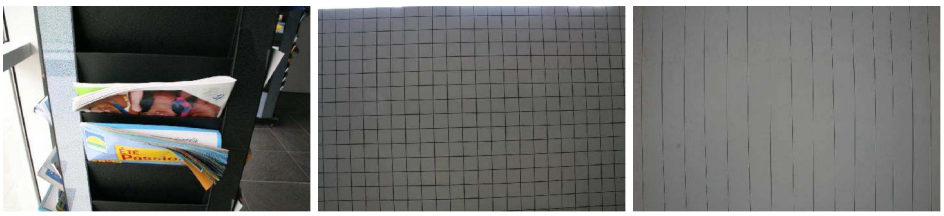
$$E = \frac{1}{L}\sum_{l=1}^{L}(S_{xx}^l S_{yy}^l - (S_{xy}^l)^2), \tag{9}$$

where S^l is the covariance matrix for the corrected points on the line l:

$$S^l = \begin{pmatrix} S_{xx}^l & S_{xy}^l \\ S_{xy}^l & S_{yy}^l \end{pmatrix} = \frac{1}{N_l}\begin{pmatrix} \sum_{i=1}^{N_l}(x_{u_{l,i}} - \bar{x}_{u_{l,i}})^2 & \sum_{i=1}^{N_l}(x_{u_{l,i}} - \bar{x}_{u_{l,i}})(y_{u_{l,i}} - \bar{y}_{u_{l,i}}) \\ \sum_{i=1}^{N_l}(x_{u_{l,i}} - \bar{x}_{u_{l,i}})(y_{u_{l,i}} - \bar{y}_{u_{l,i}}) & \sum_{i=1}^{N_l}(y_{u_{l,i}} - \bar{y}_{u_{l,i}})^2 \end{pmatrix}, \tag{10}$$

with $\bar{x}_{u_{l,i}} = \frac{1}{N_l}\sum_{i=1}^{N_l} x_{u_{l,i}}$ and $\bar{y}_{u_{l,i}} = \frac{1}{N_l}\sum_{i=1}^{N_l} y_{u_{l,i}}$. It can be proven [23] that this new energy function E is always positive and equals to 0 if and only if for each line its points are aligned. The functional E can be further written in the form of matrix-vector multiplication [23]:

$$E(\mathbf{k}) = \frac{1}{L}\sum_{l=1}^{L}\mathbf{k}^T A^l \mathbf{k}\mathbf{k}^T B^l \mathbf{k} - \mathbf{k}^T C^l \mathbf{k}\mathbf{k}^T C^l \mathbf{k}, \tag{11}$$



(a) Natural image (b) Grid pattern image (c) Harp image

Fig. 12. (Color online) Images used in the Alvarez *et al.* method [23].

with $\mathbf{k} = (k_0, k_1, k_2, \ldots, k_N)^T$ the distortion parameters in the form of vector and

$$A^l_{m,n} = \frac{1}{N_l}\sum_{i=1}^{N_l}((r_{d_{l,i}})^m x_{d_{l,i}} - (r_{d_{l,i}})^{\bar{m}} x_{d_{l,i}})((r_{d_{l,i}})^n x_{d_{l,i}} - (r_{d_{l,i}})^{\bar{n}} x_{d_{l,i}}), \quad (12)$$

$$B^l_{m,n} = \frac{1}{N_l}\sum_{i=1}^{N_l}((r_{d_{l,i}})^m y_{d_{l,i}} - (r_{d_{l,i}})^{\bar{m}} y_{d_{l,i}})((r_{d_{l,i}})^n y_{d_{l,i}} - (r_{d_{l,i}})^{\bar{n}} y_{d_{l,i}}), \quad (13)$$

$$C^l_{m,n} = \frac{1}{N_l}\sum_{i=1}^{N_l}((r_{d_{l,i}})^m x_{d_{l,i}} - (r_{d_{l,i}})^{\bar{m}} x_{d_{l,i}})((r_{d_{l,i}})^n y_{d_{l,i}} - (r_{d_{l,i}})^{\bar{n}} y_{d_{l,i}}), \quad (14)$$

with $(r_{d_{l,i}})^{\bar{m}} x_{d_{l,i}} = \frac{1}{N_l}\sum_{i=1}^{N_l}(r_{d_{l,i}})^m x_{d_{l,i}}$ and $(r_{d_{l,i}})^{\bar{m}} y_{d_{l,i}} = \frac{1}{N_l}\sum_{i=1}^{N_l}(r_{d_{l,i}})^m y_{d_{l,i}}$, $m = 1, \ldots, N$ and $n = 1, \ldots, N$. The trivial solution $\mathbf{k} = (0, 0, 0, \ldots, 0)^T$ can be avoided by setting $k_0 = 1$.

In general, minimizing $E(\mathbf{k})$ is equivalent to solve a set of equations:

$$\frac{\partial E(\mathbf{k})}{\partial k_i} = 0, \qquad i = 1, 2, \ldots, N. \quad (15)$$

When there is only one unknown parameter, the solution can be approximated by solving the root of the resulting univariate polynomial. When there are two unknown parameters, resultant-based method can be used to minimize $E(\mathbf{k})$. The case of more than two variables requires Gröbner basis techniques or the multivariate-resultants based method (see [23] for more details). To make the algorithm efficient, [23] optimizes on two parameters at one time and iterates:

1. Obtain distorted edge points that are the 2D projection of 3D straight segments.
2. Initialize $\mathbf{k} = (1, 0, \ldots, 0)^T$.
3. Choose any pair (p, q), $1 \le p, q \le N$ and fix all the other parameters, then optimize k_p and k_q by resultant-based method.
4. Update k_0 using a zoom factor such that distorted and undistorted points are as close as possible.
5. Repeat Steps 3 and 4 until all the parameters are estimated.

It is usually supposed that the edge points are already available for the plumb-line methods. But in practice, it is not a trivial problem to precisely extract aligned edge points in images. For example, the online demo [36] of the Alvarez *et al.* method [23] requires the user to click manually edge points. This is on the one hand a tedious and time-consuming work, and on the other hand, it may reduce the precision of edge points. Our method thus gives the possibility to automatize plumb-line methods. We fed four kinds of edge points to the Alvarez *et al.* method: first the manually clicked edge points of a natural image [Fig. 12(a)], second the manually clicked edge points of an image of the grid pattern [Fig. 12(b)], third the manually clicked edge points of an image of the calibration harp [Fig. 12(c)], and finally the automatically extracted edge points of an image of the calibration harp [Fig. 12(c)], as described in Section 4. These points were used as the input to the Alvarez *et al.* method to estimate an order-4 radial distortion model, which will be used to correct the distorted images in Fig. 11. The precision of this correction was finally evaluated by the method proposed in the paper (applied to a different set of images of the calibration harp from the one used to estimate the correction).

Table 2. Distortion Correction Performance of the Alvarez *et al.* Method [23] on Four Kinds of Input Edge Points: Manual Clicks on Natural Image, Manual Clicks on a Grid Pattern Image, Manual Clicks on a Calibration Harp Image, and Automatic Edge Points Extraction on the Calibration Harp Image[a]

Method	Time (mins)	d (pixels)	$\bar{d}_{\max}$
Natural image (manually)	~5	0.27	1.02
Grid pattern (manually)	~25	0.30	0.94
Calibration harp (manually)	~30	0.11	0.39
Calibration harp (automatically)	~0	0.08	0.27

[a]Compare d, $d_{\max}$ and the time to obtain the edge points.

The results in Table 2 show that the edge point extraction by our proposed method strengthens the plumb-line method in terms of precision and spares the long, tedious, and imprecise manual point-clicking task. Compared to the manual clicks with the calibration harp, the improvement in precision is moderate. Indeed, the Alvarez *et al.* method is applied on a very good quality photograph of the harp. The manual clicks were carefully placed on the lines across the domain of the image. The slight inaccuracy of the clicks was smoothed out by our method, which applies a Gaussian convolution of the edge points along the edges; see Section 4. The manual clicks on the image of the grid pattern and on the natural image give a precision that is two or three times lower than the calibration harp. For the grid pattern, the imprecision may come from the nonflatness error, the engraved straight lines on the pattern being not really straight. For the natural image, the imprecision comes from two aspects: one is again the nonstraightness error of the lines, and the other is the lack of lines at the border of the image domain, which can explain a precision decay near the image border.

7. CONCLUSION

A "calibration harp" has been proposed for camera distortion measurement, along with its associated image processing chain. This harp is both a measurement tool and a correction tool. As a measurement tool, it can be used independently to measure the residual distortion left by any distortion correction methods or any software. As a correction tool, the precise edge points detected on the harp can be used as the input to plumb-line methods and lead to a more precise distortion correction result. In the future, we aim at finding a more general distortion model to correct more severe distortion by using the calibration harp. The ideal case would be a harp-free distortion correction method. But we think the harp will always remain useful, as a final measurement tool to validate any other correction precision and/or to detect its failures.

ACKNOWLEDGMENTS

This work was supported by Agence Nationale de la Recherche ANR-09-CORD-003 (Callisto project), the Mathématiques de

l'Imagerie Satellitaire Spatiale (MISS) project of Centre National d'Etudes Spatiales, the Office of Naval Research under grant N00014-97-1-0839, and the European Research Council, advanced grant "Twelve labours."

REFERENCES

1. C. Slama, *Manual of Photogrammetry*, 4th ed. (American Society of Photogrammetry, 1980).
2. R. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," IEEE J. Robot. Autom. **3**, 323–344 (1987).
3. Z. Zhang, "A flexible new technique for camera calibration," IEEE Trans. Pattern Anal. Mach. Intell. **22**, 1330–1334 (2000).
4. J. Lavest, M. Viala, and M. Dhome, "Do we really need accurate calibration pattern to achieve a reliable camera calibration?" in *Computer Vision—ECCV'98*, Vol. 1408 of Lecture Notes in Computer Science (Springer, 1998), pp. 158–174.
5. M. H. J. Weng and P. Cohen, "Camera calibration with distortion models and accuracy evaluation," IEEE Trans. Pattern Anal. Mach. Intell. **14**, 965–980 (1992).
6. Z. Tang, "Calibration de caméra à haute précision," Ph.D. dissertation (Ecole Normale Supérieure de Cachan, 2011).
7. R. Grompone von Gioi, P. Monasse, J.-M. Morel, and Z. Tang, "Towards high-precision lens distortion correction," in *Proceedings of 17th IEEE International Conference on Image Processing* (IEEE, 2010), pp. 4237–4240.
8. G. P. Stein, "Lens distortion calibration using point correspondences," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 1997), pp. 602–608.
9. Z. Zhang, "On the epipolar geometry between two images with lens distortion," in *Proceedings of 13th International Conference on Pattern Recognition* (IEEE, 1996), pp. 407–411.
10. A. Fitzgibbon, "Simultaneous linear estimation of multiple view geometry and lens distortion," in *Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE, 2001), pp. 125–132.
11. B. Micusik and T. Pajdla, "Estimation of omnidirectional camera model from epipolar geometry," in *Proceedings of 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE, 2003), pp. 485–490.
12. H. Li and R. Hartley, "A non-iterative method for correcting lens distortion from nine-point correspondences," in *Proceedings OmniVision '05, ICCV Workshop* (2005).
13. S. Thirthala and M. Pollefeys, "The radial trifocal tensor: a tool for calibrating the radial distortion of wide-angle cameras," in *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE, 2005), pp. 321–328.
14. D. Claus and A. Fitzgibbon, "A rational function lens distortion model for general cameras," in *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE, 2005), pp. 213–219.
15. J. Barreto and K. Daniilidis, "Fundamental matrix for cameras with radial distortion," in *Proceedings of Tenth IEEE International Conference on Computer Vision* (IEEE, 2005), pp. 625–632.
16. Z. Kukelova and T. Pajdla, "Two minimal problems for cameras with radial distortion," in *Proceedings of 11th IEEE International Conference on Computer Vision* (IEEE, 2007), pp. 1–8.
17. Z. Kukelova, M. Bujnak, and T. Pajdla, "Automatic generator of minimal problem solvers," in *Computer Vision—ECCV 2008*, Vol. 5304 of Lecture Notes in Computer Science (Springer, 2008), pp. 302–315.
18. M. Byrod, Z. Kukelova, K. Josephson, T. Pajdla, and K. Astrom, "Fast and robust numerical solutions to minimal problems for cameras with radial distortion," in *Computer Vision—ECCV 2008*, Vol. 5304 of Lecture Notes in Computer Science (Springer, 2008), pp. 1–8.
19. Z. Kukelova and T. Pajdla, "A minimal solution to the autocalibration of radial distortion," in *IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2007), pp. 1–7.
20. K. Josephson and M. Byrod, "Pose estimation with radial distortion and unknown focal length," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2009), pp. 2419–2426.
21. B. Triggs, P. Mclauchlan, R. Hartley, and A. Fitzgibbon, "Bundle adjustment—a modern synthesis," *Vision Algorithms: Theory and Practice*, Vol. 1883 of Lecture Notes in Computer Science (Springer, 2000), pp. 298–372.
22. D. Brown, "Close-range camera calibration," Photogramm. Eng. **37**, 855–866 (1971).
23. L. Alvarez, L. Gomez, and J. Rafael Sendra, "An algebraic approach to lens distortion by line rectification," J. Math. Imaging Vision **35**, 36–50 (2009).
24. B. Prescott and G. McLean, "Line-based correction of radial lens distortion," Graph. Mod. Image Process. **59**, 39–47 (1997).
25. T. Pajdla, T. Werner, and V. Hlavac, "Correcting radial lens distortion without knowledge of 3-D structure," Research Report (Czech Technical University, 1997).
26. F. Devernay and O. Faugeras, "Straight lines have to be straight," Mach. Vision Appl. **13**, 14–24 (2001).
27. D. Claus and A. Fitzgibbon, "A plumbline constraint for the rational function lens distortion model," in *Proceedings of British Machine Vision Conference* (2005), pp. 99–108.
28. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision* (Cambridge University, 2004).
29. E. Rosten and R. Loveland, "Camera distortion self-calibration using the plumb-line constraint and minimal hough entropy," Mach. Vision Appl. **22**, 77–85 (2011).
30. F. Devernay, "A non-maxima suppression method for edge detection with sub-pixel accuracy," Tech. Rep. 2724, (INRIA rapport de recherche, 1995).
31. J. Canny, "A computational approach to edge detection," IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-8**, 679–698 (1986).
32. R. Deriche, "Using Canny's criteria to derive a recursively implemented optimal edge detector," Int. J. Comput. Vis. **1**, 167–187 (1987).
33. R. Grompone von Gioi, J. Jakubowicz, J. Morel, and G. Randall, "LSD: a fast line segment detector with a false detection control," IEEE Trans. Pattern Anal. Mach. Intell. **32**, 722–732 (2010).
34. R. Grompone von Gioi, J. Jakubowicz, J. Morel, and G. Randall, "LSD: a line segment detector," IPOP (2012), doi: http://dx.doi.org/10.5201/ipol.2012.gjmr-lsd.
35. J. Morel and G. Yu, "Is SIFT scale invariant?" Inverse Problems Imaging **5**, 115–136 (2011).
36. L. G. Luis Alvarez and J. R. Sendra, "Algebraic lens distortion model estimation," IPOP (2010), http://dx.doi.org/10.5201/ipol.2010.ags-alde.

Reflective Symmetry Detection by Rectifying Randomized Correspondences

Zhongwei Tang
tang@cmla.ens-cachan.fr
Mariano Tepper
mariano.tepper@duke.edu
Guillermo Sapiro
guillermo.sapiro@duke.edu

Department of Electrical and Computer Engineering
Duke University
130 Hudson Hall
Box 90291
Durham, NC 27708, USA

Abstract

We present a method for detecting bilateral or reflective symmetries in images. We pose the problem as an instance of a multiple model estimation problem. We build candidate symmetry models by randomly sampling minimal sets of SIFT matches. Since these symmetry models can be in non-frontal viewpoints, we rectify them, undoing the perspective effect. From the models with valid symmetric properties, we compute consensus sets by determining which SIFT matches are compatible with each symmetry model. We finally recombine these consensus sets, using a clustering algorithm. The method is able to detect single and multiple symmetries both in frontal and non fronto-parallel viewpoints, achieving state-of-the-art results.

1 Introduction

Symmetry is omnipresent in natural and synthetic images. Human visual perception of the world largely depends on cues provided by symmetry and repetitive patterns [3]. Automatic symmetry detection has long been an active research topic in computer vision because it also helps to enhance the performance of many vision problems, e.g., image segmentation [18], object recognition [12], person identification [6], vehicle tracking [9], and low-rank texture detection [23]. There are four types of symmetries in the 2D Euclidean space: translation, rotation, reflection and glide-reflection [21]. The human visual system is good at detecting all these symmetries, even under severely slanted viewpoints, while it remains a tough problem for computer vision. Among these, the most common form is the reflective (also called bilateral) symmetry, characterized by a line (the symmetry axis) which defines two reciprocally mirrored semi-planes. In this paper, we concentrate on detecting this type of symmetry. The concepts we propose can be nonetheless extended to the other types of symmetry.

A lot of research efforts have been dedicated to automatically detect a single, global, and dominant bilateral symmetry in images [1, 8, 13, 17]. These global methods consider the image as a whole and try to infer the global symmetry that produces the best overall fit. This poses problems when the background of the image is cluttered or the image contains multiple symmetries at different locations and scales. Although global symmetry is an important image property, encountering multiple local symmetries is far more common in practice.

