

# Recording Consistent Multiview Image Data

Markus Kettern<sup>1</sup>, Benjamin Prestele<sup>1</sup>, and Peter Eisert<sup>1,2</sup>

<sup>1</sup> Fraunhofer Institut für Nachrichtentechnik Heinrich Hertz Institut

<sup>2</sup> Humboldt Universität zu Berlin, Institut für Informatik

{markus.kettern | benjamin.prestele | peter.eisert}@hhi.fraunhofer.de

**Abstract.** Recording consistent multiview image data for 3D reconstruction is a challenging task, especially with low cost cameras. We propose a mobile setup of circularly arranged cameras for recording a rotational image sequence around the subject. It provides calibration of the cameras with a novel, multi-functional calibration target as well as software color adaption. The output of the calibrated cameras can be rectified to form a smooth rotational image sequence using the calibration data.

## 1 Introduction

Acquiring consistent multiview image data is an important field of research and has great practical relevance in many image and video-based applications. However, a great deal of preparation is required. A multi-camera mount and precise fixation of the cameras is needed as well as methods for calibration of the cameras. After recording, colors should be adjusted to compensate for differences in the imaging characteristics of the cameras and lighting effects.

We propose a complete and mobile setup for automatic acquisition of multiview image data consisting of a circular array of cameras able to record a sequence of images in which the viewpoint rotates around the recorded subject. The setup has been developed to record human heads but can be used with other subjects as well. It is designed for using low-cost cameras and does not require precise mounting since robust calibration routines are employed which compensate for construction inaccuracies. The colors of the recorded images are adjusted by a software color adaption in order to correct for color differences that cannot be compensated for by adjusting the camera hardware settings.

The obtained calibration data can be used in different ways, especially for 3D reconstruction methods. Here however we propose utilizing it for a direct 2-dimensional rectification of the recorded images, resulting in a smooth image sequence displaying a rotation around the subject. First, the center point of the camera array in 3D space is determined from the calibration data. Then, for each camera a transformation is computed to adjust its viewing direction to point directly towards this center point and is applied to the recorded image.

After a brief overview of related work in the next section, we introduce our hardware setup in section 3. A description of the calibration follows in section 4, where the software-based methods of color adaption and image rectification are treated in section 5. Finally, results and an outlook on future uses for this setup are presented in section 6.

## **2 Related Work**

A standard reference for geometrical camera calibration containing most theoretical aspects and basis for most calibration applications is [6]. An easy way of calibrating an array of synchronized cameras by waving a small light source is proposed in [10] and [1]. Here however, we follow an analysis by synthesis approach which was first presented in [2] and has the advantage of being very robust and precise when initialized properly. Also, it does not require the cameras to be synchronized and can be carried out with one shot per camera, where multiple shots can be used to improve precision of the calibration results.

For color adaption, [7] propose a second order polynomial transform within RGB space. This transform can handle nonlinear color distortions, but its global application does not seem adequate to compensate for the differences in color responses of low cost cameras. A histogram-based approach more similar to our method and used to improve the compression of multiview video sequences is presented in [3]. A description close to our method has been published in [4].

Projective rectification of images taken by a calibrated camera setup has been treated in [5], [8] and ultimately [6].

### 3 The Hardware Setup

The camera setup proposed here consists of 35 low cost cameras mounted in circular manner and spanning about  $130^\circ$  around the subject to be recorded. The diameter of this circle is approximately 2m so that the recording distance is about 1m.

The calibration target is a cylinder big enough to cover the spatial range of possible subject placement. It is displayed in figure 1. The colored stripe sequence in the center (1) allows an image search algorithm to robustly detect the calibration target. Its rotation around the vertical axis is determined from the positions of the colored boxes above and below these central stripes (2). The white boxes (3) serve the model-based calibration methods to precisely measure the model fitting error.



Figure 1: Multi-functional calibration target

### 4 Camera Calibration

For the calibration of the camera array, a model-based analysis by synthesis approach is used which demands a good and robust initial estimation of the calibration target's position and orientation. While initialization is often done manually, we have developed an initialization routine which provides the desired data without human interaction.

