Mosaicing a Large Number of Widely Dispersed, Noisy, and Distorted Images: A Bayesian Approach

Frank Dellaert Sebastian Thrun Chuck Thorpe

March 99
CMU-RI-TR-99-34

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

We gratefully acknowledge the support of the following sponsoring institutions: the National Science Foundation (CAREER grant number IIS-9876136 and regular grant number IIS-9877033); DARPA-ATO via TACOM (contract number DAAE07-98-C-L032); DARPA-ISO via Rome Labs (contract number F30602-98-2-0137); the Federal Highway Administration via the program "Sensor-Friendly Roadways and Vehicles"; NHTSA via the program "Run-Off-Road Collision Countermeasures"; PennDOT via the program "Bus Side Collision Countermeasures"; SAIC corporation; KIRIN brewery company; SONY corporation; and the Robotics Institute. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing official policies or endorsements, either expressed or implied, of the United States Government or any of the sponsoring institutions.

Abstract

In this paper we extend existing mosaicing algorithms to deal with image sequences that are captured over a wide spatial area, exhibit large geometric and photometric distortions, and contain significant additive noise and other contaminations, such as light reflections. The paper focuses on three main contributions. (1) We extend the camera model used for mosaicing to deal, not only with geometric lens distortion, but with vignetting and permanent occluders as well. (2) We introduce a novel method for global image alignment based on a technique from the robotics literature, together with a novel optimization strategy, the folding algorithm, to guarantee global convergence. (3) We utilize techniques developed in the super-resolution literature for restoration of the final image mosaic from the contaminated input images. The estimation of camera parameters, the global alignment and the final mosaic are all derived within a unifying Bayesian framework, starting with the single objective of obtaining the maximum a posteriori estimate of the final mosaic, given the input images. Our approach is illustrated with results on a complex and challenging image sequence obtained from a state of the art robotics application.

1 Introduction

In some applications, the need arises to build large-scale image mosaics from hundreds of degraded images. For example, a project we are involved with concerns a robotic tour-guide, Minerva, which needs to reliably navigate in a large museum environment. Other considerations motivated the use of a computer-vision based solution, in which the robot navigates by means of a visual map of the ceiling. This involves first building such a map from images collected over time, a non-trivial task by any means. To illustrate this, we show 6 images in Fig.1 from a collection of 250, taken in the National Museum of American History. In this paper we address the problem of building large-scale mosaic from large sets of images like these.

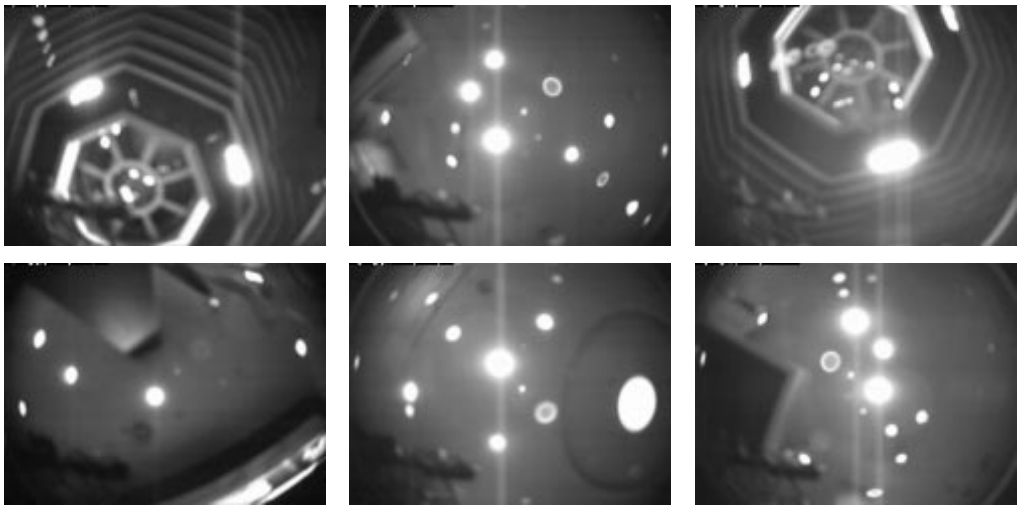

Figure 1: Some degraded images of a ceiling in one of the Smithsonians, taken by the robotic tour-guide Minerva.