Current research is thus oriented to methods based on local features [11, 20]. Finding multiple symmetries can be posed as a multiple model selection problem. Multi-RANSAC [24] and randomized Hough transform [22] are two basic model estimation tools that can be used to detect multiple symmetries. For example, a simple voting strategy was employed for bilateral symmetry detection through a fold-then-cut plane generation and synthesis [10]. Loy and Eklundh presented a more efficient method based on the Hough transform [11], where each SIFT match votes either for a bilateral symmetry axis or a center of rotational symmetry in the parameter space. Later, the work was extended to make use of the recent advance in invariant features [14] to deal with bilateral symmetry under perspective [4, 5]. Even though recent work based on a region growing scheme seems to work better [2], Loy and Eklundh's work [11] is still considered as the baseline state-of-the-art algorithm.

The Hough transform can naturally cope with multiple symmetries. However, it is not as robust against outliers as RANSAC [7] and is sensitive to the selection and discretization of the parameter space. The J-linkage algorithm was recently proposed to detect multiple primitive geometric structures in noisy and outliers-corrupted data [19]. It combines ideas from RANSAC (robustness against outliers) and the Hough transform (multiple structures detection through voting schemes). The input of the algorithm is a set X of geometric objects (e.g., 2D points). J-linkage randomly samples minimal sets of objects from X and creates candidate models (e.g., if the model is a line, two points are sufficient to define it). It then records, for each candidate model, its consensus set, i.e., the subset of X compatible with that model. This is encoded as a binary preference matrix, whose rows indicate the sampled models an object belongs to, and whose columns indicate the consensus set for each sampled model. The objects are finally clustered using the rows as vector features via agglomerative hierarchical clustering [19]. The clusters that are populated enough, can then be used to robustly estimate the final models.

We propose a method to detect multiple bilateral symmetries at different scales and viewpoints. It is based on the J-linkage framework, presenting specific and novel tools in each step of the algorithm to adapt it to our problem. These tools, described in Section 2, include: (1) a rectification step that allows us to transform each symmetry to the 2D Euclidean space, in which validation can be easily performed, while coping with skewed symmetries; (2) a simple validation criterion to determine valid rectified symmetries and their consensus set; (3) a non-uniform adaptive sampling strategy, specifically designed to deal with a large set of matches corrupted by a high percentage of outliers. In Section 3 we present experimental results showing state-of-the-art results. Finally, we provide some final remarks in Section 4.

2 Multiple symmetry detection with J-linkage

We start by detecting affine-invariant keypoints and obtaining affine-invariant SIFT descriptors with the ASIFT algorithm [15]. We denote by $\text{SIFT}(\mathbf{p}_j)$ the descriptor at keypoint $\mathbf{p}_j$. SIFT also provides the orientation θ_j and scale $s(\mathbf{p}_j)$ at $\mathbf{p}_j$. The elements in each SIFT descriptor vector $\text{SIFT}(\mathbf{p}_j)$ are rearranged to obtain their mirrored version, where the line with orientation θ_j is used as the mirroring axis [11]. We then match the SIFT descriptors with their mirrored versions using a simple nearest-neighborhood search. Let $\mathcal{M} = \{m_i\}_{i=1,\cdots,N}$ be this set of N matches. Each match m_i consists of two keypoints, expressed in homogeneous coordinates, $m_i = \{\mathbf{p}_i, \mathbf{p}'_i\}$, with $\mathbf{p}_i = (x_i, y_i, 1)^T$ and $\mathbf{p}'_i = (x'_i, y'_i, 1)^T$. We denote by $\overline{m}_i$ the line segment connecting $\mathbf{p}_i$ and $\mathbf{p}'_i$. We also denote by $\mathbf{c}_i$ the midpoint of $\overline{m}_i$.

The $\mathbb{R}^{3\times3}$ matrix $\mathbf{H}$ such that $h_{33} = 1$ (where h_{ij} is the value at the i-th row and j-th

column) is a 2D homography. This type of transforms models perspective transformations. Notice that a transformed point $\mathbf{q} = (x_q, y_q, z_q)^T = \mathbf{Hp}$ is normalized such that $z_q = 1$, although the normalization to achieve this is implicit in this paper.

J-linkage-based symmetry detection. Given N matches, we estimate K symmetries by randomly sampling K valid minimal seed sets from $\mathcal{M}$. We will see that two matches are sufficient to define a minimal seed set. The exact nature of this process will be covered in the following subsections. We thus obtain K symmetries, each with an associated consensus set (the subset of $\mathcal{M}$ compatible with each symmetry). A binary $N \times K$ matrix is thus built, where the entry (i, j) is 1 if the i-th match is in the consensus set of the j-th symmetry, and 0 otherwise. Each row of this matrix indicates which symmetries are preferred by each match and is considered as a binary feature vector for that match. Using these features, agglomerative hierarchical clustering based on the Jaccard distance [19] is used to cluster the matches. Finally, each, large enough, cluster corresponds to a local symmetry.

2.1 Estimating a fronto-parallel symmetry by rectifying match pairs

If the symmetric region only undergoes a similarity transform (rotation, translation and zoom), one match m_i is enough to determine the symmetry axis [11], which is orthogonal to the segment $\overline{\mathbf{p}_i\mathbf{p}'_i}$ and passing through $\mathbf{c}_i$. However, in practice, the symmetry is not necessarily observed in a frontal view and can thus undergo some perspective distortion. In this more general case, two matches $m_1 = \{\mathbf{p}_1, \mathbf{p}'_1\}$ and $m_2 = \{\mathbf{p}_2, \mathbf{p}'_2\}$ intersecting at the vanishing point $\mathbf{v} = (v_x, v_y, 1)^T$ are necessary to determine the symmetric axis (see Fig. 1(a)). We will undo this perspective distortion, that is, rectify the symmetry, thus being able to determine the compatibility of the remaining matches in the simpler 2D Euclidean space.

A homography $\mathbf{H}$ can be decomposed into

$$\mathbf{H} = \mathbf{ARH}_0 = \mathbf{AR}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ h_{31} & h_{32} & 1 \end{bmatrix}, \tag{1}$$

where $\mathbf{H}_0$ and $\mathbf{A}$ are a projective transform and a shear, respectively, and $\mathbf{R}$ is a rotation. We will compute $\mathbf{H}$ and use it to bring the symmetry into a fronto-parallel setting. This process is depicted in Fig. 1(a).

We begin by computing $\mathbf{H}_0$. This transform needs to send the vanishing point $\mathbf{v}$ to infinity, that is,

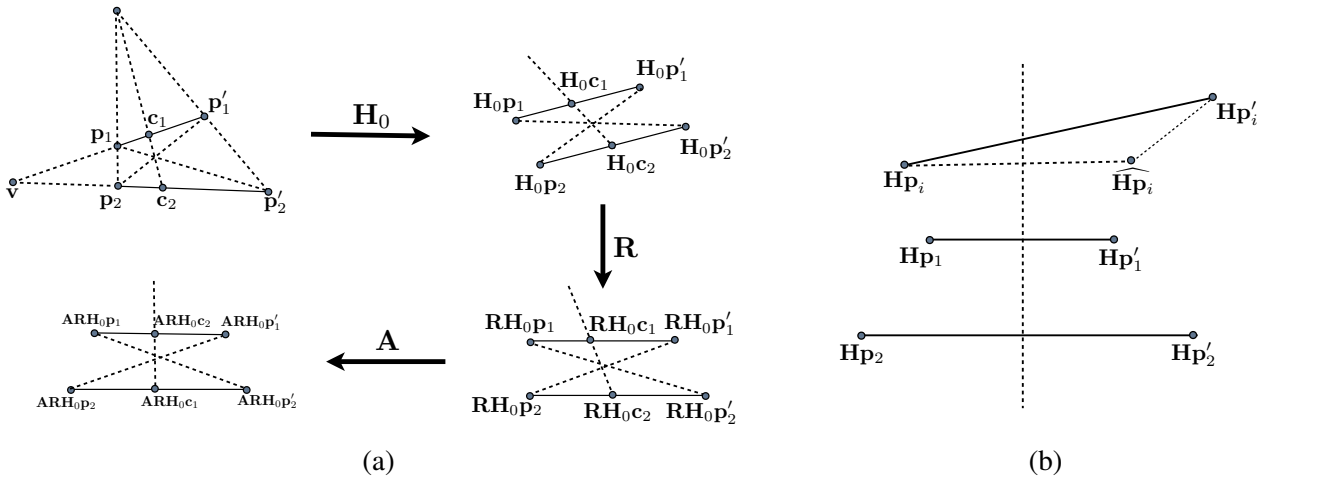
$$h_{31}v_x + h_{32}v_y + 1 = 0. \tag{2}$$

For choosing the remaining degree of freedom we add the constraint that the required scale change factor at the four keypoints of m_1 and m_2 is as close as to 1 as possible. Intuitively, among all possible $\mathbf{H}_0$, we select the one that is closer to the identity matrix, i.e., the one introducing the least perspective effect. This amounts to computing

$$\min_{h_{31},h_{32}} \sum_{i=1,2} (h_{31}x_i + h_{32}y_i)^2 + (h_{31}x'_i + h_{32}y'_i)^2. \tag{3}$$

Geometrically, this choice can be interpreted by rotating the camera as little as possible to make the vanishing point at the infinity. By plugging Eq. (2) in this minimization, we obtain a one-variable least-square problem which can be easily solved.

For computing $\mathbf{R}$, we simply constrain the vanishing point (now at infinity) to lie on the x axis. Finally, $\mathbf{A}$ is obtained by estimating the shear that aligns the two corresponding midpoints $\mathbf{c}_1, \mathbf{c}_2$ with the y axis.



(a) (b)

Figure 1: (a) Finding the rectification homography ($\mathbf{H} = \mathbf{ARH}_0$) from a pair of matches. (b) Compatible matches test given a rectified symmetry.

The homography $\mathbf{H}$ is then applied to the keypoints of all the matches $m_i \in \mathcal{M}$. In this rectified plane, all the matches compatible with the reflective symmetry defined by m_1, m_2 should be parallel with the seed line segments $\overline{m}_1, \overline{m}_2$ and have their midpoint on the symmetric axis. Obviously, this is an ideal scenario and in practice we need to relax this criterion. Let $\widehat{\mathbf{Hp}_i}$ (resp. $\widehat{\mathbf{Hp}'_i}$) be the point that is actually symmetric to $\mathbf{Hp}_i$ (resp. $\mathbf{Hp}'_i$). We then measure the ratio between the segment with endpoints $\mathbf{Hp}'_i$, $\widehat{\mathbf{Hp}_i}$ (resp. $\mathbf{Hp}_i$, $\widehat{\mathbf{Hp}'_i}$) and the segment with endpoints $\mathbf{Hp}_i$, $\widehat{\mathbf{Hp}_i}$ (resp. $\mathbf{Hp}'_i$, $\widehat{\mathbf{Hp}'_i}$). See Fig. 1(b) for a graphical representation. If both ratios are smaller than a precision parameter η, the match m_i is added to the consensus set.

2.2 Validating the seed matches

We need to ensure that the two randomly sampled seed matches m_1 and m_2 lead to a valid symmetry model. We perform the following sanity checks in order to reject invalid models:

- Since we are dealing with reflective symmetries, the segments $\overline{m}_1$ and $\overline{m}_2$ must not intersect before and after rectification;
- The keypoint scales must become approximately similar after rectification,

$$|\hat{s}(\mathbf{p}_i) - \hat{s}(\mathbf{p}'_i)| / \max(\hat{s}(\mathbf{p}_i), \hat{s}(\mathbf{p}'_i)) < \delta, \; i = 1,2, \tag{4}$$

  where $\hat{s}(\mathbf{p}_i)$ denotes the scale at $\mathbf{p}_i$ after rectification and δ is a precision parameter;
- The sum of the orientations $\hat{\theta}_i, \hat{\theta}'_i$ at $\mathbf{p}_i, \mathbf{p}'_i$ after rectification, must be approximately equal to π (these orientations are not very robust). We check the condition [16] (see Fig. 2)

$$1 + \cos(\hat{\theta}_i + \hat{\theta}'_i) < \varepsilon, \; i = 1,2, \tag{5}$$

  where ε is a precision parameter.

We could have used these sanity checks to constrain more the homography $\mathbf{H}$. But since the scale and orientation of the keypoints is not very precise, the estimation of the homography can be instable.

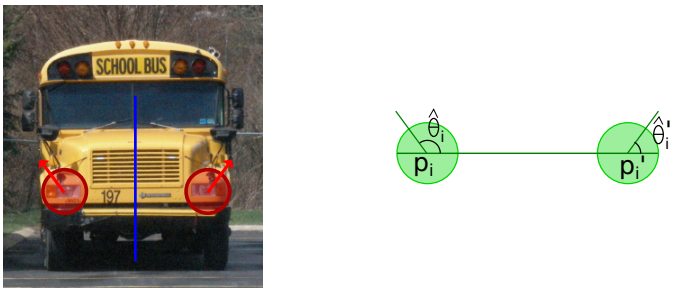


Figure 2: A schematic representation of two matching keypoints. On the right, the match $m_i = \{\mathbf{p}_i, \mathbf{p}'_i\}$ with corresponding angles $\hat{\theta}_i$, $\hat{\theta}'_i$ (i.e., θ_i, θ'_i transformed by the rectification).

2.3 Adaptively sampling seed matches

One of the keys for the success of the J-linkage algorithm is choosing a proper non-uniform sampling. The rationale behind this is the intent to oversample the *true* symmetries in the image, thus obtaining stable row features that robustify the clustering process. Among all of the sampled model seeds, it is desirable that there is at least a certain number of outlier-free model seeds sampled around each underlying true model. In [19, 24], non-uniform sampling is used to detect multiple simple geometric models in point clouds, like lines, circles, or planes, by assigning a higher probability to neighboring points.

We use the following adaptive non-uniform sampling strategy. The first match is sampled according to the following mixed probability, $\forall m_i = \{\mathbf{p}_i, \mathbf{p}'_i\} \in \mathcal{M}$,

$$\Pr(m_i) = \frac{1}{Z_1}\exp\left(-\frac{1}{\sigma_d^2}(\|\mathrm{SIFT}(\mathbf{p}_i) - \mathrm{SIFT}(\mathbf{p}'_i)\| - d_0)^2 - \frac{1}{\sigma_l^2}(\|\mathbf{p}_i - \mathbf{p}'_i\| - l_0)^2\right), \quad (6)$$

where d_0 and l_0 indicate the scale in the descriptor domain and image domain at which we prefer to detect the symmetry, Z_1 is a normalization factor such that $\sum_i \Pr(m_i) = 1$, and, finally, σ_d and σ_l decide how strict the preferences are. Ideally, we should have $d_0 = 0$, but since some parts of the image match closely than others, we might end up missing some symmetries. We thus relax this constraint.

In practice, an image can contain multiple symmetries at different scales. Automatically updating the parameters of Eq. (6) to find all possible symmetries is not an easy task. This motivates us to update the sampling probability of the first match along the sampling process. Each time two seed matches are sampled and its consensus set $\mathcal{M}_0$ is computed, we decrease the probability of all these matches by a factor κ close to 1, followed by the renormalization of the probability:

$$\text{probability update (followed by renormalization)}: \quad \Pr(m_i) \leftarrow \kappa \Pr(m_i),\ m_i \in \mathcal{M}_0. \quad (7)$$

This guarantees that the sampling will not solely focus on symmetries with "high quality" matches and "lower quality" matches will also be visited.

Once the first match is sampled following the above adaptive non-uniform sampling, we sample the second match according to the conditional probability:

$$\Pr(m_j|m_i) = \frac{1}{Z_2}\exp\left(-\frac{1}{\sigma_c^2}(\|\mathbf{c}_i - \mathbf{c}_j\| - c_0)^2\right),\ m_i \in \mathcal{M},\ m_j \in \mathcal{M} \setminus \{m_i\}, \quad (8)$$

where Z_2 is a normalization factor such that $\sum_j \Pr(m_j|m_i) = 1$, $\mathbf{c}_i$ is the midpoint of line segment $\overline{m}_i$, and σ_c, c_0 control the shape of the probability function.

Algorithm 1: Non-uniform adaptive sampling.

Data: matches $\mathcal{M} = \{m_i\}, i = 1, \cdots, N$
Result: preference matrix P
Initialize the sampling probability $\Pr(m_i)$ and $\Pr(m_j|m_i)$ according to Eq. (6) and (8) ;
Sampling counter $K_{counter} \leftarrow 0$;
$flag\_uniform\_sampling = 0$;
Consensus set to be uniformly sampled $\mathcal{C}_0 \leftarrow \emptyset$;
while $K_{counter} < K$ **do**
  $K_{trial} \leftarrow 0$;
  while $K_{trial} < K_{max\_trial}$ **do**
    $K_{trial} \leftarrow K_{trial} + 1$;
    if $flag\_uniform\_sampling = 0$ **then**
      Sample two seed matches according to probability $\Pr(m_i)$ and $\Pr(m_j|m_i)$;
    else
      Sample two seed matches uniformly in the consensus set $\mathcal{C}_0$;
    Rectify m_1, m_2 (see Section 2.1) ;
    If the seeds do not form a valid model (see Section 2.2) continue sampling;
  If the seeds do not form a valid model (see Section 2.2) terminate the algorithm;
  Compute the consensus set $\mathcal{C}$ for the rectified m_1, m_2 (see Fig. 1(b)) ;
  Add the consensus set $\mathcal{C}$ to preference matrix P;
  Update the probability $\Pr(m_i)$ according to Eq. (7) ;
  $K_{counter} \leftarrow K_{counter} + 1$;
  if $flag\_uniform\_sampling = 1$ **then**
    $K_{uniform\_sampling} \leftarrow K_{uniform\_sampling} + 1$;
  if $flag\_uniform\_sampling = 0 \ \wedge \ |\mathcal{C}| > T$ **then**
    $flag\_uniform\_sampling \leftarrow 1$;
    $K_{uniform\_sampling} \leftarrow 0$;
    Initialize the set to be uniformly sampled: $\mathcal{C}_0 \leftarrow \mathcal{C}$;
  if $flag\_uniform\_sampling = 1 \ \wedge \ |\mathcal{C}| > |\mathcal{C}_0|$ **then**
    Update the set to be uniformly sampled: $\mathcal{C}_0 \leftarrow \mathcal{C}$;
  if $flag\_uniform\_sampling = 1 \ \wedge \ K_{uniform\_sampling} = K_u$ **then**
    $flag\_uniform\_sampling \leftarrow 0$;

If the current consensus set $\mathcal{C}$ is big enough, i.e., $|\mathcal{C}| > T$ for some T, we switch to uniformly sample K_u seeds *inside* $\mathcal{C}$. This ensures that good models are oversampled.

The overall adaptive non-uniform sampling procedure is summarized in Alg. 1. Given the preference matrix, we run the J-linkage algorithm to cluster the matches.

3 Experimental results

For all experiments, we randomly select $N = 3000$ keypoints/features from the ones provided by ASIFT [15]. These features are matched to their mirrored version by nearest neighbors matching (one neighbor per feature for single symmetry detection and four neighbors per feature for multiple symmetry detection). We initialize the probabilities in eqs. (6) and (8)

	Synthetic Single				Synthetic Multiple			
	LE [11]	LHXS [10]	CL [2]	Proposed	LE [11]	LHXS [10]	CL [2]	Proposed
TP/GT	92%	62%	100%	**100%**	35%	28%	77%	**67%**
FP/GT	15%	0%	15%	**0%**	4%	8%	33%	**10%**

	Real Single				Real Multiple			
	LE [11]	LHXS [10]	CL [2]	Proposed	LE [11]	LHXS [10]	CL [2]	Proposed
TP/GT	84%	29%	94%	**97%**	43%	18%	68%	**65%**
FP/GT	68%	3%	69%	**39%**	44%	0%	17%	**16%**

Table 1: Performance comparison of several methods on the PSU dataset [12]. TP, FP, and GT respectively denote the number of true positives, false positives, and ground truth symmetries. The percentage of the methods are taken from [12] and [2].

with the values

$$d_0 = \frac{1}{2N} \sum_{m_i \in \mathcal{M}} \|\text{SIFT}(\mathbf{p}_i) - \text{SIFT}(\mathbf{p}'_i)\|, \qquad \sigma_d^2 = \frac{1}{10} \max_{m_i \in \mathcal{M}} \|\text{SIFT}(\mathbf{p}_i) - \text{SIFT}(\mathbf{p}'_i)\|^2, \tag{9}$$

$$l_0 = \frac{1}{10} \left(w^2 + h^2\right)^{1/2}, \qquad \sigma_l^2 = \frac{1}{10} \max_{m_i \in \mathcal{M}} \|\mathbf{p}_i - \mathbf{p}'_i\|^2, \tag{10}$$

$$c_0 = \frac{1}{20} \left(w^2 + h^2\right)^{1/2}, \qquad \sigma_c^2 = \frac{1}{10} \max_{m_i, m_j \in \mathcal{M}} \|\mathbf{c}_i - \mathbf{c}_j\|^2, \tag{11}$$

where w and h are the width and height of the image, respectively. The probability $\Pr(m_i)$ is updated using $\kappa = 0.98$ in Eq. (7). We check the validity of every non-uniformly sampled match pair by setting $\delta = 0.2$ and $\varepsilon = 0.25$ in eqs. (4) and (5). The matches compatible with the sampled symmetry models with parameter $\eta = 0.04$ (Fig. 2b) are considered as inliers and added to the consensus sets. We also set $K = 4000$, $T = 10$, and $K_u = 30$. After the J-linkage clustering, only symmetries containing at least 10 matches are kept. We finally apply a non-maximum suppression on the clusters whose symmetry axes are close.

We first compare the proposed method with three recent ones [2, 10, 11] on the PSU dataset [12], which is composed of 88 images.[1] The symmetries are either frontal or slightly skewed. This dataset covers four types of images: synthetic single reflection, synthetic multiple reflection, natural single reflection, and natural multiple reflection. We should point out that even though PSU provides a reliable benchmark dataset for comparing symmetry detection algorithms, the ground truth it provides is not always complete and/or accurate. The results in Table 1 (see some examples in Fig. 3) show that our method is much better than Loy and Eklundh's (LE [11]) and Liu *et al.*'s (LHXS [10]). Compared with Cho *et al.*'s (CL [2]) method, we have better performance for single symmetry detection. For multiple symmetries detection, our method is more conservative, sometimes detecting fewer symmetries, which explains the decrease in both the true positive and false positives rates. This conservative strategy can be explained by the strict criterion used to only create precise consensus sets. Nonetheless, the proposed method results are highly competitive. Notice that, as a post-processing, we could also adopt Loy and Eklundh's region growing strategy [11], in order to expand and further improve the detected symmetries.

[1] The original PSU dataset contained 91 images [2, 12] but the original download link is broken. We obtained a version of the dataset containing 88 images at http://vivid.cse.psu.edu/texturedb/gallery/album05. Since code is not available for the methods in [2, 10], the results are not exactly comparable but serve as a performance indicator.

In non fronto-parallel symmetries, our algorithm is also able to obtain good results. We include several examples in Fig. 4. The algorithms that were used for comparing in the frontal setting are not designed for this type of symmetries and would only succeed in those images where the perspective effect is weak.

Recently, the TILT algorithm [23] was introduced, combining the concept of low-rank factorization with perspective estimation. Given a manually selected region of interest, it computes the homography that produces the sub-image with the lowest possible rank. In Fig. 5, we show that the detected symmetries with their consensus sets (supporting matches) can be used to automatically select input regions of interest for the TILT algorithm. This brings forth the relationship between these two concepts: symmetric regions have necessarily a low rank. Even though the regions are not completely flat, the TILT algorithm is capable of transforming the regions to be a quasi frontal view. Notice that using the rectifying homography as an initialization of the TILT optimization might also help further improve and stabilize the TILT results by obtaining a visually better local minimum.

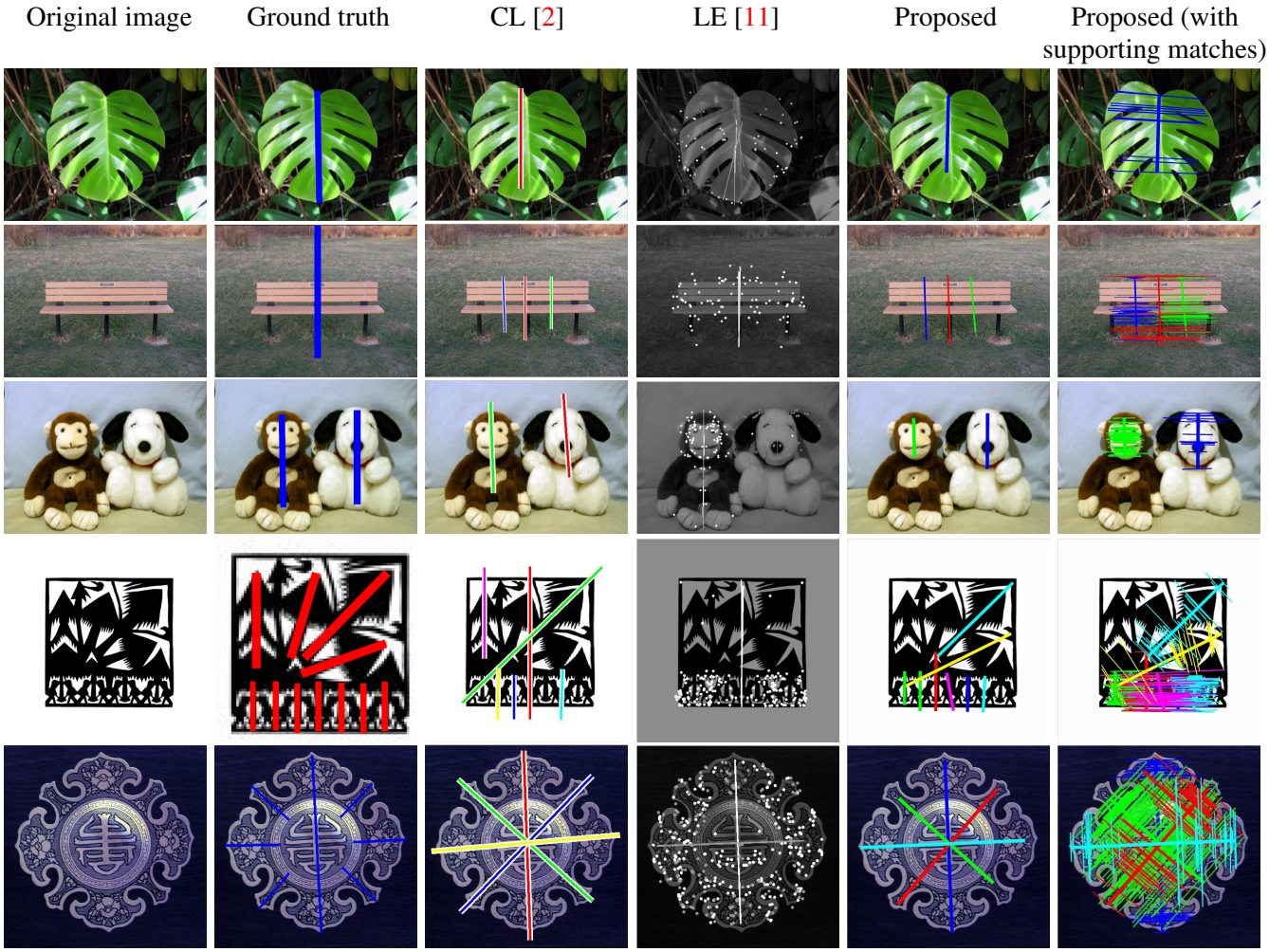


Figure 3: Several symmetry detection results on the PSU dataset. The images on the third row are borrowed from [2].

4 Conclusion

We presented a method for detecting bilateral symmetries in images. The method detects symmetries in a rectified image domain by sampling symmetry seeds in a non-uniform adaptive manner, and then building candidate consensus sets. Features are built from the consen-

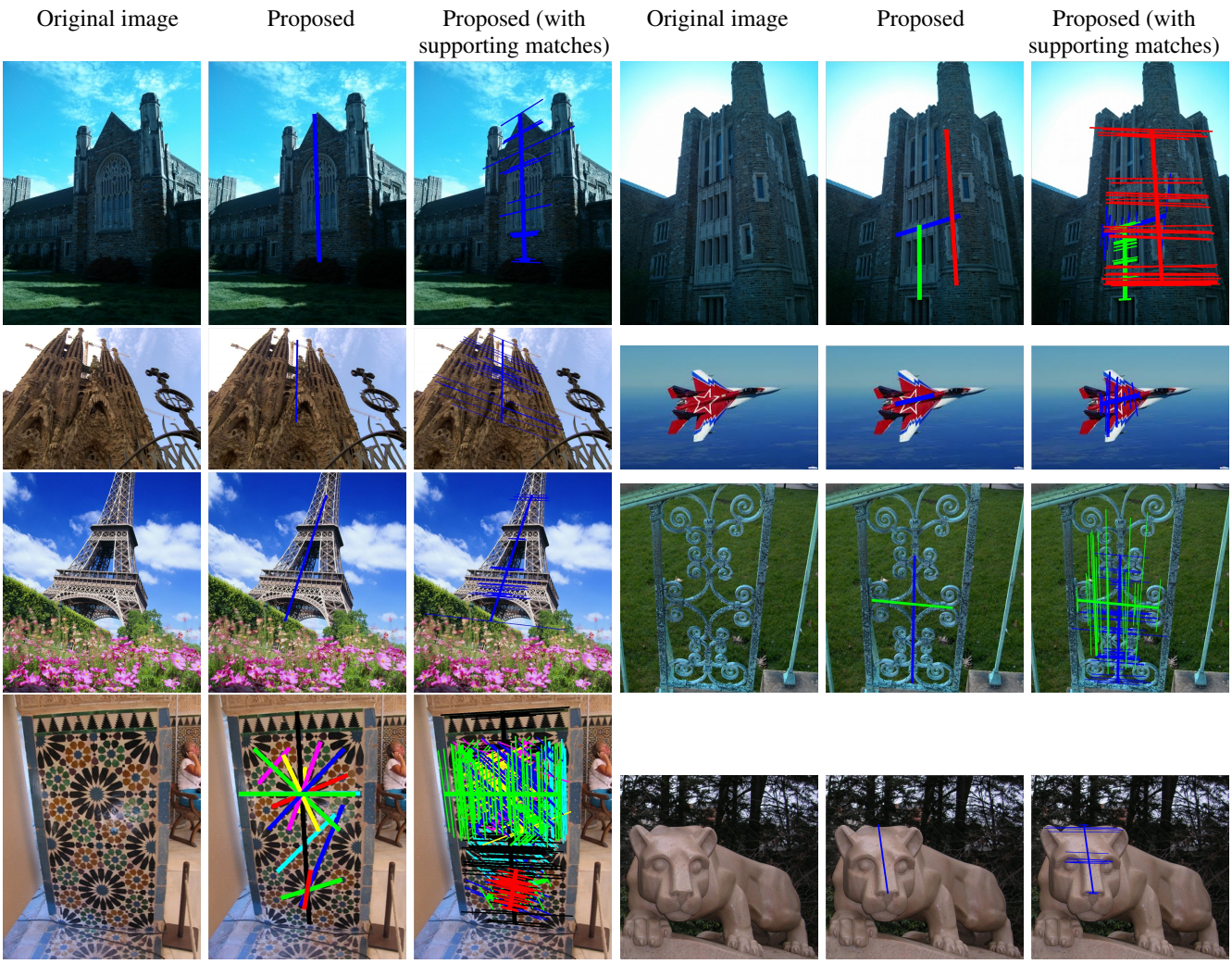


Figure 4: Several non fronto-parallel symmetry examples. Our algorithm is able to recover the symmetries, even when the perspective effect is non-negligible.

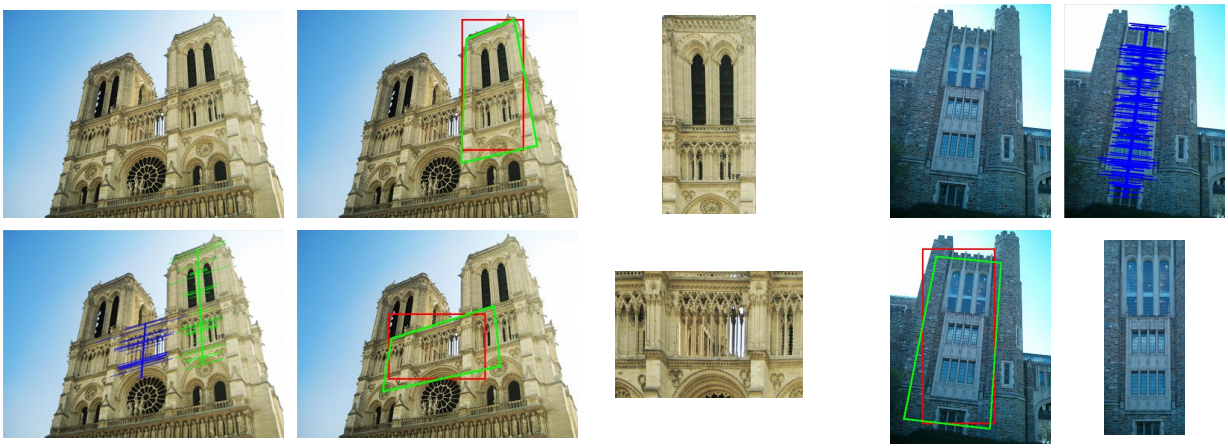


Figure 5: Relationship between low rank and symmetry in two examples of images with detected symmetries. The red bounding boxes represent the regions of the detected symmetry, which are used as automatic input to the TILT algorithm. The green bounding boxes represent the low-rank sub-images obtained with the TILT algorithm, projected back to the original image. We also show the low-rank components before re-projection.

sus sets and the final symmetries are detected via an agglomerative clustering algorithm. The method is able to detect single and multiple symmetries both in frontal and skewed (non fronto-parallel) viewpoints, achieving state-of-the-art results. We plan on extending the method to detect other symmetry types. We are also investigating a more unified way to combine the concept of symmetry and low-rank.

References

[1] G. Birkoff. *Aesthetic Measure*. Havard University Press, Cambridge, MA, 1932.

[2] M. Cho and K. Mu Lee. Bilateral symmetry detection via symmetry-growing. In *BMVC*, 2009.

[3] R.W. Conners. Developing a quantitative model of human preattentive vision. *IEEE Trans Syst Man Cybern*, 19(6):1384–1407, 1989.