Existing mosaicing algorithms would fail on data sets like the one in Fig.1 for three reasons:

Lens Distorted Images The lens used to capture the images in Fig.1 introduces both geometric distortions (such as radial distortion) and photometric distortion (such as vignetting [6]) Although radial distortion has been addressed in the context of mosaicing in [10], in this paper we extend the image formation model to deal with vignetting and permanent occluders.

Large Number of Widely Dispersed Images The images in Fig.1 belong to

a set of 250 images, capturing an area of the ceiling that spans 40m by 60m. Although finding a good local alignment between the images is easy, the problem of globally aligning images that are so widely dispersed is not easily solved. Global alignment schemes [13, 11, 3] have been proposed, but these pose the problem as a single global optimization. We found that these approaches do not work when the images cover a very large area, as they fall into local minima. We propose a local-to-global algorithm, derived from a global localization technique in the robotics literature [8], which does not suffer from this problem.

Noisy Images Almost no mosaicing algorithms explicitly deal with noise issues. Combining multiple images into one mosaic is motivated mostly from esthetic principles, and is concerned with 'blending' the images such that the seams between them become invisible [2, 9, 14]. In reality, images such as those in Fig.1 suffer from a number of contaminations: in this image sequence the focus was slightly off, causing blur. The camera is often aimed straight at ceiling-mounted lights which introduces 'washing-out', as well as light reflections that move over the image. Finally, the images themselves contain significant additive noise. There is a large literature on image-reconstruction from multiple degraded images in the super-resolution (SR) literature [7, 12, 5, 4]. In this paper, we use techniques developed in the SR literature for restoration of the final image mosaic from the contaminated input images.

Finally, the estimation of camera parameters, global alignment and final mosaic are all derived within a unifying Bayesian framework, starting with the single objective of obtaining the *maximum a posteriori* (MAP) estimate of the final mosaic given the input images.

The remainder of this paper is structured as follows: in the next section, Section 2, we explain the image formation model we use, and then proceed to derive the three sub-problems of mosaicing (camera, poses, mosaic) within a single Bayesian framework. In Section 3, the solution to each of these three sub-problems is described in detail. We illustrate each approach with the results we obtained for our test-bed application, the museum ceiling mosaic.

2 A Bayesian Framework

In this section we present a Bayesian framework in which we can address the three difficulties of distorted, spatially dispersed, and noisy images. It extends the Bayesian approach for image-restoration described in [5]. In this approach, the first step is to formulate a model for the problem, described next.

2.1 The Image Formation Model

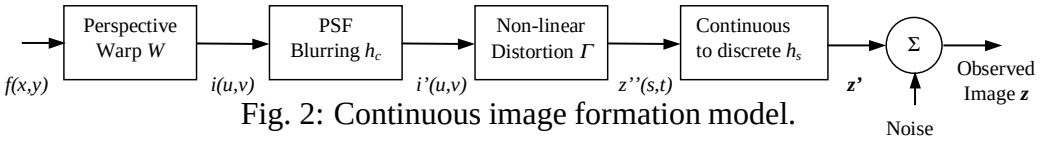

Fig. 2: Continuous image formation model.

The image formation model we use extend the models used in [5, 4] to accommodate the unique problems associated with mosaicing. In particular, we model arbitrary perspective warps, account for distortion and allow for non-Gaussian noise. A non-Gaussian noise model is indispensable in order to deal with contaminations such as light reflections, and provides the theoretical basis for the robust estimator used in Section 3.3. In this paper, we model the scene as a planar surface with an unknown, continuous texture $f(x, y)$. Other models in which a cylindrical, spherical or more general projection models (e.g. as in [9]) could be envisioned, but we will not treat them here.

A block diagram of the continuous image formation model we use is shown in Fig.2. W is the perspective warp between the mosaic texture $f(x, y)$ and the image $i(u, v)$. After a point-spread function (PSF) blurring operation with kernel h_c in the ideal image plane, a non-linear function Γ is applied to model lens distortion effects. The sampling and area averaging operation in the CCD array is represented by a space-variant blurring kernel h_s. Finally, adding (not necessarily Gaussian) noise yields the observed image $\mathbf{z}$.