[4] H. Cornelius and G. Loy. Detecting bilateral symmetry in perspective. In *POCV*, 2006.

[5] H. Cornelius, M. Perdoch, J. Matas, and G. Loy. Efficient symmetry detection using local affine frames. In *SCIA*, 2007.

[6] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *CVPR*, 2010.

[7] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM*, 24(6): 381–395, 1981.

[8] Y. Keller and Y. Shkolnisky. An algebraic approach to symmetry detection. In *ICPR*, 2004.

[9] A. Kuehnle. Symmetry-based recognition of vehicle rears. *Pattern Recognit Lett*, 12 (4):249–258, 1991.

[10] Y. Liu, J. Hays, Y-Q. Xu, and H-Y. Shum. Digital papercutting. Technical Sketch. In *SIGGRAPH*, 2005.

[11] G. Loy and J-O. Eklundh. Detecting symmetry and symmetric constellations of features. In *ECCV*, 2006.

[12] S. Lee M. Park, P-C. Chen, S. Kashyap, A. Butt, and Y. Liu. Performance evaluation of state-of-the-art discrete symmetry detection algorithms. CVPR, 2008.

[13] G. Marola. On the detection of the axes of symmetry of symmetric and almost symmetric planar images. *IEEE Trans Pattern Anal Mach Intell*, 11(1):104–108, 1989.

[14] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *BMVC*, 2002.

[15] J.M. Morel and G.Yu. ASIFT: A new framework for fully affine invariant image comparison. *SIAM J Imaging Sci*, 2(2):438–469, 2009.

[16] D. Reisfeld, H. Wolfson, and Y. Yeshurun. Context-free attentional operators: The generalized symmetry transform. *Int J Comput Vis*, 14:119–130, 1995.

[17] C. Sun and D. Si. Fast reflectional symmetry detection using orientation histograms. *Real-Time Imaging*, 5:63–74, 1999.

[18] Y. Sun and B. Bhanu. Reflection symmetry-integrated image segmentation. *IEEE Trans Pattern Anal Mach Intell*, 34(9):1827–1841, 2012.

[19] R. Toldo and A. Fusiello. Robust multiple structures estimation with j-linkage. *ECCV*, 2008.

[20] T. Tuytelaars, A. Turina, and L. Van Gool. Noncombinatorial detection of regular repetitions under perspective skew. *IEEE Trans Pattern Anal Mach Intell*, 25(4):418–432, 2003.

[21] H. Weyl. *Symmetry*. Princeton University Press, 1952.

[22] L. Xu, E. Oja, and P. Kultanen. A new curve detection method: Randomized Hough transform (RHT). *Pattern Recognit Lett*, 11(5):331–338, 1990.

[23] Z. Zhang, A. Ganesh, X. Liang, and Y. Ma. TILT: Transform invariant low-rank textures. *Int J Comput Vis*, 99(1):1–24, 2012.

[24] M. Zuliani, C.S. Kenney, and B.S. Manjunath. The multiransac algorithm and its application to detect planar homographies. In *ICIP*, 2005.

Multi image noise estimation and denoising

A. Buades * Y. Lou † J.M Morel ‡ Z. Tang‡

Abstract

Photon accumulation on a fixed surface is the essence of photography. In the times of chemical photography this accumulation required the camera to move as little as possible, and the scene to be still. Yet, most recent reflex and compact cameras propose a burst mode, permitting to capture quickly dozens of short exposure images of a scene instead of a single one. This new feature permits in principle to obtain by simple accumulation high quality photographs in dim light, with no motion or aperture blur. It also gives the right data for an accurate noise model. Yet, both goals are attainable only if an accurate cross-registration of the burst images has been performed. The difficulty comes from the non negligible image deformations caused by the slightest camera motion, in front of a 3D scene, and from the light variations or motions in the scene. This chapter proposes a numerical processing chain permitting to achieve jointly the two mentioned goals: an accurate noise model for the camera, which is used crucially to obtain a state of the art multi-images denoising. The key feature of the proposed processing chain is a reliable multi-image noise estimator, whose accuracy will be demonstrated by three different procedures. Thanks to the signal dependent noise model obtained from the burst itself, a faithful detection of the well registered pixels can be made. The denoising by simple accumulation of these pixels, which are an overwhelming majority, permits to extend the Nicéphore Niepce photon accumulation method to image bursts. The denoising performance by accumulation is shown to reach the theoretical limit, namely a $\sqrt{n}$ denoising factor for n frames. Comparison with state of the art denoising algorithms will be shown on several bursts taken with reflex cameras in dim light.

1 Introduction

The accumulation of photon impacts on a surface is the essence of photography. The first Nicephore Niepce photograph [20] was obtained after an eight hours exposure. The serious objection to a long exposure is the variation of the scene due to changes in light, camera motion, and incidental motions of parts of the scene. The more these variations can be compensated, the longer the exposure can be, and the more the noise can be reduced. It is a frustrating experience for professional photographers to take pictures under bad lighting conditions with a hand-held camera. If the camera is set to a long exposure time, the photograph gets blurred by the camera motions and aperture. If it is taken with short exposure, the image is dark, and enhancing it reveals the noise. Yet, this dilemma can be solved by taking a burst of images, each with short-exposure time, as shown in Fig. 1, and by averaging them after registration. This observation is not new and many algorithms have been proposed, mostly for stitching and super-resolution. These algorithms have thrived in the last decade, probably thanks to the discovery of

*MAP5, CNRS - Universite Paris Descartes, 45 rue Saints Peres, 75270 Paris Cedex 06, France.
†Mathematics Department, University of California at Los Angeles, Los Angeles, U.S.A.
‡CMLA, ENS Cachan, 61 av. President Wilson, Cachan 94235, France.

a reliable algorithm for image matching, the SIFT algorithm [53]. All of the multi-image fusion algorithms share three well separated stages, the search and matching of characteristic points, the registration of consecutive image pairs and the final accumulation of images. All methods perform some sort of multi-image registration, but surprisingly do not propose a procedure to check if the registration is coherent. Thus, there is a non-controlled risk that the accumulation blurs the final accumulation image, due to wrong registrations. Nevertheless, as we shall see, the accurate knowledge of noise statistics for the image sequence permits to detect and correct all registration incoherences. Furthermore, this noise statistics can be most reliably extracted from the burst itself, be it for raw or for JPEG images. In consequence, a stand alone algorithm which denoises any image burst is doable. As experiments will show, it even allows for light variations and moving objects in the scene, and it reaches the $\sqrt{n}$ denoising factor predicted for the sum of the n independent (noise) random variables.

We call in the following "burst", or "image burst" a set of digital images taken from the same camera, in the same state, and quasi instantaneously. Such bursts are obtained by video, or by using the burst mode proposed in recent reflex and compact cameras. The camera is supposed to be held as steady as possible so that a large majority of pixels are seen through the whole burst. Thus, no erratic or rash motion of the camera is allowed, but instead incident motions in the scene do not hamper the method.

There are other new and promising approaches, where taking images with different capture conditions is taken advantage of. Liu et al. [86] combine a blurred image with long-exposure time, and a noisy one with short-exposure time for the purpose of denoising the second and deblurring the first. Beltramio and Levine [10] improve the dynamic range of the final image by combining an underexposed snapshot with an overexposed one. Combining again two snapshots, one with and the other without flash, is investigated by Eisemann *et. al.* [33] and Fattal *et. al* [37]. Another case of image fusion worth mentioning is [8], designed for a 3D scanning system. During each photography session, a high-resolution digital back is used for photography, and separate macro (close-up) and ultraviolet light shots are taken of specific areas of text. As a result, a number of folios are captured with two sets of data: a "dirty" image with registered 3D geometry and a "clean" image with the page potentially deformed differently to which the digital flattening algorithms are applied.

Our purpose here is narrower. We only aim at an accurate noise estimation followed by denoising for an image burst. No super-resolution will be attempted, nor the combination of images taken under different apertures, lightings or positions. The main assumption on the setting is that a hand-held camera has taken an image burst of a still scene, or from a scene with a minority of moving objects. To get a significant denoising, the number of images can range from 9 to 64, which grants a noise reduction by a factor 3 to 8. Since the denoising performance grows like the square root of the number of images, it is less and less advantageous to accumulate images when their number grows. But impressive denoising factors up to 6 or 8 are reachable by the simple algorithm proposed here, which we shall call *average after registration* (AAR). Probably the closest precursor to the present method is the multiple image denoising method by Zhang *et. al.* [89]. Their images are not the result of a burst. They are images taken from different points of views by different cameras. Each camera uses a small aperture and a short exposure to ensure minimal optical defocus and motion blur, to the cost of very noisy output. A global registration evaluating the 3D depth map of the scene is computed from the multi-view images, before applying a patch based denoising inspired by NL-means [15]. Thus the denoising strategy is more complex than the simple accumulation after registration which is promoted in the present chapter. Nevertheless, the authors remark that their denoising performance stalls when the number of frames grows, and write that this difficulty should be overcome. Yet, their observed denoising performance curves grow approximately like the square root of the number of frames, which indicates that the real performance of the algorithm is due to the accumulation. The method proposed here therefore goes back to accumulation, as the essence of photography.

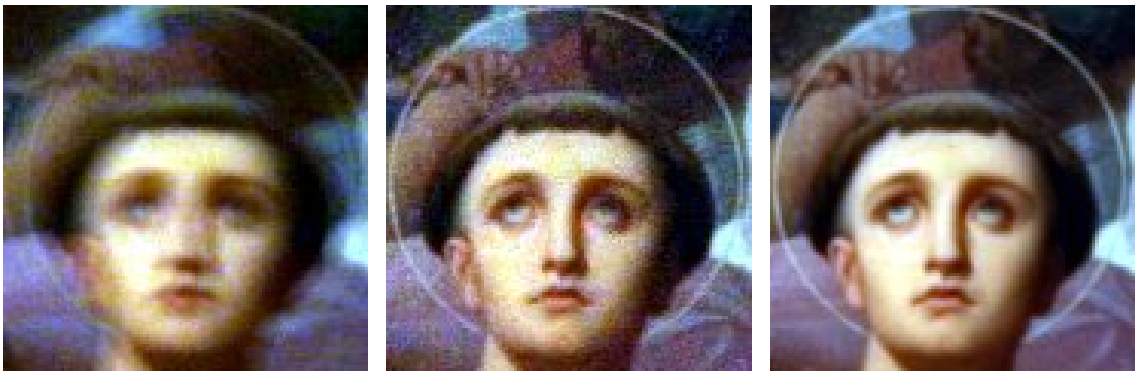


Figure 1: From left to right: one long-exposure image (time = 0.4 sec, ISO=100), one of 16 short-exposure images (time = 1/40 sec, ISO = 1600) and the average after registration. All images have been color-balanced to show the same contrast. The long exposure image is blurry due to camera motion. The middle short-exposure image is noisy, and the third one is some **four times** less noisy, being the result of averaging 16 short-exposure images. Images may need to be zoomed in on a screen to compare details and textures.

It uses, however, a hybrid scheme which decides at each pixel between accumulation and block denoising, depending on the reliability of the match. The comparison of temporal pixel statistics with the noise model extracted from the scene itself permits a reliable conservative decision so as to apply or not the *accumulation after registration* (AAR). Without the accurate nonparametric noise estimation, this strategy would be unreliable. Therefore estimating accurately the noise model in a burst of raw or JPEG images is the core contribution of this chapter. A more complex and primitive version of the hybrid method was announced in the conference paper [17]. It did not contain the noise estimation method presented here.

Plan and Results The chapter requires a rich bibliographical analysis for the many aspects of multi-image processing (Section 3). This survey shows that most super-resolution algorithms do in fact much more denoising than they do super-resolution, since they typically only increase the size of the image by a factor 2 or 3, while the number of images would theoretically allow for a 5 to 8 factor. Section 2 reviews the other pilar of the proposed method, the noise estimation literature. (This corpus is surprisingly poor in comparison to the denoising literature.)

Section 4 is key to the proposed technique, as it demonstrates that a new variant of static noise blind estimate gives results that exactly coincide with Poisson noise estimates taken from registered images in a temporal sequence. It is also shown that although JPEG images obtained by off-the-shelf cameras have no noise model, a usable substitute to this noise model can be obtained: It simply is the variance of temporal sequences of registered images.

Section 5 describes the proposed multi-image denoising method, which in some sense trivializes the denoising technology, since it proposes to go back as much as possible to a mere accumulation, and to perform a more sophisticated denoising only at dubiously registered pixels. Section 6 compares the proposed strategy with two state of the art multi-images denoising strategies.

2 Noise Estimation, a Review

As pointed out in [51], "Compared to the in-depth and wide literature on image denoising, the literature on noise estimation is very limited". Following the classical study by Healey et al. [44], the noise in CCD sensors can be approximated by an additive, white and signal dependent noise model. The noise model and its variance reflect different aspects of the imaging chain at the CCD, mainly dark noise and shot noise. Dark noise is due to the creation of spurious electrons generated by thermal energy which become indistinguishable from photoelectrons. Shot noise is a result of the quantum nature of light and characterizes the uncertainty in the number of photons stored at a collection site. This number of photons follows a Poisson distribution so that its variance equals its mean. The overall combination of the different noise sources therefore leads to an affine noise variance $a + bu$ depending on the original signal value u. Yet, this is only true for the raw CCD image. Further processing stages in the camera hardware and software such as the white balance, the demosaicking, the gamma correction, the blur and color corrections, and eventually the compression, correlate the noise and modify its nature and its standard deviation in a non trivial manner. There is therefore no noise model for JPEG images. However, as we shall see, a signal dependent noise variance model can still be estimated from bursts of JPEG images (section 4.2.) It is enough to perform reliably the average after registration (AAR).

2.1 Additive Gaussian Noise Estimation

Most computer vision algorithms should adjust their parameters according to the image noise level. Surprisingly, there are few papers dealing with the noise estimation problem, and most of them only estimate the variance of a signal independent additive white Gaussian noise (AWGN). This noise statistics is typically measured on the highest-frequency portion of the image spectrum, or on homogenous image patches. In the AWGN case a spectral decomposition through an orthonormal transform such as wavelets or the DCT preserves the noise statistics. To estimate the noise variance, Donoho et. al [29] consider the finest scale wavelet coefficients, followed by a median filter to eliminate the outliers. Suppose $\{y_i\}_{i=1,\cdots N}$ be N independent Gaussian random variables of zero-mean and variance σ^2, then

$$E\{\text{MED}(|y_i|)\} \approx 0.6745\sigma.$$

It follows immediately that the noise standard deviation σ is given by

$$\tilde{\sigma} = \frac{1}{0.6745}\text{MED}(|y_i|) = 1.4826\text{MED}(|y_i|).$$

The standard procedure of the local approaches is to analyze a set of local estimates of the variance. For example, Rank et. al [71] take the maximum of the distribution of image derivatives. This method is based on the assumption that the underlying image has a large portion of homogeneous regions. Yet, if an image is highly textured, the noise variance can overestimated. To overcome this problem, Ponomarenko et. al [68] have proposed to analyze the local DCT coefficients. A segmentation-based noise estimation is carried out in [1], which considers both i.i.d. and spatially correlated noise.

The algorithm in [69] is a modification of the early work [68] dealing with AVIRIS (Airborne Visible Infrared Imaging Spectrometer) images, in which the evaluation of the noise variance in sub-band images is addressed. The idea is to divide each block into low frequency and high frequency components by thresholding, and to use K blocks of the smallest variance of the low frequency coefficients to calculate a noise variance, where K is adaptively selected so that it is smaller for highly-textured images.

[25] proposed an improvement of the estimate of the variance of AWGN by using transforms creating a sparse representation (via BM3D [22]) and using robust statistics estimators

(MAD and ICI). For a univariate data set $X_1, X_2, ..., X_n$, the MAD is defined as the median of the absolute deviations from the data's median: $\text{MAD} = \text{median}_i\,(\,|X_i - \text{median}_j(X_j)|\,)$. The algorithm is as follows.

1. for each 8×8 block, group together up to 16 similar non-overlapping blocks into 3D array. The similarity between blocks in evaluated by comparing corresponding blocks extracted from a denoised version by BM3D.
2. apply a 3-D orthonormal transform (DCT or wavelet) on each group and sort the coefficients according to the zig-zag scan.
3. collect the first 6 coefficients $c_1, \cdots, c_6$ and define their empirical energy as the mean of the magnitude of the (up to 32) subsequent coefficients:

$$E\{|c_j|^2\} = \text{mean}\{|c_{j+1}^2, \cdots, c_{j+32}^2|\}$$

4. Sort the coefficients from all the groups (6 coefficients per group) according to their energy
5. do MAD and Intersection of Confidence Intervals (ICI) [42] to achieve the optimal bias-variance trade-off in the MAD estimation.

All the above mentioned algorithms give reasonable estimates of the standard deviation **when the noise is uniform**. Yet, when applying these algorithms to estimate signal dependent noise, the results are poor. The work of C. Liu *et. al.* [52] estimates the upper bound on the noise level fitting to a camera model. The noise estimation from the raw data is discussed in [39, 40]. The former is a parametric estimation by fitting the model to the additive Poissonian-Gaussian noise from a single image, while the latter measures the temporal noise based on an automatic segmentation of 50 images.

2.2 Poisson Noise Removal

This chapter deals with real noise, which in most real images (digital cameras, tomography, microscopy and astronomy) is a Poisson noise. The Poisson noise is inherent to photon counting. This noise adds up to a thermal noise and an electronic noise which are approximately AWGN. In the literature algorithms considering the removal of AWGN are dominant but, if its model is known, Poisson noise can be approximately reduced to AWGN by a so called variance stabilizing transformation (VST). The standard procedure follows three steps,

1. apply VST to make the data homoscedastic
2. denoise the transformed data
3. apply the inverse VST.

The square-root operation is widely used as a VST,

$$f(z) = b\sqrt{z + c}. \tag{1}$$

It follows from the asymptotic unit variance of $f(z)$ that the parameters are given by $b = 2$ and $c = 3/8$, which is the Anscombe transform [2]. A multiscale VST (MS-VST) is studied in [88] along with the conventional denoising schemes based on wavelets, ridgelets and curvelets depending on morphological features (isotropic, line-like, curvilinear, etc) of the given data. It is argued in [55] that the inverse transformation of VST is crucial to the denoising performance. Both the algebraic inverse

$$\mathcal{I}_A(D) = \left(\frac{D}{2}\right)^2 - \frac{3}{8}.$$

and the asymptotically unbiased inverse

$$\mathcal{I}_B(D) = \left(\frac{D}{2}\right)^2 - \frac{1}{8},$$

in [2] are biased for low counts. The authors [55] propose an exact unbiased inverse. They consider an inverse transform $\mathcal{I}_C$ that maps the value $E\{f(z)|y\}$ to the desired value $Ez|y$ that

$$E\{f(z)|y\} = 2\sum_{z=0}^{\infty}\left(\sqrt{z+\frac{3}{8}}\cdot\frac{y^z\exp^{-y}}{z!}\right)$$

where $f(z)$ is the forward Anscombe transform (1). In practice, it is sufficient to compute the above equation for a limited set of values y and approximate $\mathcal{I}_C$ by $\mathcal{I}_B$ for large values of y. Furthermore, the state-of-the-art denoising scheme BM3D [39] is applied in the second step.

There are also wavelets based methods [65, 48] or Bayesian [78, 54, 49] removing Poisson noise. In particular, the wavelet-domain Wiener filter [65] uses a cross-validation that not only preserves important image features, but also adapts to the local noise level of the spatially varying Poisson process. The shrinkage of wavelet coefficients investigates how to correct the thresholds [48] to explicitly account for effects of the Poisson distribution on the tails of the coefficient distributions. A recent Bayesian approach by Lefkimmiatis et al. [49] explores a recursive quad-tree image representation which is suitable for Poisson noise degradation and then follows an expectation-maximization technique for parameter estimation and Hidden Markov tree (HMT) structures for inter-scale dependencies. The common denominator to all such methods is that we need an accurate Poisson model, and this will be thoroughly discussed in Section 4.

It is, however, a fact that the immense majority of accessible images are JPEG images which contain a noise altered by a long chain of processing algorithms, ending with compression. Thus the problem of estimating noise in a single JPEG image is extremely ill-posed. It has been the object of a thorough study in [51]. This chapter proposes a blind estimation and removal method of color noise from a single image. The interesting feature is that it constructs a "noise level function" which is signal dependent, obtained by computing empirical standard deviations image homogeneous segments. Of course the remanent noise in a JPEG image is no way white or homogeneous, the high frequencies being notoriously removed by the JPEG algorithm. On the other hand, demosaicking usually acts as a villainous converter of white noise into very structured colored noise, with very large spots. Thus, even the variance of smooth regions cannot give a complete account of the noise damage, because noise in JPEG images is converted in extended flat spots. We shall, however, retain the idea promoted in [51] that, in JPEG images, a signal dependent model for the noise variance can be found. In section 4.2 a simple algorithm will be proposed to estimate the color dependent variance in JPEG images from multi-images. All in all, the problem of estimating a noise variance is indeed much better posed if several images of the same scene by the same camera, with the same camera parameters, are available. This technique is classic in lab camera calibration [44].

3 Multi-Images and Super Resolution Algorithms

Photo stitching Probably one of the most popular applications in image processing, photo stitching [14, 13] is the first method to have popularized the SIFT method permitting to register into a panorama a set of image of a same scene. Another related application is video stabilization [7]. In these applications no increase in resolution is gained, the final image has roughly the same resolution as the initial ones.

Super-Resolution Super-resolution means creating a higher resolution, larger image from several images of the same scene. Thus, this theme is directly related to the denoising of image bursts. It is actually far more ambitious, since it involves a deconvolution. However, we shall see that most super-resolution algorithms actually make a moderate zoom in, out of many images, and therefore mainly perform a denoising by some sort of accumulation. The convolution model in the found references is anyway not accurate enough to permit a strong deconvolution.

A single-frame super-resolution is often referred to as interpolation. See for example [83, 84]. But several exemplar-based super-resolution methods involve other images which are used for learning, like in Baker and Kanade [4] who use face or text images as priors. Similarly, the patch-example-based approaches stemming from the seminal paper [41], use a nearest-neighbor search to find the best match for local patches, and replace them with the corresponding high-resolution patches in the training set, thus enhancing the resolution. To make the neighbors compatible, the belief-propagation algorithm to the Markov network is applied, while another paper [26] considered a weighted average by surrounding pixels (analogue to nonlocal means [15]). Instead of a nearest-neighbor search, Yang et. al [81] proposed to incorporate the sparsity in the sense that each local patch can be sparsely represented as a linear combination of low-resolution image patches; and a high-resolution image is reconstructed by the corresponding high-resolution elements. The recent remarkable results of [85] go in the same direction. The example-based video enhancement is discussed in [11], where a simple frame-by-frame approach is combined with temporal consistency between successive frames. Also to mitigate the flicker artifacts, a stasis prior is introduced to ensure the consistency in the high frequency information between two adjacent frames.

Focus on Registration In terms of image registration, most of the existing super-resolution methods rely either on a computationally intensive optical flow calculation, or on a parametric global motion estimation. The authors of [92] discuss the effects of multi-image alignment on super-resolution. The flow algorithm they employ addresses two issues: flow consistency (flow computed from frame A to frame B should be consistent with that computed from B to A) and flow accuracy. The flow consistency can be generalized to many frames by computing a consistent bundle of flow fields. Local motion is usually estimated by optical flow, other local deformation models include Delaunay triangulation of features [8] and B-splines [60]. Global motion, on the other hand, can be estimated either in the frequency domain or by feature-based approaches. For example, Vandewalle et. al. [80] proposed to register a set of images based on their low-frequencies, aliasing-free part. They assume a planar motion, and as a result, the rotation angle and shifts between any two images can be precisely calculated in the frequency domain. The standard procedure for feature-based approaches is (1) to detect the key points via Harris corner [19, 3] or SIFT [87, 72], (2) match the corresponding points while eliminating outliers by RANSAC and (3) fit a proper transformation such as a homography. The other applications of SIFT registration are listed in Tab. 2.

Reconstruction after Registration A number of papers focus on image fusion, assuming the motion between two frames is either known or easily computed. Elad and Feuer [34] formulate the super-resolution of image sequences in the context of Kalman filtering. They assume that the matrices which define the state-space system are known. For example, the blurring kernel can be estimated by a knowledge of the camera characteristics, and the warping between two consecutive frames is computed by a motion estimation algorithm. But due to the curse of dimensionality of the Kalman filter, they can only deal with small images, e.g. of size 50×50. The work [56] by Marquina and Osher limited the forward model to be spatial-invariant blurring kernel with the down-sampling operator, while no local motion was present. They solved a TV-based reconstruction with Bregman iterations.

A joint approach on demosaicing and super-resolution of color images is addressed in [35], based on their early super-resolution work [36]. The authors use the bilateral-TV regularization for the spatial luminance component, the Tikhonov regularization for the chrominance component and a penalty term for inter-color dependencies. The motion vectors are computed via a hierarchical model-based estimation [9]. The initial guess is the result of the Shift-And-Add method. In addition, the camera PSF is assumed to be a Gaussian kernel with various standard deviation for different sets of experiments.

Methods Joining Denoising, Deblurring, and Motion Compensation Super-resolution and motion deblurring are crossed in the work [5]. First the object is tracked through the sequence, which gives a reliable and sub-pixel segmentation of a moving object [6]. Then a high-resolution is constructed by merging the multiple images with the motion estimation. The deblurring algorithm, which mainly deals with motion blur [46], has been applied only to the region of interest. The recent paper on super-resolution by L. Baboulaz and P. L. Dragotti [3] presents several registration and fusion methods. The registration can be performed either globally by continuous moments from samples, or locally by step edge extraction. The set of registered images is merged into a single image to which either a Wiener or an iterative Modified Residual Norm Steepest Descent (MRNSD) method is applied [63] to remove the blur and the noise. The super-resolution in [72] uses SIFT + RANSAC to compute the homography between the template image and the others in the video sequence, shifts the low-resolution image with subpixel accuracy and selects the closest image with the optimal shifts.

Implicit Motion Estimation More recently, inspired by the nonlocal movie denoising method, which claims that "denoising images sequences does not require motion estimation" [16], researchers have turned their attention towards super-resolution without motion estimation [32, 31, 70]. Similar methodologies include the steering kernel regression [76], BM3D [24] and its many variants. The forward model in [24] does not assume the presence of the noise. Thus the authors pre-filter the noisy LR input by V-BM3D [21]. They up-sample each image progressively m times, and at each time, the initial estimate is obtained by zero-padding the spectra of the output from the previous stage, followed by filtering. The overall enlargement is three times the original size. Super-resolution in both space and time is discussed in [73, 74], which combine multiple low-resolution video sequences of the same dynamic scene. They register any two sequences by a spatial homography and a temporal affine transformation, followed by a regularization-based reconstruction algorithm.

A Synoptic Table of Super-Resolution Multi-Images Methods Because the literature is so rich, a table of the mentioned methods, classified by their main features, is worth looking at. The methods can be characterized by a) their number k of fused images, which goes from 1 to 60, b) the zoom factor, usually 2 or 3, and therefore far inferior to the potential zoom factor $\sqrt{k}$, c) the registration method, d) the deblurring method, e) the blur kernel. A survey of the table demonstrates that a majority of the methods use many images to get a moderate zoom, meaning that the denoising factor is important. Thus, these methods denoise in some sense by accumulation. But, precisely because all of them aim at super-resolution, none of them considers the accumulation by itself.

Tables 1 and 2 confirm the dominance of SIFT+RANSAC as a standard way to register multi-images, as will also be proposed here in an improved variant. Several of the methods in Table 1 which do not perform SIFT+RANSAC, actually the last four rows, are "implicit". This means that they adhere to the dogma that denoising does not require motion estimation. It is replaced by multiple block motion estimation, like the one performed in NL-means and BM3D. However, we shall see in the experimental section that AAR (average after registration) has a still better performance than such implicit methods. This is one of the main questions that arose in this exploration, and the answer is clear cut: denoising by accumulation, like in ancient photography times still is a valid response in the digital era.

4 Noise Blind Estimation

In this section we return to noise estimation and will confront and cross-validate a single frame noise estimation with a multi-images noise estimation.

Table 1: comparison of Super Resolution algorithms

Ref.	# of images V.S. factor		Registration	Deblurring	blur kernel
[41] [4]	1	2 to 16	KNN to training set	NO	
[26]	1	2 3		MAP penalty	3×3 5×5
[81]	1	to 4	sparse w.r.t. traning	back-projection	Not mention
[19]	15	2	Harris+RANSAC	Tiknonov	Not mention
[18]	25	3	PCA	NO	
[92]	40	2	consistent flow bundle	NO	
[80]	4	2	frequency domain	NO	
[34]	100	2	assume known motion	Kalman filter	3×3 average
[36, 35]	30	3	hierarchical estimates [9]	bilateral-TV	Gaussian
[72]*	15, 60	2	SIFT+RANSAC	NO	
[87]	20	4	SIFT+RANSAC	Least-square	Gauss($\sigma = 3$)
[5]	10	2	region tracking [6]	motion analysis [46]	motion blur
[3]	20, 40	8	moment-based or Harris + RANSAC	Wiener or MRNSD [63]	B-spline of degree 7
[32] [31]	1 20	2 3	implicit: NLM	NO	
[70] [76]	30	3	implicit: NLM kernel regression	TV bilateral-TV	3×3 average
[24]	9	3	Video-BM3D	zero-padding spectra	3×3 average

Table 2: Multi-image SIFT for registration

	Application	# of images	Registration	Blending method
[8]*	manuscript	Not mention	SIFT + RANSAC	Delaunay triangulation
[60]	registration	30 ultrasound 60 MRI	SIFT + threshold + least-square for affine	B-splines deformation
[82] [45]	Mosaic	200 10	SIFT + RANSAC	weighted average
[50]	stitching	6	SIFT + RANSAC	weighted average
[91]	head tracking	1020	SIFT + RANSAC	NA (track 3D motion)

4.1 Single Image Noise Estimation

Most noise estimation methods have in common that the noise standard deviation is computed by measuring the derivative or equivalently the wavelet coefficient values of the image. As we mentioned, Donoho et al. [30] proposed to estimate the noise standard deviation as the median of absolute values of wavelet coefficients at the finest scale. Instead of the median, many authors [12, 47] prefer to use a robust median.

Olsen [67] and posteriorly Rank et al. [71] proposed to compute the noise standard deviation by taking a robust estimate on the histogram of sample variances of patches in the derivative image. In order to minimize the effect of edges small windows were preferred, with 3×3 or 5×5 pixels. The sample variance of small patches or the point-wise derivatives provide a non robust measure and require a considerable number of samples with few outliers to guarantee the correct selection of the standard deviation. We observed that the opposite point of view, that is, the use of larger windows 15×15 pixels to 21×21 pixels permits a more robust estimation. However, since larger windows may contain more edges a much smaller percentile will be preferred to the median, in practice the 1% or the 0.5%.

Noise in real photograph images is signal dependent. In order to adapt the noise estimation strategies, the gray level image histogram will be divided adaptively into a fixed number of bins having all the same number of samples. This is preferable to classical approaches where the gray range is divided into equal intervals. Such a uniform division can cause many bins to be almost empty.

To evaluate if a signal dependent noise can be estimated from a single image, 110 images were taken with a Nikon D80, with ISO 100 and very good illumination conditions. These are the best conditions we can expect to have a low noise standard deviation. These color images were converted to gray level by averaging the three color values at each pixel. Finally factor 3 sub-sampling was applied by averaging square groups of nine pixels. These operations having divided the noise standard deviation by slightly more than five, these images can be considered as noise free. Finally, a signal dependent noise was added to them, with variance $8 + 2u$ where u was the noiseless grey level.

The uniform and adaptive divisions of the grey level range in a fixed number of 15 bins were compared, and several noise estimation methods were applied to estimate the noise standard deviation inside each bin. The performance of all methods are compared in Table 3 showing the average and standard deviation of the errors between the estimated and original noise curves. The best estimate is obtained by applying the proposed strategy using the variance of large patches rather than small ones or point derivatives. These measurements also confirm that the division of the grey level range into bins with fixed cardinality is preferable to the fixed length interval division. This experiment confirms that a signal dependent noise can be estimated with a high accuracy.

Ground Truth? In order to evaluate the performance of such a noise estimation algorithm in real images we need a ground truth to compare with. This ground truth can be obtained for a given camera by taking a sequence of images of the same pattern, after fixing the camera on a pedestal. All camera parameters remain unchanged for all photographs of the sequence, thus avoiding different exposure times or apertures. The temporal average and standard deviation of the whole sequence of images can therefore be computed without any further registration. The use of a piecewise constant image reduces the effect of small vibrations of the camera, see Fig. 2. The noise in each channel is estimated independently. Each color range is divided adaptively into a fixed number of bins taking into account the color channel histogram. Inside each bin a percentile is used to estimate the standard deviation.

Fig. 3 displays the ground truth estimated curves with this strategy, both in RAW and JPEG format for two different ISO settings. The ground truth curves are compared with the ones estimated in the first image of the sequence by the proposed single image noise estimation

	MAD	RMAD	MVPD	MVPD2
$\overline{e}$	1.81	2.87	1.58	0.75
std(e)	1.14	2.59	1.06	0.61

a) Uniform gray division

	MAD	RMAD	MVPD	MVPD2
$\overline{e}$	1.66	1.87	1.36	0.73
std(e)	1.04	1.17	0.90	0.35

b) Adaptive gray division

Table 3: A signal dependent noise with variance $8 + 2u$ is added to 110 noise free images. The uniform and adaptive strategies for dividing the grey level range in a fixed number of 15 bins are compared. For each strategy, the following noise estimation methods in each bin are compared: median of absolute derivatives (MAD), robust median of absolute derivatives (RMAD), median of sample variance of patches 3×3 of the derivative image (MVPD) and 0.005 percentile of sample variance of patches 21×21 of the derivative image (MVPD2). Are displayed the average and standard deviation of the errors between the estimated and original noise curves for the 110 images.

algorithm. For the RAW case, the single image and ground truth estimated curves are nearly identical. Fig. 2 shows a lack of red in the RAW image of the calibration pattern, even if this pattern is actually gray. This effect is corrected by the white balance as observed in the JPEG image.