To formulate a discrete model for purposes of estimation, we employ a discretized *mosaic texture* $\mathbf{f}$. In the case one is not interested in color information, $\mathbf{f}$ is a column vector of grayscale values, one for each mosaic pixel $\mathbf{k}$. The discrete, pixel-level model derived from Fig.2 for pixel p in image i is given by:

$$\mathbf{z}_{ip} = a_i + b_i \gamma_p \sum_k \mathbf{f}_k \, \rho(W^{-1}(\Gamma^{-1}(p)), k) + \mathbf{v}_{ip} \tag{1}$$

where (a_i, b_i) models the contrast-brightness adjustment for image i, γ_p is a space variant vignetting factor (see Fig.3), and ρ is a space-variant blurring kernel that is convolved with the texture $\mathbf{f}$. For each pixel p, this blurring kernel is centered on its *pre-image* $W^{-1}(\Gamma^{-1}(p))$ in texture space, and its shape ρ depends on how the texture is locally oriented with respect to the image.

We can write the above more concisely in matrix notation. Making the dependence on the registration parameters $\mathbf{x}$ and the camera parameters θ explicit, and

grouping all pixels in image i into the column-vector $\mathbf{z}_i$:

$$\mathbf{z}_i = \mathbf{H}_i(\mathbf{x}_i, \theta)\,\mathbf{f} + \mathbf{v}_i \tag{2}$$

where $\mathbf{H}_i$ is a sparse matrix containing the weights of the space-variant resampling operation, parametrized by θ and the registration parameters $\mathbf{x}_i$ for image i. The $b_i\gamma_p$ factors are also incorporated in $\mathbf{H}_i$, and we resort to a notational trick to account for the brightness shift a_i: a constant component with value 1 is appended to $\mathbf{f}$ and the corresponding entry in $\mathbf{H}_i$ is set to a_i.

In the ceiling mosaic example, the registration parameters $\mathbf{x}$ can be identified with the camera poses. The perspective warp W_i for each image is completely determined by the camera pose $\mathbf{x}_i$ and the camera parameters θ. It can be expressed as a 3×3 matrix that transforms homogeneous texture coordinates into homogeneous image coordinates. An analytic expression for this matrix, parameterized by the robot pose, the camera mount, and the intrinsic camera parameters can easily be derived.

2.2 Bayesian Mosaicing

In this section we show how camera parameter estimation, global pose alignment and mosaic reconstruction may all be derived from the MAP criterion:

$$\hat{\mathbf{f}} = \arg\max_{\mathbf{f}} P(\mathbf{f}|\mathbf{z}) \tag{3}$$

where $\mathbf{z} = [\mathbf{z}_1^T \ldots \mathbf{z}_N^T]^T$ is the set of input images. From the image formation model in (2), it is clear that we need to know about the camera parameters θ and the camera poses $\mathbf{x}$ in order to estimate $\mathbf{f}$. Since they are *unknown*, we need to integrate over them:

$$\hat{\mathbf{f}} = \arg\max_{\mathbf{f}} \int_{\mathbf{x},\theta} P(\mathbf{f}|\mathbf{x}, \theta, \mathbf{z})\; P(\mathbf{x}|\theta, \mathbf{z})\; P(\theta|\mathbf{z}) \tag{4}$$

Due to the low dimensionality of $\mathbf{x}$ and θ, they are overdetermined by the information in the images $\mathbf{z}$ and any prior information we have. Thus, their posterior distributions $P(\mathbf{x}|\theta, \mathbf{z})$ and $P(\theta|\mathbf{z})$ are peaked around their MAP estimate, and we can approximate (4) by

$$\hat{\mathbf{f}} \approx \arg\max_{\mathbf{f}}\; P(\mathbf{f}|\hat{\mathbf{x}}, \hat{\theta}, \mathbf{z}) \tag{5}$$

This requires us to solve the additional problems of obtaining the MAP estimate for θ and $\mathbf{x}$:

$$\hat{\theta} = \arg\max_{\theta} \; P(\theta|\mathbf{z}) \tag{6}$$

$$\hat{\mathbf{x}} \approx \arg\max_{\mathbf{x}} P(\mathbf{x}|\hat{\theta}, \mathbf{z}) \tag{7}$$

This then constitutes the overall strategy of our approach:

1. Find the MAP estimate for the camera parameters θ. This is done using a method similar to the one used by Sawhney and Kumar [10] (Section 3.1).

2. Find the MAP estimate for the camera poses $\mathbf{x}$ given $\hat{\theta}$. This is achieved using a variant of a method due to Lu and Milios [8] (Section 3.2).

3. Find the MAP estimate for the mosaic $\mathbf{f}$, given $\hat{\theta}$ and $\hat{\mathbf{x}}$. This is done using the image formation model discussed above, while using an edge-preserving image prior due to Bouman et al. [1] (Section 3.3).

In the next section we describe each of these in turn.

3 Estimation of the Mosaic

3.1 Estimation of the Camera Parameters

We find the MAP estimate for the camera parameters, the four intrinsic parameters $(\alpha_u, \alpha_v, u_0, v_0)$ and a distortion coefficient κ, using the multi-image alignment technique by Sawhney and Kumar [10]. Their reasoning is that, when pixels are projected to the same point in the reference frame, their values should agree. Thus, the error minimized is the sum of the variances for each pixel in the reference frame. We employ only a subset of frames at the beginning of the sequence, for which the camera poses are known.

In addition to distortion, we also model vignetting and permanent occluders. In the museum ceiling example, this was done by calculating the average image from 500 recorded images, which is displayed at the left in Fig.3. The binary mask on the right of Fig.3 is calculated by screening out pixels with substantially lower than average variance, which indicates a permanent artifact is present at that pixel. In a second step, the vignetting coefficients γ_p are calculated for use in in (1): for each pixel, $\gamma_p = I_p/I_0$, where I_p is the average pixel value for pixel p, and I_0 is the average pixel value for the image center.

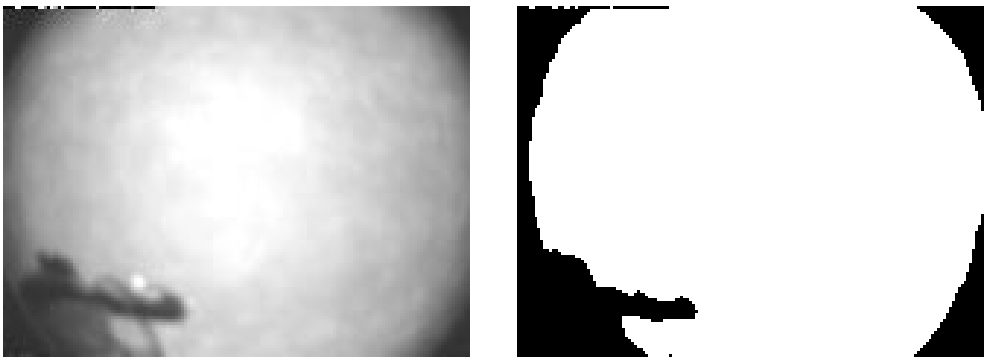

Fig. 3: Left: vignetting effect. Right: the binary mask used to screen out permanent artifacts.

3.2 Estimation of the Global Alignment

A main contribution of this paper is the introduction of a novel algorithm to globally align the input images into a consistent mosaic. In contrast to earlier methods using bundle block adjustment [13, 11], this algorithm is based on a technique from the robotics community due to Lu and Milios [8], used to create a globally consistent world model from laser range scans. In this section, we (a) apply it within the context of mosaicking, (b) provide it with a stronger theoretical foundation, based on the concept of *local views*, and (c) present a novel strategy to guarantee global convergence, the *folding algorithm*.

An important insight is that, while estimating the global mosaic is hard, it is easy to estimate what the mosaic looks like *locally*. For example, for image $\mathbf{z}_i$, this can be done simply by projecting the camera image $\mathbf{z}_i$ into the reference frame, obtaining a *local view* $\hat{\mathbf{g}}_i$. We can then align the nearby camera images $\{\mathbf{z}_j, j \in \mathcal{N}_i\}$ to this local view (where $\mathcal{N}_i$ is the local neighborhood around image i). If we do this for every camera image, we obtain a series of relative pose measurements $\bar{\mathbf{d}}$, after which a MAP estimate for the global poses $\mathbf{x}$ can easily be obtained.