The ground truth noise curves estimated from the JPEG images do not agree at all with the classical noise model. This is due to the various image range nonlinear transformations applied by the camera hardware during the image formation, which modify the nature and standard deviation of the noise. The ground truth and single image estimated curves in the JPEG case have a similar shape but a different magnitude. The main new feature is that the interpolation and low pass filtering applied to the originally measured values have strongly altered the high frequency components of the noise. Thus, **the noise statistics are no longer computable from a local patch of the image. The estimation of such a noise curve can only be accomplished by computing the temporal variance in a sequence of images of the same scene.**

4.2 Multi-Image Noise Estimation

A temporal average requires the images of the sequence to be perfectly registered. Yet, this registration rises a serious technical objection: 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 in any of the following situations:

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

The computation of an homography between a pair of images needs the accurate correspondence of at least four points in each image. Finding key points in images and matching them is a fundamental step for many computer vision and image processing applications. One of the most robust is the Scale Invariant Feature Transform (SIFT) [53], which we will use. Other

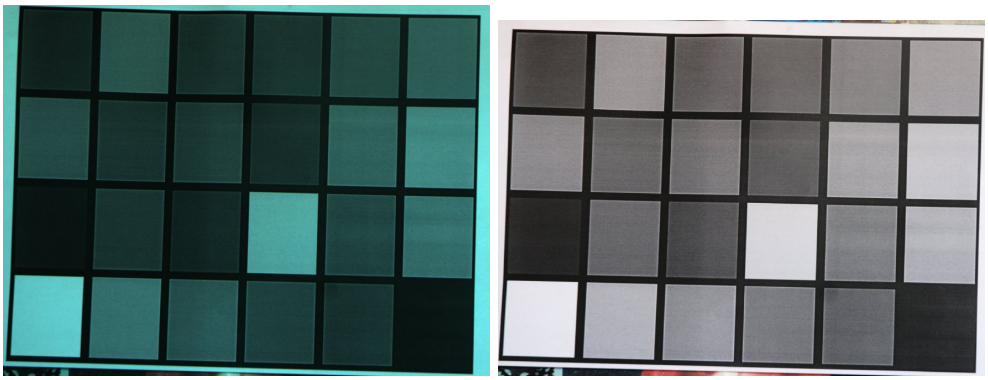


Figure 2: Calibration pattern used for noise ground truth estimation. Left: raw image. Right: JPEG image. Even if the calibration pattern is nearly gray the raw image looks blue because the red is less present. This effect is corrected by the white balance applied by the camera image chain leading to the jpeg image.

possible methods allowing for large baselines are [57, 58, 66, 75, 62, 61], but we are here using images taken with only slight changes of view point.

Because wrong matches occur in the SIFT method, an accurate estimate of the dominant homography will require the elimination of outliers. The standard method to eliminate outliers is RANSAC (RANdom SAmple Consensus) [38]. However, it is efficient only when outliers are a small portion of the whole matching set. For this reason several variants have been proposed to improve the performance of outlier elimination, the principal being [77, 90, 79, 64, 59]. The main difference between our approach and the classic outlier elimination is the fact that we dispose of a whole sequence of images and not just of a pair. Instead of choosing a more elaborate version than RANSAC, we preferred to exploit the sequence redundancy in order to improve the registration stage.

The goal is to estimate a dominant homography for the whole set of images, which are typically a few dozens. Only matches which are common to the whole sequence must be kept. In other terms, the keypoints of the first image are kept only if they are matched with another keypoint in any other image of the sequence. This constraint eliminates most of the outliers (see Algorithm 1). In order to apply such a strategy, we assume that the images overlap considerably. Recall that the purpose is not to make a mosaic or a panorama, but to estimate the noise curve and eventually to denoise the sequence.

A temporal average and standard deviation is computed for the registered sequence. The average values are used to build a histogram and to divide the grey level range adaptively. Inside each bin, the median value of the corresponding standard deviations is taken.

Algorithm 1: Hybrid Accumulation After Registration Algorithm

Input Initial set of images $I_0, I_1, \cdots, I_n$, obtained from a burst

SIFT
Apply the SIFT algorithm between to each pair (I_0, I_j), $j = 1, \cdots, n$. Call S_j the set of matches.
Retain from S_j only the matches for which the matching key point in I_0 has a match in all other images.

RANSAC
Set number of agreed points, m, to 0.
while the number of trials does not exceed N **do**
  Pick up 4 random points from S_0
  for (each $j > 0$) **do**
    Compute the homography using these 4 points and the corresponding ones in S_j
    Add to m the number of points in S_j which agree with this homography up to the precision p.
  end for
  If $m > maxim$, then $maxim = m$ and save the set of agreed points in the whole sequence
end while
Compute for each pair, the homography H_j with the selected points.

FUSION
Apply the homography H_j to each image obtaining $\bar{I}_j$, $j = 1, \cdots, n$.
Average the transformed images obtaining the mean $\mu(x, y)$. Compute also $\sigma(x, y)$, the temporal standard deviation.
Estimate the noise curve using $\sigma(x, y)$, getting $\sigma_n(u)$ the standard deviation associated to each color u.
Obtain the final estimate:

$$(1 - w(\mu, \sigma))\mu(x, y) + w(\mu, \sigma)\, NL(I_0)(x, y),$$

where NL is the NL-means algorithm (Buades et al. [15]) and the function $w(\nu, \sigma)$ is defined by

$$w(\nu, \sigma) = \begin{cases} 0 & \text{if } \sigma < 1.5\sigma_n(\mu) \\ \frac{\sigma - 1.5\sigma_n(\mu)}{1.5\sigma_n(\mu)} & \text{if } 1.5\sigma_n(\mu) < \sigma < 3\sigma_n(\mu) \\ 1 & \text{if } \sigma > 3\sigma_n(\mu) \end{cases}$$

Fig. 4 displays three frames from an image sequence with a rotating pattern and a fixed pedestal. The noise curves estimated from the first image with the single image algorithm and those from the registered and averaged sequence are displayed in the same figure. The estimated curves in the raw image coincide if either of both strategies is applied. However, as previously observed these are quite different when we take into account the JPEG image.

Images taken with indoor lights often show fast variations of the contrast and brightness, like those in Fig. 5. This brightness must be rendered consistent through all the images, so that the standard deviation along time is due to the noise essentially and not to the changes of lights. For this reason, a joint histogram equalization must conservatively be applied before the noise estimation chain. The Midway equalization method proposed in [28, 27] is the ideal tool to do so, since it forces all images to adopt a joint *midway* histogram which is indeed a kind of barycenter of the histograms of all images in the burst. Fig. 5 illustrates the noise estimation after and before color equalization.

5 Average after Registration Denoising

The core idea of the average after registration (AAR) denoising method is that the various values at a cross-registered pixels 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.

There is a strong argument in favor of denoising by simple averaging of the registered samples instead of block-matching strategies. If a fine non-periodic texture is present in an image, it is 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. Now, state of the art denoising methods (e.g. BM3D) are based on nonlocal block matching, which is at risk to confound the repeated noise-like textures with real noise. A registration process which is far more global than block matching, using strong features elsewhere in the image, should permit a safer denoising by accumulation, provided the registration is sub-pixel accurate and the number of images sufficient.

A simple test illustrates this superior noise reduction and texture preservation on fine non periodic textures. A image was randomly translated by non integer shifts, and signal dependent noise was added to yield an image sequence of sixteen noisy images. Figure 6 shows the first image of the sequence and its denoised version obtained by accumulation after registration (AAR). The theoretical noise reduction factor with sixteen images is four. This factor is indeed reached by the accumulation process. Table 4 displays the mean square error between the original image and the denoised one by the different methods. Block based algorithms such as NLmeans [15] and BM3D [23], have a considerably larger error, even if their noise reduction could be theoretically superior due to their two dimensional averaging support. But fine details are lost in the local comparison of small image blocks.

	Barbara	Couple	Hill
noisy	11.30	11.22	10.27
NLM	4.52	3.73	4.50
BM3D	4.33	3.39	3.90
AR	**3.55**	**3.03**	**2.73**

Table 4: Mean square error between the original image and the denoised one by the various considered methods applied on the noisy image sequences in Figure 6. The block based algorithms, NLmeans [15] and BM3D [23] have a considerably larger error, even if their noise reduction could be in theory superior, due to their two dimensional averaging support. AAR is close to the theoretical reduction factor four.

As mentioned in the introduction, the registration by using the SIFT algorithm and computing a homography registration is by now a standard approach in the image fusion literature. The main difference of the proposed approach with anterior work is that the mentioned works do not account for registration errors. Yet, in general, the images of a 3D scene are **not** related by a homography, but by an epipolar geometry [43]. Even if the camera is well-calibrated, a 3D point-to-point correspondence is impossible to obtain without computing the depth of the 3D scene. However, as we mentioned, a camera held steadily in the hand theoretically produces images deduced from each other by a homography, the principal image motion being due to slight rotations of the camera. Nonetheless, we should not expect that a simple homography will be perfectly accurate everywhere in each pair, but only on a significant part. A coherent registration will be obtained by retaining only the SIFT matches that are common to the whole burst. Therefore the registration applies a joint RANSAC strategy, as exposed in Algorithm 1. This ensures that the same background objects are used in all images to compute the corresponding homographies.

The main new feature of the algorithm is this: The averaging is applied only at pixels where the observed standard deviation after registration is close to the one predicted by the estimated noise model. Thus, there is no risk whatsoever associated with AAR, because it only averages sets of samples whose variability is noise compatible.

At the other pixels, the conservative strategy is to apply a state of the art video denoising algorithm such as the spatiotemporal NL-means algorithm or BM3D. To obtain a smooth transition between the averaged pixels and the NL-means denoised pixels, a weighting function is used. This function is equal to 0 when the standard deviation of the current pixel is lower than 1.5 times the estimated noise standard deviation, and equal to 1 if it is larger than 3 times the estimated noise standard deviation. The weights are linearly interpolated between 1.5 and 3.

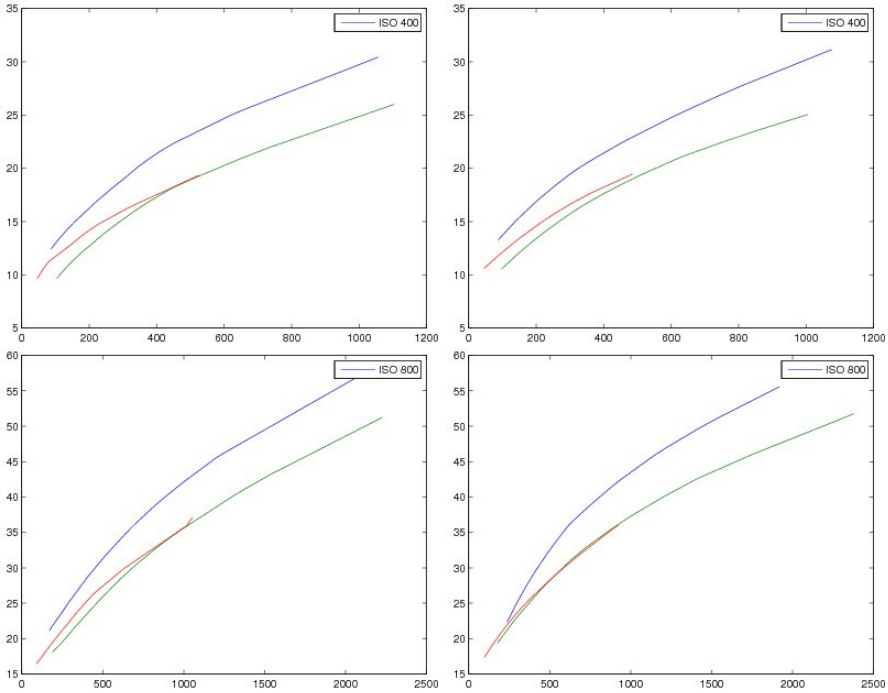
6 Discussion and Experimentation

We will compare the visual quality of restored images from real burst sequences. The focus is on JPEG images, which usually contain non white noise and color artifacts. As we illustrated in the previous sections, the variability of the color at a certain pixel cannot be estimated from a single image but from a whole sequence. We will compare the denoised images by using AAR as well as the classical block based denoising algorithms, NL-means. Fig. 7 shows the results obtained on three different bursts. Each experiment shows in turn: a) three images extracted from the burst, b) the burst average after registration performed at *all* points, followed by a mask of the image regions in which the temporal standard deviation is significantly larger than the standard deviation predicted by the noise estimate. At all of these points a block based denoising estimate is used instead of the temporal mean. The final combined image, obtained by an hybridization of the average registration and NL-Means or BM3D, is the right image in each second row.

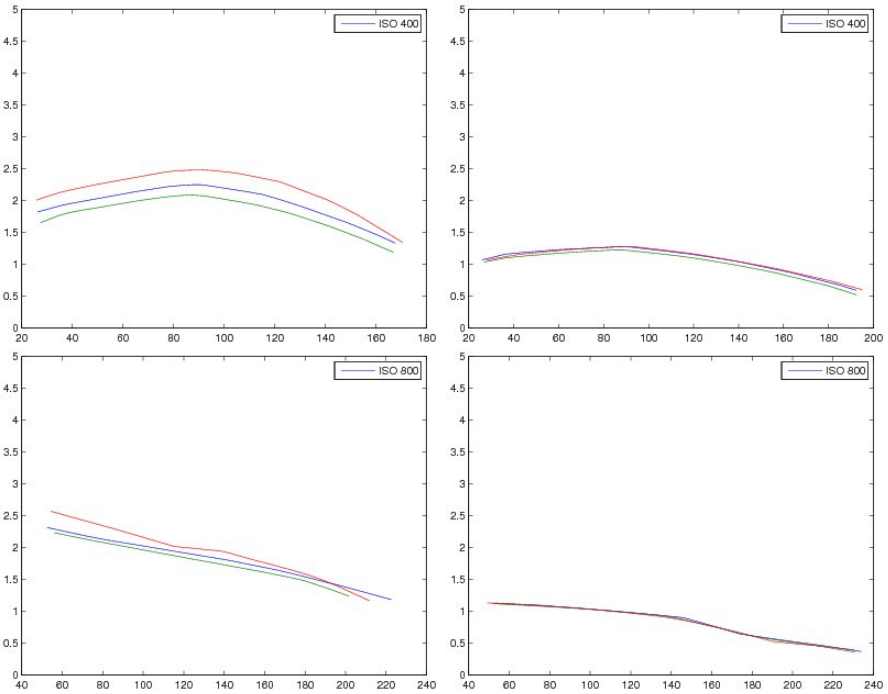
The first experimental data was provided by the company DxO Labs. It captures a rotating pattern with a fixed pedestal. In this case, the dominant homography is a rotation of the main circular pattern, which contains more SIFT points than the pedestal region. Since the proposed algorithm only finds a dominant homography, which is the rotation of the pattern, the simple average fails to denoise the region of the fixed pedestals and of the uniform background. As shown in the white parts of the mask, these regions are detected because they have an excessive temporal standard deviation. They are therefore treated by NL-means or BM3D in the final hybrid result. The whole pattern itself is restored by pure average after registration.

The second burst consists of two books, a newspaper and a moving mouse. Since the dominant homography is computed on still parts, the books and the background, the moving mouse is totally blurred by the averaging after registration, while the rest of the scene is correctly fused. As a consequence, AAR uses the average everywhere, except the part swept by the mouse.

The last burst is a sequence of photographs with short exposure time of a large painting taken in Musée d'Orsay, *Martyrs chrétiens entrant l'amphithéâtre* by Léon Bénouville. Making good photographs of paintings in the dim light of most museums is a good direct application for the proposed algorithm, since the images of the painting are related by a homography even with large changes of view point, the painting being flat. As a result, the average is everywhere favored by the hybrid scheme. Details on the restored images and comparison with BM3D are shown in Fig. 8-10. Dim light images are displayed after their color values have been stretched to $[0, 255]$.



a) RAW images



b) JPEG images

Figure 3: Ground truth and single image noise estimates for the RAW and JPEG images of Fig. 2. The estimated curve by the temporal average and standard deviation coincide with the one estimated from the first image by the proposed single image noise estimation algorithm. This is not the case for the JPEG images. The ground truth and single image estimated curves in the JPEG case have a similar shape but a different magnitude. The interpolation and low pass filtering applied to the original measured values have altered the high frequency components of the noise and have correlated its low frequencies. This means that the noise statistics are no longer computable from a local patch of the image. The estimation of a noise curve can only be accomplished by computing the temporal variance in a sequence of images of the same scene.