In order to obtain a relative pose measurement $\bar{\mathbf{d}}_{\mathbf{ij}}$ we use standard methods to register image $\mathbf{z}_j$ to the local view $\hat{\mathbf{g}}_i$, i.e., we minimize $-log\, P(\mathbf{z}_j|\mathbf{d}_{ij}, \hat{\mathbf{g}}_i, \theta)$ or:

$$\bar{\mathbf{d}}_{\mathbf{ij}} = \underset{\mathbf{d}_{ij}}{\arg\min} \frac{1}{2} \sum_p \Big[\frac{\mathbf{z}_p - \mathbf{H}_p(\mathbf{d}_{ij}, \theta)\hat{\mathbf{g}}_i}{\sigma_p} \Big]^2 \tag{8}$$

where the index p ranges over the pixels in image j, $\mathbf{H}_p$ denotes row p of $\mathbf{H}$, and we have used the imaging model of (2). For purposes of registration, we

assume a Gaussian noise model with pixel variances σ_p. The above defines a non-linear least-squares problem in $\mathbf{d}_{ij} = [\Delta_x \, \Delta_y \, \Delta_\theta]^T$, and we use the Gauss-Newton method to find $\bar{\mathbf{d}}_{\mathbf{ij}}$. We have found it helpful to precede this with a small random-sampling in the neighborhood of the ini tial estimate, to minimize the risk of getting stuck in a local minimum.

Then, to estimate the global camera poses $\mathbf{x}$, we follow [8] and introduce these relative poses $\bar{\mathbf{d}}$ as additional measurements:

$$\hat{\mathbf{x}} \approx \arg\max_{\mathbf{x}} P(\mathbf{x}|\hat{\theta}, \mathbf{z}, \bar{\mathbf{d}}) \tag{9}$$

$$= \arg\max_{\mathbf{x}} P(\bar{\mathbf{d}}|\mathbf{x}) P(\mathbf{x}) \tag{10}$$

In the above we have assumed that the poses $\mathbf{x}$ are conditionally independent of the images and camera parameters given $\bar{\mathbf{d}}$, and the final expression is obtained using Bayes' rule. If $P(\bar{\mathbf{d}}|\mathbf{x})$ and $P(\mathbf{x})$ are normally distributed, the MAP estimate for $\mathbf{x}$ is obtained by:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left[\frac{1}{2} (\bar{\mathbf{d}} - h(\mathbf{x}))^T \mathbf{C}^{-1} (\bar{\mathbf{d}} - h(\mathbf{x})) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_0)^T \mathbf{P}^{-1} (\mathbf{x} - \mathbf{x}_0) \right] \tag{11}$$

where $\bar{\mathbf{d}} = \{\bar{\mathbf{d}}_k, k = 1..m\}$ is a column vector that groups all the relative pose measurements, $\mathbf{C}$ is the measurement noise covariance matrix associated with $\bar{\mathbf{d}}$, and $\mathbf{x}_0$ and $\mathbf{P}$ are the mean and covariance matrix of the prior on $\mathbf{x}$. We have used the prior only for the purpose of fixing one reference pose to a global coordinate frame, necessary because the likelihood specifies only *relative* pose information. In order to minimize (11), $\mathbf{C}$ can be estimated using Monte-Carlo simulation.

The Folding Algorithm

In order to guarantee convergence, we introduce a novel strategy for gradually incorporating relative pose measurements. Because of the non-linear nature of the problem, and because the quality of the image registration depends substantially on the initial estimate used, it is important that measurements are incorporated in the order that they are to be trusted. It is intuitively clear that frames nearby in the image sequence will have the best initial alignment to start with. Thus, we first measure all relative poses between neighboring frames in the image sequence, then optimize (11), then allow frames two frames away, optimize again, etc. By working our way outwards from every frame to its neighbors, the desired ordering is achieved.

Figure 4: Three snapshots of the global alignment process using the *folding algorithm*, described in the text. Left: initial, bad alignment, using odometry only. Middle: intermediate result. Right: final global alignment.

For the museum ceiling application, the resulting *folding algorithm* is illustrated in Fig.4. Since the robot recorded odometry information during image capture, we have used it to initialize the process. Because of the accumulated odometry error, the resulting mosaic is far from globally aligned, as can be seen in the leftmost panel. An intermediate step in the optimization process, where neighboring frames up to 65 frames away have been integrated, is shown in the middle panel. The final result is shown at right. The optimization at each step was done using Gauss-Newton non-linear optimization. Note that the dimensionality of the problem is quite large: in the above example, there were 250 unknown poses, i.e., 750 unknowns, with about 2000 relative pose measurements between them. Thus, the Jacobian of h in (11) in that case is 6000×750, although it is quite sparse.

3.3 Estimation of the Restored Mosaic

As the final contribution of the paper, we introduce the use of super-resolved image restoration, in conjunction with the appropriate imaging model from Section 2.1, to obtain the final mosaic estimate $\hat{\mathbf{f}}$.

As the image prior $P(\mathbf{f})$ we use an edge preserving image prior, the generalized Gaussian Markov Random Field (GGMRF) prior due to Bouman et al. [1]. In this model, the quadratic energy term associated with a Gaussian noise model is replaced with a softer version, $\rho(\Delta) = |\Delta|^p$:

$$P(\mathbf{f}) \propto exp\left[-\gamma^p\Big(\sum_s a_s|\mathbf{f}_s|^p + \sum_{\{s,r\}\in\mathcal{C}} b_{s,r}|\mathbf{f}_s - \mathbf{f}_r|^p\Big)\right]$$

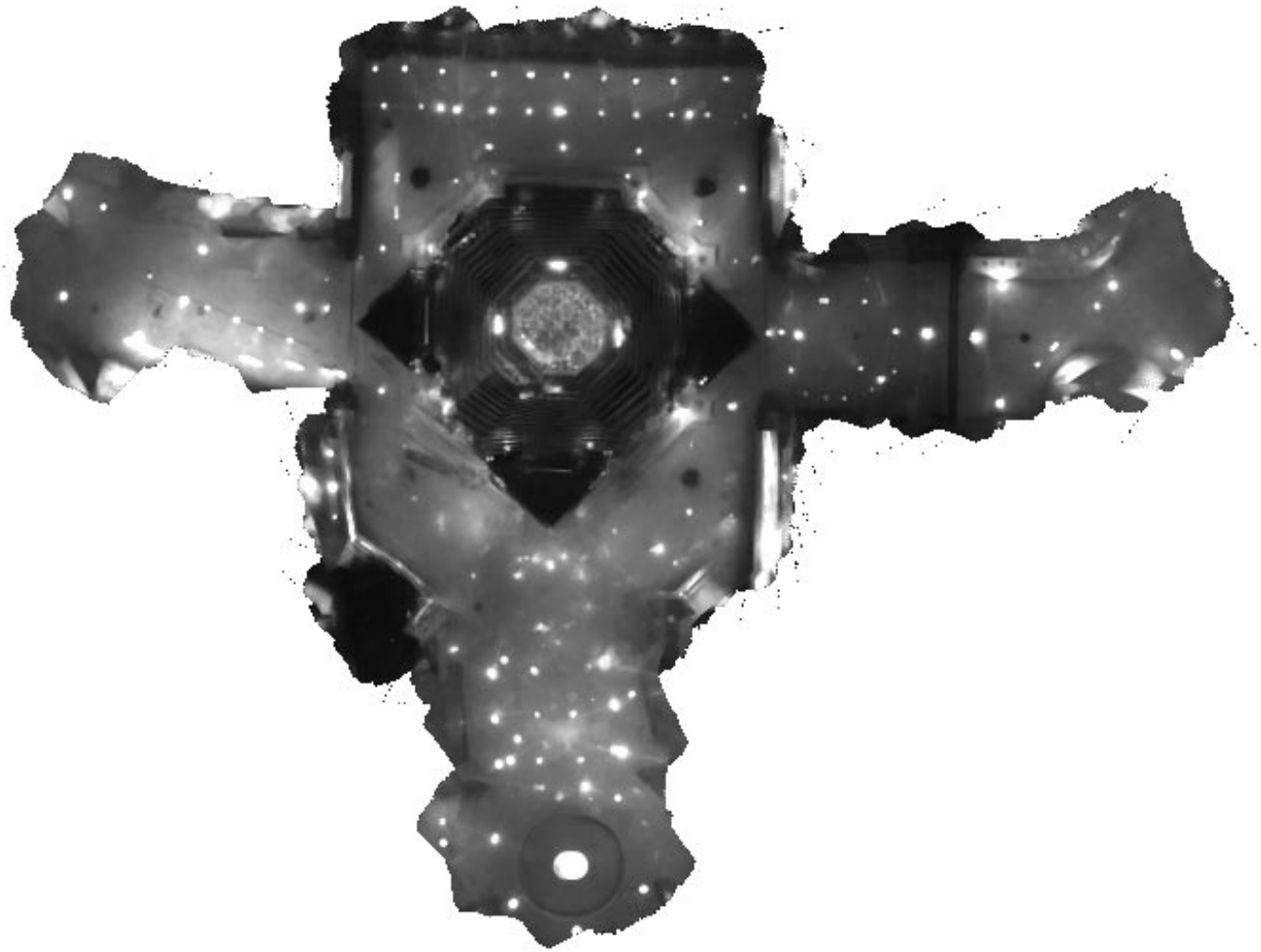

Figure 5: The final museum ceiling mosaic obtained from 250 input images, covering an area of about 60 by 40 meter.