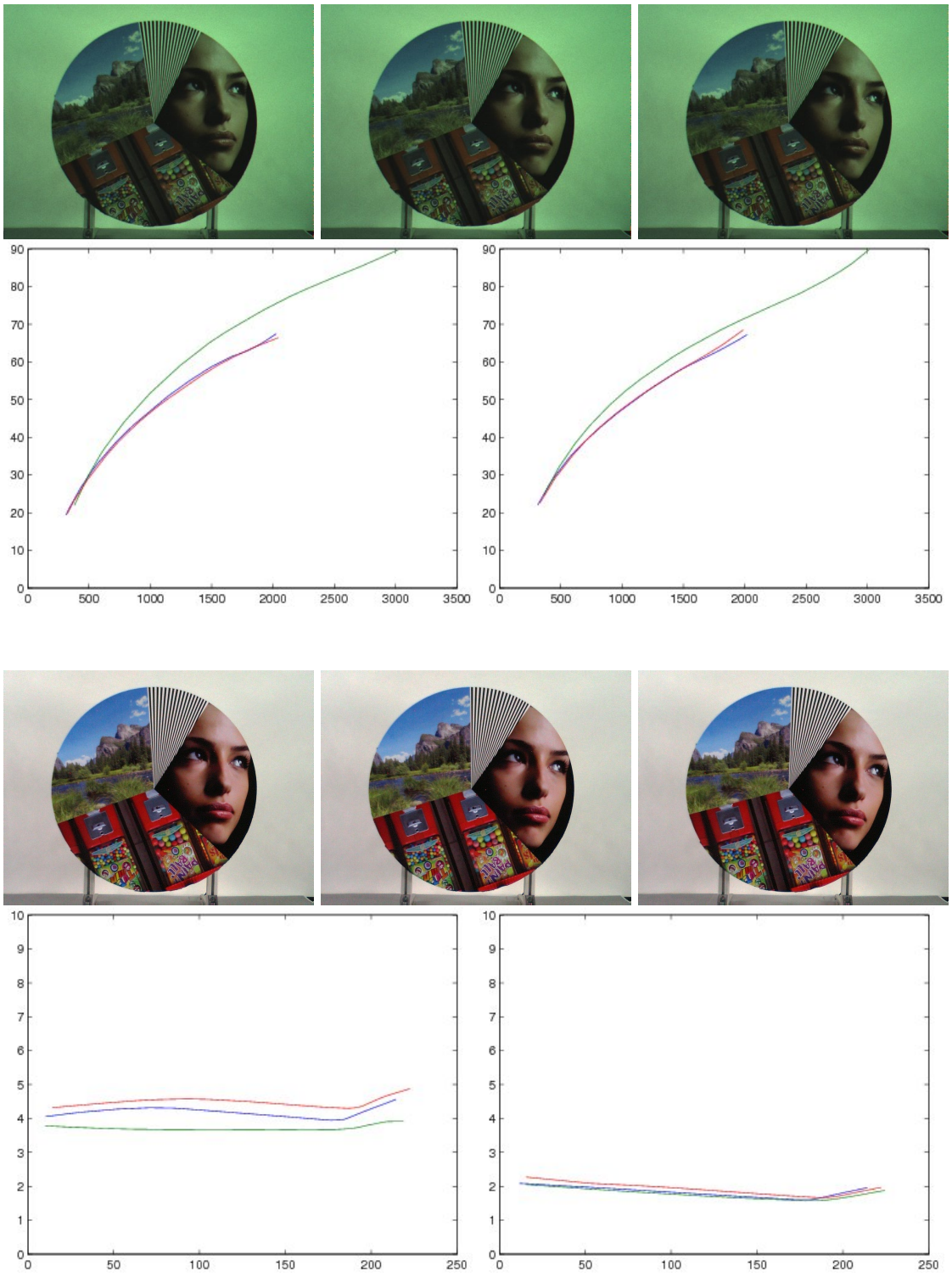


Figure 4: Three frames from an image sequence with a rotating pattern and a fixed pedestal both in RAW (top) and JPG (bottom). The estimated curves in the raw image coincide if either of both strategies is applied. However, as previously observed these are quite different when we take into account the JPEG image

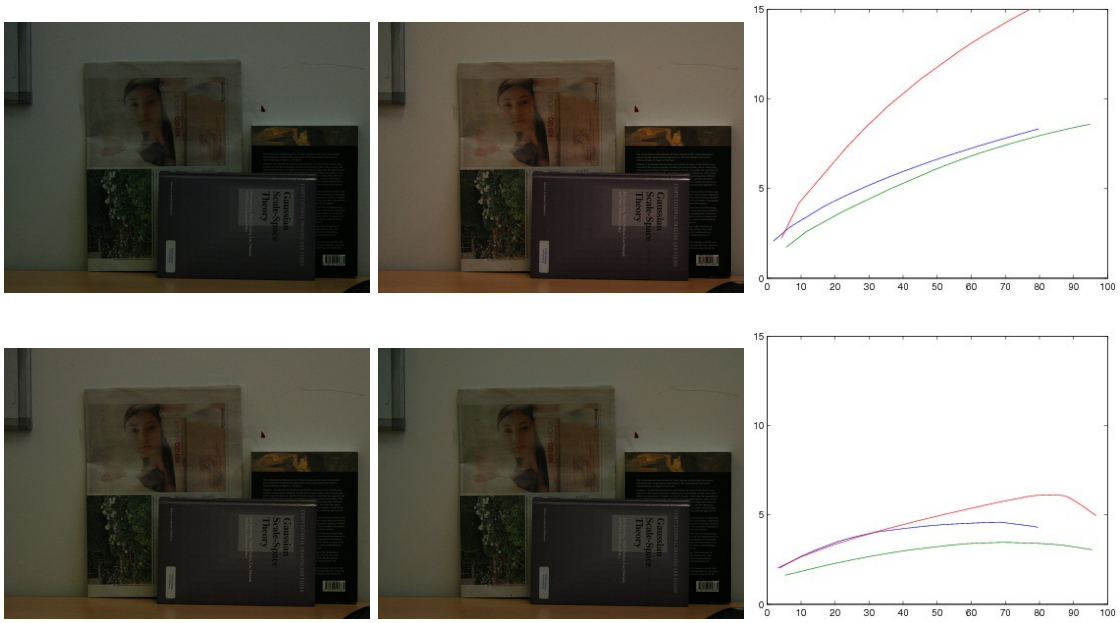


Figure 5: Top: two frames of an image sequence with variations of brightness. Noise curve estimated by temporal average and standard deviation after registration. Bottom: the same two frames of the sequence after a joint histogram equalization [27] and estimated noise curves. The second estimation is correct. The first was not, because of the almost imperceptible lighting conditions.

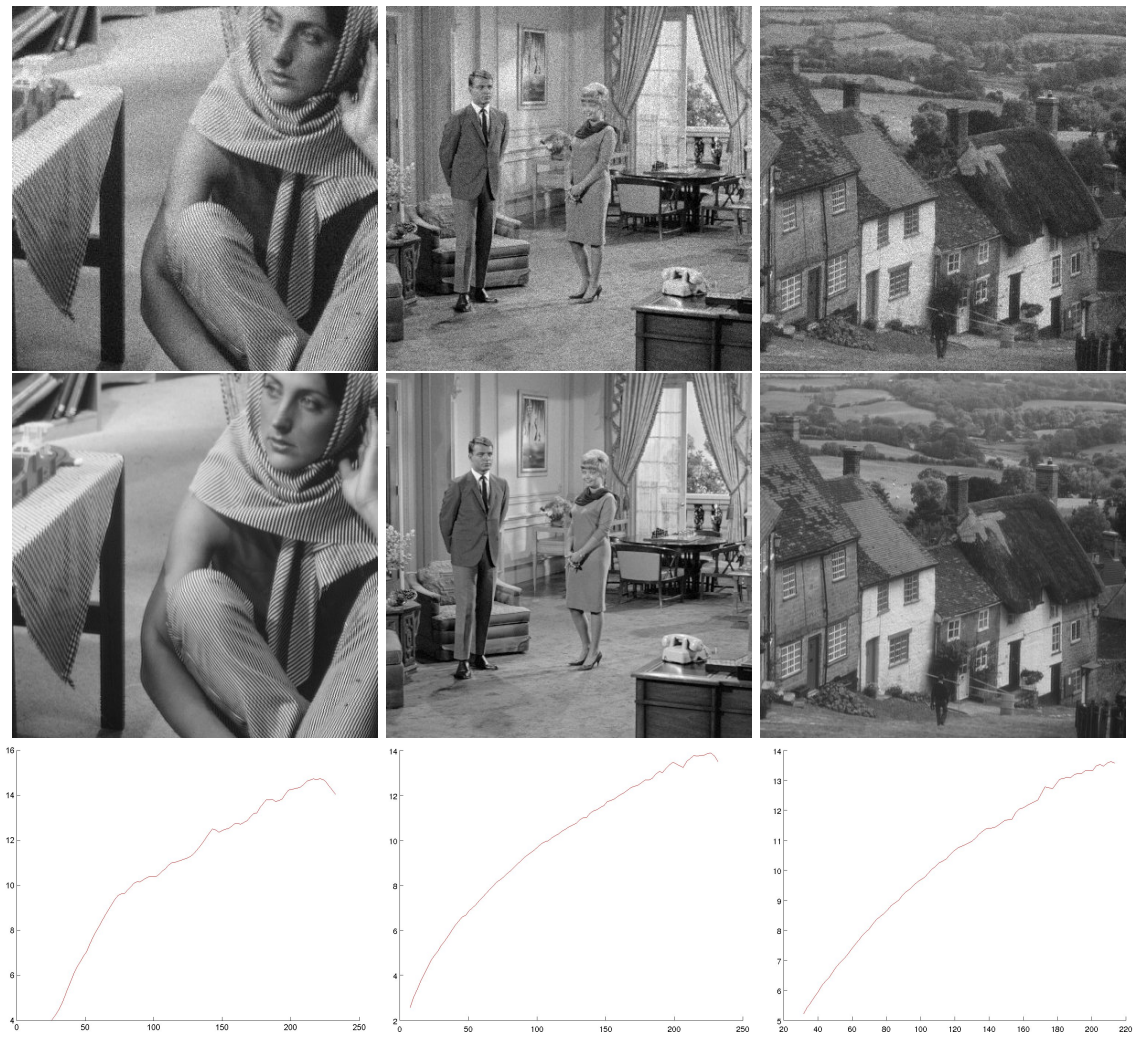


Figure 6: Noise curve. From top to bottom: one of the simulated images by moving the image and adding Poisson noise, denoised by accumulation after registration and the noise curve obtained by the accumulation process using the sixteen images. The standard deviation of the noise (Y-axis) fits to the square root of the intensity (X-axis).

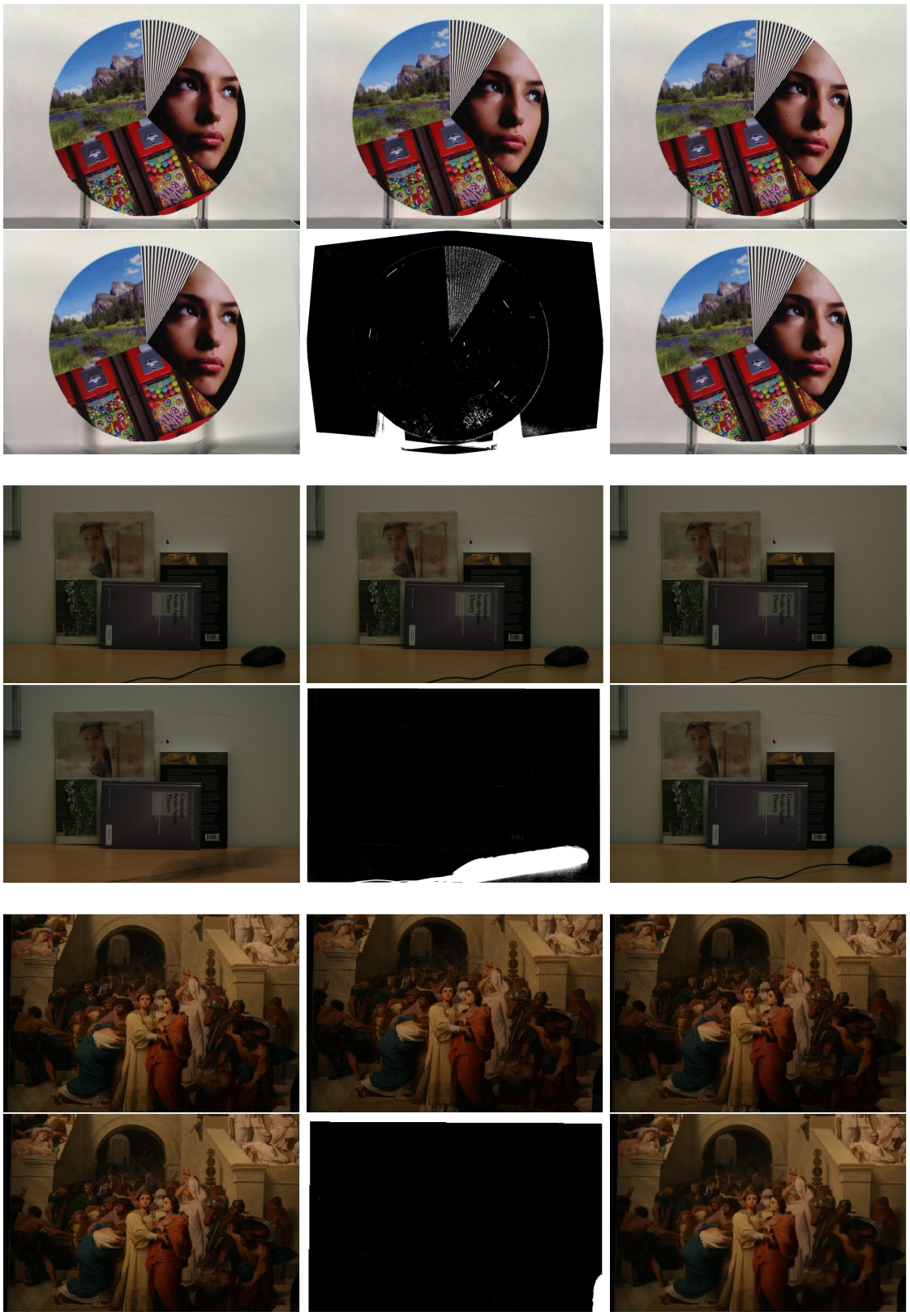


Figure 7: In each double row: three images of a sequence in the first row. In the second row on the left the average after registration, in the middle the mask of points with a too large temporal standard deviation, and on the right the restored image by hybrid method. These experiments illustrate how the hybrid method detects and corrects the potential wrong registrations due to local errors in the global homography.

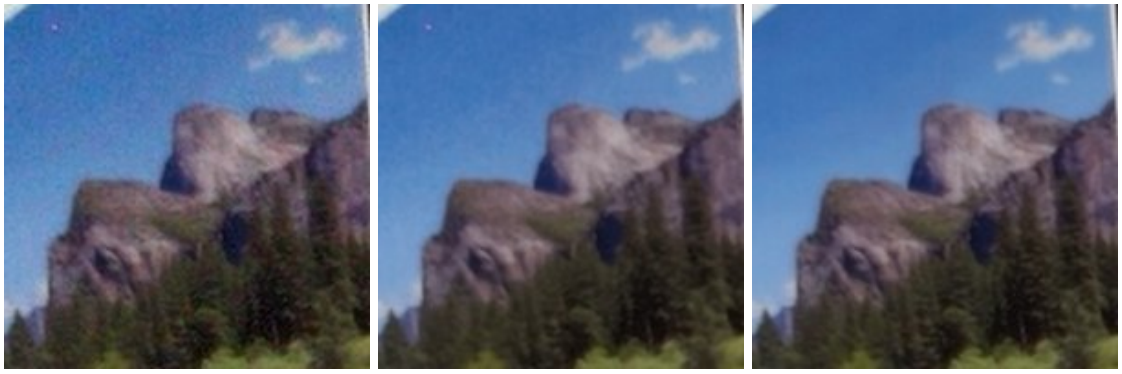


Figure 8: Detail from image in Fig. 7. From left to right: original image, NL-means (BM3D gives a similar result) and hybrid AAR. The images may need to be zoomed in on a screen to compare details and textures. Compare the fine texture details in the trees and the noise in the sky.

Figure 9: Detail from image in Fig. 7. From left to right: original image, BM3D (considered the best state of the art video denoiser) and AAR. The images are displayed after their color values have been stretched to $[0, 255]$. The images may need to be zoomed in on a screen to compare details and textures. Notice how large color spots due to the demosaicking and to JPEG have been corrected in the final result.

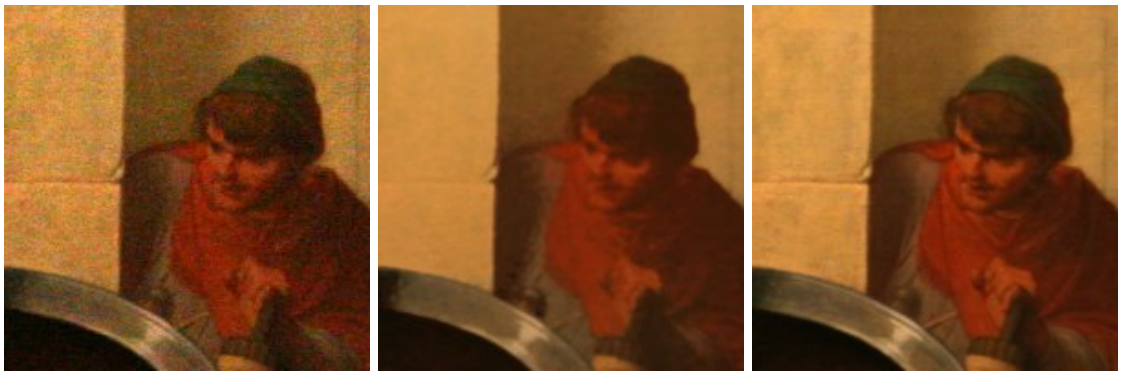


Figure 10: Detail from image in Fig. 7. From left to right: original image, BM3D (considered the best state of the art video denoiser) and AAR. Images are displayed after their color values have been stretched to $[0, 255]$. The images may need to be zoomed in on a screen to compare details and textures. Compare details on the face and on the wall texture.