where $1 \leq p \leq 2$, γ is a scale parameter, $\mathcal{C}$ is the collection of neighborhood cliques, and a_s and $b_{s,r}$ are weights parameters. It can be seen that both the sum of absolute differences and the Gaussian MRF are special cases, respectively with $p = 1$ and $p = 2$.

We also replace the Gaussian error model used before for registering the images with a robust M-estimator based on the same $|\Delta|^p$ function. Thus, the likelihood of the mosaic $\mathbf{f}$ given the images $\mathbf{z}$ is written as:

$$P(\mathbf{z}|\mathbf{f}, \hat{\mathbf{x}}, \hat{\theta}) \propto exp\left[-\frac{1}{p}\sum_j \left|\frac{\mathbf{z}_j - \mathbf{H}_j(\hat{\mathbf{x}}, \hat{\theta})\mathbf{f}}{\sigma_j}\right|^p\right] \tag{12}$$

where j indexes over all pixels in all images, $\mathbf{H}_j$ is row j of the measurement matrix $\mathbf{H}$, and σ_j is the noise standard deviation for pixel j. It can easily be seen that for $p = 2$ this is equivalent to a Gaussian model with diagonal noise covariance matrix $\mathbf{R}$, with $\mathbf{R}_{jj} = \sigma_j$.

The final expression we minimize combines the negative log-likelihood from (12) with the prior energy term in (3.3) (with $a_s = 0$):

$$\hat{\mathbf{f}} \quad = \quad \arg\max_{\mathbf{f}} \frac{1}{p}\sum_j \left|\frac{\mathbf{z}_j - \mathbf{H}_j(\hat{\mathbf{x}}, \hat{\theta})\mathbf{f}}{\sigma_j}\right|^p \; + \; \gamma^p \sum_{\{s,r\}\in\mathcal{C}} b_{s,r}|\mathbf{f}_s \; - \; \mathbf{f}_r|^p \tag{13}$$

The resulting mosaic for the museum sequence is shown in Fig.5. As in [1], we use a value of $p = 1.2$, which strikes a compromise between ease of optimization and the extent of edge-preservation. We use a second-order neighborhood system, with $a_s = 0$, $b_{s,r} = 1$ for cardinal neighbors, and $b_{s,r} = 0.7$ for diagonal neighbors. The optimization is done using gradient descent.

The use of the robust estimator yields a significant improvement on the weighted averaging implied by a Gaussian model. In Fig.6, a measured image on the left is compared to the image predicted from the mosaic, using the space-variant imaging model. As can be seen, many of the light-reflections contaminating the actual images are not present in the predicted image. This happens because the robust estimator is more akin to a majority vote (in the $p = 1$ case it will be the median) than to averaging, and light reflections are not consistent between frames.

4 Conclusion

In summary, the following contributions were made:

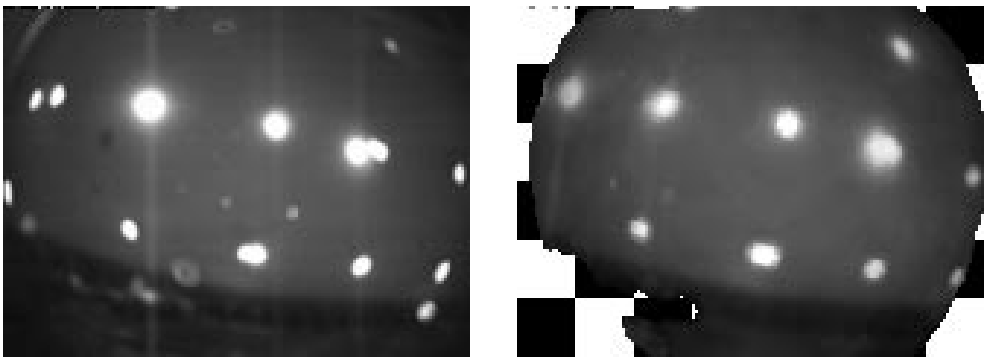

Fig. 6: Reflections in the actual image (left) have been successfully removed in the predicted image (right).

1. We have extended the camera model used for mosaicing to deal not only with geometric lens distortion, but with vignetting and permanent occluders as well.

2. We have introduced a novel method for global image alignment, based on a technique by Lu and Milios [8], which we have formalized using the concept of *local views*. We have introduced a novel optimization strategy, the *folding algorithm*, to guarantee global convergence.

3. To the best of our knowledge, we are the first to deal with the issues of additive noise and other contaminations in the context of mosaicing. In particular, we use an edge-preserving MRF prior to infer information that can never be recovered from the images. Also, through the use of a robust M-estimator we have eliminated most of the light reflections that were present in the image sequence.

4. We have posed the image mosaicing problem within a Bayesian image restoration framework, and shown how estimation of camera parameters, global pose alignment and mosaic reconstruction can be theoretically motivated within a unified Bayesian view.

In addition to these theoretical and technical contributions, we have demonstrated how these techniques were successfully employed to construct a globally aligned mosaic for the museum ceiling sequence. To our knowledge, this is one of the most complex and challenging sequences for which mosaicing has been attempted, given that the robot path twists back onto itself many times, and only

low-quality, contaminated images are available. The resulting mosaic was used with great success in a state of the art mobile robotics application.

Still, there are many warrants for future work. The most important approximation made in this paper concerns the planar model for the scene. Though it is an open question whether a more sophisticated model is needed in the mobile robot application, other applications might certainly benefit from dealing with parallax and 3D structure. Also, in the present paper we have not focussed on achieving super-resolution in the final mosaic: the resolution of input images and the final mosaic was comparable. Nevertheless, in developing the our approach, the resolution was never constrained in any way. In ongoing work we are experimenting with lower as well as higher than image resolution mosaics, in order to explore the promise of true arbitrary resolution mosaicing.

Acknowledgments

We are greatly indebted to Simon Baker for his detailed and useful comments on an earlier version of this paper. We would also like to thank the members of the TE-Lab at CMU for the use of their equipment and generous support. Thanks also to Pete Rander for lending us a wide-angle lens.

References

[1] Charles Bouman and Ken Sauer. A generalized gaussian image model for edge-preserving MAP estimation. *IEEE Trans. Image Processing*, 2(3):296–310, 1993.

[2] P.J. Burt and E.H. Adelson. A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics*, 2(4):217–236, 1983.

[3] J. Davis. Mosaics of scenes with moving objects. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 1998.

[4] M. Elad and A. Feuer. Restoration of single super-resolution image from several blurred, noisy and down-sampled measured images. *IEEE Trans. Image Processing*, 6(12):1646–58, December 1997.

[5] R.C. Hardie, K.J. Barnard, and E.E. Armstrong. Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Trans. Image Processing*, 6(12):1621–1633, December 1997.

[6] B.K.P. Horn. *Robot Vision*. MIT Press, Cambridge, MA, 1986.

[7] M. Irani and S. Peleg. Motion analysis for image enhancement: Resolution, occlusion, and transparency. *J. of Visual Communication and Image Representation*, 4(4):324–335, 1993.

[8] F. Lu and E Milios. Globally consistent range scan alignment for environment mapping. Technical report, Department of Computer Science, York University, April 1997.

[9] S. Peleg and J. Herman. Panoramic mosaics by manifold projection. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 338–343, San Juan, Puerto Rico, June 1997.

[10] H. S. Sawhney and R. Kumar. True multi-image alignment and its application to mosaicing and lens distortion correction. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 450–456, San Juan, Puerto Rico, June 1997.

[11] H.S. Sawhney, Steve Hsu, and R. Kumar. Robust video mosaicing through topology inference and local to global alignment. In *Eur. Conf. on Computer Vision (ECCV)*, 1998.

[12] R. R. Schultz and R. L. Stevenson. Extraction of high-resolution frames from video sequences. *IEEE Trans. Image Processing*, 5(6):996–1011, June 1996.

[13] H.-Y. Shum and R. Szeliski. Construction and refinement of panoramic mosaics with global and local alignment. In *Int. Conf. on Computer Vision (ICCV)*, pages 953–958, Bombay, January 1998.

[14] Y. Xiong and K. Turkowski. Registration, calibration and blending in creating high-quality panoramas. In *IEEE Workshop on Applications of Computer Vision (WACV)*, pages 69–75, 1998.