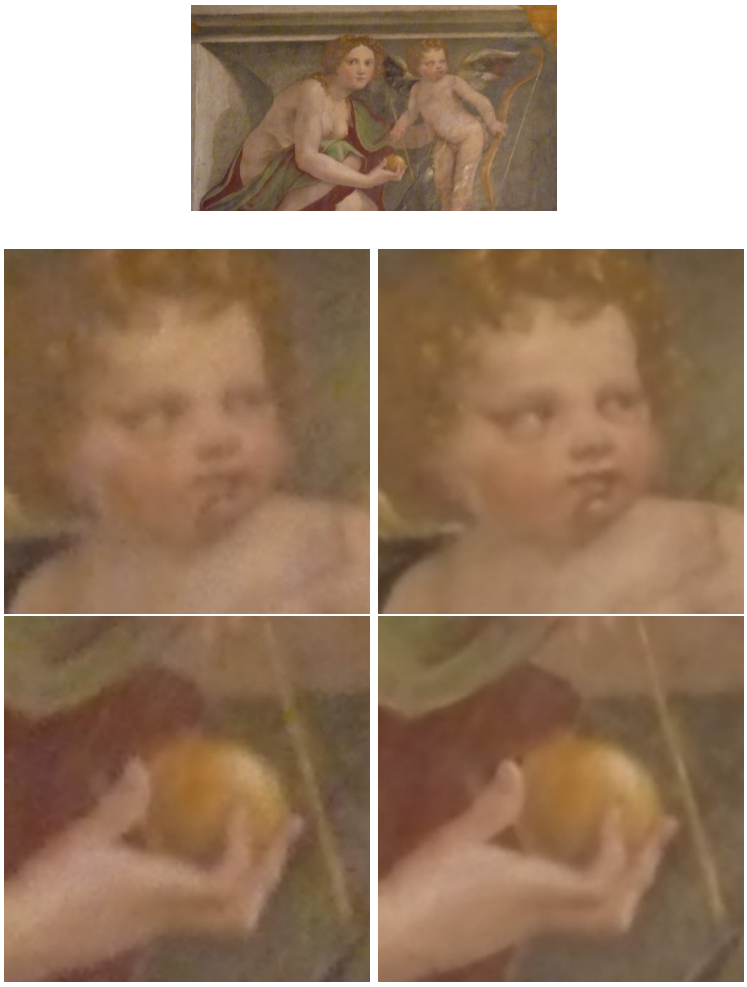


Figure 11: Top: initial image of the burst containing six images. Bottom: details on the initial and hybrid AAR images.

References

[1] Sergey K. Abramov, Vladimir V. Lukin, Benoit Vozel, Kacem Chehdi, and Jaakko T. Astola. Segmentation-based method for blind evaluation of noise variance in images. *Journal of Applied Remote Sensing*, 2(1), 2008.

[2] F. J. Anscomb. The transformation of poisson, binomial and negative-binomial data. *Biometrika*, 35(3):246–254, 1948.

[3] L. Baboulaz and P. L. Dragotti. Exact feature extraction using finite rate of innovation principles with an application to image super-resolution. *IEEE Transactions on Image Processing*, 18(2):281–298, 2009.

[4] Simon Baker and Takeo Kanade. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1167–1183, 2002.

[5] B. Bascle, Andrew Blake, and Andrew Zisserman. Motion deblurring and super-resolution from an image sequence. In *European Conference on Computer Vision*, pages 573–582, 1996.

[6] Bénédicte Bascle and Rachid Deriche. Region tracking through image sequences. In *International Conference on Computer Vision*, pages 302–307, 1995.

[7] S. Battiato, G. Gallo, G. Puglisi, and S. Scellato. Sift features tracking for video stabilization. In *International Conference on Image Analysis and Processing*, pages 825–830, 2007.

[8] R. Baumann and W. B. Seales. Robust registration of manuscript images. In *ACM/IEEE-CS joint conference on Digital libraries*, pages 263–266, New York, NY, USA, 2009. ACM.

[9] James R. Bergen, P. Anandan, Keith J. Hanna, and Rajesh Hingorani. Hierarchical model-based motion estimation. In *European Conference on Computer Vision*, pages 237–252, 1992.

[10] M. Bertalmio and S. Levine. Fusion of bracketing pictures. In *Conference for Visual Media Productio*, 2009.

[11] Christopher M. Bishop, Andrew Blake, and Bhaskara Marthi. Super-resolution enhancement of video. In *International Conference on Artificial Intelligence and Statistics*, 2003.

[12] M.J. Black and G. Sapiro. Edges as outliers: Anisotropic smoothing using local image statistics. *Lecture notes in computer science*, pages 259–270, 1999.

[13] M. Brown, R. Szeliski, and S. Winder. Multi-image matching using multi-scale oriented patches. *Conference on Computer Vision and Pattern Recognition*, pages 510–517, 2005.

[14] Matthew Brown and David G. Lowe. Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, pages 59–73, 2007.

[15] A. Buades, B. Coll, and J. M. Morel. A review of image denoising algorithms, with a new one. *SIAM Multiscale Modeling Simulation*, 4(2):490–530, 2005.

[16] A. Buades, B. Coll, and J.M. Morel. Nonlocal image and movie denoising. *International Journal of Computer Vision*, 76(2):123–139, 2008.

[17] T. 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*, pages 1–15, August 2009.

[18] D. Capel and A. Zisserman. Super-resolution from multiple views using learnt image models. In *Conference on Computer Vision and Pattern Recognition*, volume 2, pages II–627–II–634, 2001.

[19] David Capel and Andrew Zisserman. Automated mosaicing with super-resolution zoom. In *Conference on Computer Vision and Pattern Recognition*, pages 885–891, 1998.

[20] C. Chevalier, G. Roman, and J.N. Niepce. *Guide du photographe*. C. Chevalier, 1854.

[21] K. Dabov, A. Foi, and K. Egiazarian. Video denoising by sparse 3d transform-domain collaborative filtering. In *European Signal Processing Conference*, 2007.

[22] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space. In *International Conference on Image Processing*, 2007.

[23] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3d transform-domain collaborative filtering. *IEEE Transations on Image Processing*, 16(8), 2007.

[24] A. Danielyan, A. Foi, V. Katkovnik, and K. Egiazarian. Image and video super-resolution via spatially adaptive block-matching filtering. In *International Workshop on Local and Non-Local Approximation in Image Processing (LNLA)*, 2008.

[25] Aram Danielyan and A. Foi. Noise variance estimation in nonlocal transform domain. In *International Workshop on Local and Non-Local Approximation in Image Processing (LNLA)*, 2009.

[26] Dmitry Datsenko and Michael Elad. Example-based single document image super-resolution: a global map approach with outlier rejection. In *Multidimensional Systems and Signal Processing*, number 18, pages 103–121, 2007.

[27] J. Delon. Midway image equalization. *Journal of Mathematical Imaging and Vision*, 21(2):119–134, 2004.

[28] J. Delon. Movie and video scale-time equalization application to flicker reduction. *IEEE Transactions on Image Processing*, 15(1):241–248, Jan. 2006.

[29] D. Donoho and J. Johnstone. Ideal spatial adaption via wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

[30] David Donoho and Iain M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90:1200–1224, 1995.

[31] M. Ebrahimi and E.R. Vrscay. Multi-frame super-resolution with no explicit motion estimation. In *International Conference on Image Processing, Computer Vision, and Pattern Recognition*, 2008.

[32] Mehran Ebrahimi and Edward Vrscay. Solving the inverse problem of image zooming using "self-examples". *Image Analysis and Recognition*, pages 117–130, 2007.

[33] E. Eisemann and F. Durand. Flash photography enhancement via intrinsic relighting. *ACM Transactions on Graphics*, 23(3):673–678, 2004.

[34] Michael Elad and Arie Feuer. Super-resolution reconstruction of continuous image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:459–463, 1999.

[35] S. Farsiu, M. Elad, and P. Milanfar. Multiframe demosaicing and super-resolution of color images. *IEEE Transactions on Image Processing*, 15(1):141–159, Jan. 2006.

[36] Sina Farsiu, Dirk Robinson, Michael Elad, and Peyman Milanfar. Fast and robust multiframe super-resolution. *IEEE Transactions on Image ProcessinG*, 13:1327–1344, 2003.

[37] Raanan Fattal, Maneesh Agrawala, and Szymon Rusinkiewicz. Multiscale shape and detail enhancement from multi-light image collections. In *ACM SIGGRAPH*, page 51, 2007.

[38] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24:381–395, 1981.

[39] A. Foi, S. Alenius, V. Katkovnik, and K. Egiazarian. Noise measurement for raw-data of digital imaging sensors by automatic segmentation of non-uniform targets. *IEEE Sensors Journal*, 7(10):1456–1461, 2007.

[40] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian. Practical poissonian-gaussian noise modeling and fitting for single image raw-data. *IEEE Transaction on Image Processing*, 17(10):1737–1754, 2008.

[41] William T. Freeman, Thouis R. Jones, and Egon C Pasztor. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22:56–65, 2002.

[42] A. Goldenshluger and A. Nemirovski. On spatial adaptive estimation of nonparametric regression. *Mathematical Methods of Statistics*, 6:1737–1754, 1997.

[43] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[44] G.E. Healey and R. Kondepudy. Radiometric ccd camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):267–276, 1994.

[45] Marko Heikkilä and Matti Pietikäinen. An image mosaicing module for wide-area surveillance. In *ACM international workshop on Video surveillance & sensor networks*, pages 11–18, New York, NY, USA, 2005. ACM.

[46] Michal Irani and Shmuel Peleg. Motion analysis for image enhancement: Resolution, occlusion, and transparency. *Journal of Visual Communication and Image Representation*, 4:324–335, 1993.

[47] C. Kervrann and J. Boulanger. Local adaptivity to variable smoothness for exemplar-based image regularization and representation. *International Journal of Computer Vision*, 79(1):45–69, 2008.

[48] E. D. Kolaczyk. Wavelet shrinkage estimation of certain poisson intensity signals using corrected thresholds. *Statistica Sinica*, 9:119–135, 1999.

[49] Stamatios Lefkimmiatis, Petros Maragos, and George Papandreou. Bayesian inference on multiscale models for poisson intensity estimation: Application to photo-limited image denoising. *IEEE Transactions on Image Processing*, 18(8):1724–1741, 2009.

[50] Yanfang Li, Yaming Wang, Wenqing Huang, and Zuoli Zhang. Automatic image stitching using sift. In *International Conference on Audio, Language and Image Processing (ICALIP)*, pages 568–571, July 2008.

[51] C. Liu, R. Szeliski, S.B. Kang, C.L. Zitnick, and W.T. Freeman. Automatic estimation and removal of noise from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):299–314, 2008.

[52] Ce Liu, William T. Freeman, Richard Szeliski, and Sing Bing Kang. Noise estimation from a single image. *Conference on Computer Vision and Pattern Recognition*, 1:901–908, 2006.

[53] David G Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[54] H. Lu, Y. Kim, and J.M.M. Anderson. Improved poisson intensity estimation: denoising application using poisson data. *IEEE Transations on Image Processing*, 13(8):1128–1135, Aug. 2004.

[55] M. Makitalo and A. Foi. On the inversion of the anscombe transformation in low-count poisson image denoising. In *International Workshop on Local and Non-Local Approximation in Image Processing (LNLA)*, 2009.

[56] Antonio Marquina and S. Osher. Image super-resolution by tv-regularization and bregman iteration. *Journal of Scientific Computing*, 37(3):367–382, 2008.

[57] Krystian Mikolajczyk and Cordelia Schmid. An affine invariant interest point detector. *European Conference of Computer Vision*, pages 128–142, 2002.

[58] Krystian Mikolajczyk and Cordelia Schmid. Scale and affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.

[59] L. Moisan and B. Stival. A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. *International Journal of Computer Vision*, 57:201–218, 2004.

[60] M. Moradi, P. Abolmaesumi, and P. Mousavi. Deformable registration using scale space keypoints. In *SPIE Medical Imaging 2006: Image Processing*, volume 6144, pages 61442G1–61442G8, 2006.

[61] J.M. Morel and G.Yu. Asift: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2(2):438–469, 2009.

[62] P. Musé, F. Sur, F. Cao, Y. Gousseau, and J.-M. Morel. An *A Contraio* decision method for shape element recognition. *International Journal of Computer Vision*, 69(3):295–315, 2006.

[63] James Nagy and Zdenek Strakos. Enforcing nonnegativity in image reconstruction algorithms. In *SPIE Mathematical Modeling Estimation and Imaging*, pages 182–190, 2000.

[64] D. Nistér. Preemptive ransac for live structure and motion estimation. *Machine Vision and Applications*, 16(5):321–329, 2005.

[65] Robert D. Nowak and Richard G. Baraniuk. Wavelet-domain filtering for photon imaging systems. *IEEE Transactions on Image Processing*, 8(5):666–678, 1997.

[66] Chum O., Urban M., Matas J., and Pajdla T. Robust wide baseline stereo from maximally stable extremal regions. *British Machine Vision Conference*, pages 384–396, 2002.

[67] S. I. Olsen. Estimation of noise in images: an evaluation. *CVGIP: Graph. Models Image Process*, 55(4):319–323, 1993.

[68] N. N. Ponomarenko, V. V. Lukin, S. K. Abramov, K. O. Egiazarian, and J. T. Astola. Blind evaluation of additive noise variance in textured images by nonlinear processing of block dct coefficients. In *Image Processing: Algorithms and Systems II*, volume 5014 of SPIE Proceedings, pages 178–189, 2003.

[69] N. N. Ponomarenko, V. V. Lukin, M. S. Zriakhov, A. Kaarna, and J. T. Astola. An automatic approach to lossy compression of aviris images. *IEEE International Geoscience and Remote Sensing Symposium*, 2007.

[70] M. Protter, M. Elad, H. Takeda, and P. Milanfar. Generalizing the non-local-means to super-resolution reconstruction. *IEEE Transactions on Image Processing*, 18(1):36–51, 2009.

[71] K. Rank, M. Lendl, and R. Unbehauen. Estimation of image noise variance. In *IEEE Proceedings- Vision, Image and Signal Processing*, volume 146, pages 80–84, 1999.

[72] Yeol-Min Seong and HyunWook Park. Superresolution technique for planar objects based on an isoplane transformation. *Optical Engineering*, 47, 2008.

[73] E. Shechtman, Yaron Caspi, and Michal Irani. Increasing space-time resolution in video. In *European Conference on Computer Vision*, pages 753–768, 2002.

[74] Eli Shechtman, Yaron Caspi, and Michal Irani. Space-time super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):531–545, 2005.

[75] F. Sur, F. Cao, P. Musé, and Y. Gousseau. Unsupervised thresholds for shape matchings. *International Conference on Image Precessing*, 2, 2003.

[76] H. Takeda, P. Milanfar, M. Protter, and M. Elad. Super-resolution without explicit subpixel motion estimation. *IEEE Transactions on Image Processing*, 18(9):1958–1975, 2009.

[77] Chi-Keung Tang, Gerard G. Medioni, and Mi-Suen Lee. N-dimensional tensor voting and application to epipolar geometry estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(8):829–844, 2001.

[78] Klaus E. Timmermann and Robert D. Nowak. Multiscale modeling and estimation of poisson processes with application to photon-limited imaging. *IEEE Transactions on Information Theory*, 45(3):846–862, 1999.

[79] P. Torr and A. Zisserman. Mlesac: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78:138–156, 2000.

[80] Patrick Vandewalle, Sabine Süsstrunk, and Martin Vetterli. A frequency domain approach to registration of aliased images with application to super-resolution. *EURASIP Journal on Applied Signal Processing*, 2006:1–14, March 2006.

[81] Jianchao Yang, J. Wright, T. Huang, and Yi Ma. Image super-resolution as sparse representation of raw image patches. In *Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2008.

[82] Zhan-Long Yang and Bao-Long Guo. Image mosaic based on sift. *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pages 1422–1425, 2008.

[83] G. Yu and S. Mallat. Sparse super-resolution with space matching pursuit. In *Proc. of Signal Processing with Adaptive Sparse Structured Representation (SPARS)*, 2009.

[84] G. Yu and S. Mallat. Super-resolution with sparse mixing estimators. Technical report, CMAP, Ecole Polytechnique, 2009.

[85] G. Yu, G. Sapiro, and S. Mallat. Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity. *Arxiv preprint arXiv:1006.3056*, 2010.

[86] Lu Yuan, Jian Sun, Long Quan, and Heung-Yeung SHum. Image deblurring with blurred/noisy image pairs. In *SIGGRAPH*, 2007.

[87] Z. Yuan, P. Yan, and S. Li. Super resolution based on scale invariant feature transform. In *International Conference on Audio, Language and Image Processing*, pages 1550–1554, 2008.

[88] B. Zhang, M. Fadili, and J. L. Starck. Wavelet, ridgelets and curvelets for poisson noise removal. *IEEE Transactions on Image Processing*, 17(7):1093–1108, 2008.

[89] Li Zhang, Sundeep Vaddadi, Hailin Jin, and Shree Nayar. Multiple view image denoising. In *Conference on Computer Vision and Pattern Recognition*, 2009.

[90] Z. Zhang, R. Deriche, O. D. Faugeras, and Q.T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, 78(1-2):87–119, 1995.

[91] Gangqiang Zhao, Ling Chen, Jie Song, and Gencai Chen. Large head movement tracking using sift-based registration. In *International conference on Multimedia*, pages 807–810, New York, NY, USA, 2007. ACM.

[92] W. Zhao and Harpreet S. Sawhney. Is super-resolution with optical flow feasible? In *European Conference on Computer Vision*, pages 599–613, 2002.