## 4.1 Initialization

The initial estimation of the calibration target's position in the image is obtained by inspecting a sparse selection of the image columns for the presence of the colored stripe sequence in the center of the target. For comparison of color sequences, we introduce an RGB ratio vector

$$\mathbf{q}_i = \left[ \frac{r_i}{r_{i+1}} \quad \frac{g_i}{g_{i+1}} \quad \frac{b_i}{b_{i+1}} \quad \frac{r_i}{g_i} \quad \frac{g_i}{b_i} \quad \frac{b_i}{r_i} \right]^T \quad (1)$$

where  $r_i$  denotes the average red-channel value of color  $i$  etc.

This vector has the advantage of being insensitive against global illumination and tint of an image as well as being highly discriminative between different color pairs. Sets of colored windows whose average colors exhibit RGB-ratios similar to the ones of the target color sequence are determined by searching across vertical offsets and different window sizes within each column.

This yields an objective function whose minimum values lead to an initial guess of the position and the stripe size of the color sequence for each image column inspected. The horizontal extent of the calibration target is determined not only by the value of the objective function but also by thresholding the difference in vertical position and stripe size between neighboring columns. This provides a selection process very robust to noise and false detections.

Now, we want to improve the localization of the vertical transitions between the colors as well as the left and right borders of the calibration target. To this end, the image is filtered with a  $[-0.5, 0, 0.5]^T$ -kernel in each color channel to detect horizontal edges between different colors and assign each pixel a probability of lying on a certain color transition. Then, dynamic programming is used to find the horizontal path with maximal transition probabilities and minimal vertical offsets between neighboring pixels.

The left and right borders of the calibration target are detected by fitting a coarse model of a horizontal line through the white boxes on the calibration target to the image. The pathway of this line is extrapolated from the stripe transitions we just determined. Since the model has only two parameters (rotation and tilt) and a range of results for scale and position is known from the initial detections of the color sequence, it can be fit efficiently via a coarse-to-fine search. Figure 2 displays the transitions detected and trimmed by this algorithm as white curves on an image from a low cost camera.

To improve model fitting in 4.2, the image is converted to grayscale and

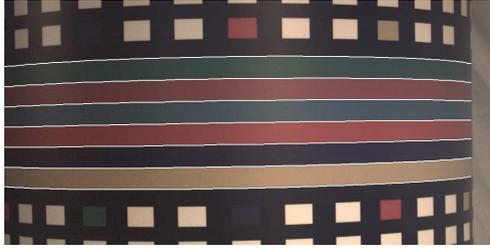


Figure 2: Color transitions as detected by the proposed algorithm

adaptive contrast enhancement is performed in order to highlight the white boxes and blacken out the other image contents. We use a histogram-based sliding-window algorithm which turns the darkest pixels to black by a threshold and stretches the histogram so that all pixels with intensity higher than 25% over the new black level become white. A suitable window size is estimated from the known scale of the calibration target.

The rotation of the calibration target around its vertical axis is determined by exploiting the uniqueness of combinations of the colored boxes visible in the image. The error of assigning each color of the detected central stripes to each box is calculated and the combination minimizing the sum of these assignment costs is sought among the combinations available on the calibration target. Effects of illumination and color response are minimized by sampling the color of each stripe directly above or below the respective box.

## 4.2 Model-based Geometrical Calibration

For multi-view camera calibration, we employ an analysis-by-synthesis approach and thus avoid the need for sub-pixel localization of calibration features. Instead we exploit the entire image for accurate estimation of intrinsic and extrinsic camera parameters [2]. This however requires an approximate initialization to prevent it from converging to false local minima in the model fitting process which arise from aliasing of the white boxes on the calibration target. Utilizing the positions of the colored boxes obtained from the initialization algorithm described in the previous section, a 3D model of the calibration target is fit to the provided grayscale image. This yields coarse extrinsic parameters (orientation and position of the body) and an initial estimation of intrinsic parameters. Furthermore, this fitting is sufficient for the following algorithm to converge to the correct camera parameters.

The geometrical calibration is based on an analysis-by-synthesis framework which matches synthesized images of the calibration target with the images recorded by the cameras. The views of the calibration target are rendered using a simple ray tracing technique. Given a set of intrinsic and extrinsic camera parameters, a corresponding synthetic view is created. From the mismatches between the synthetic image and the real view, parameter updates for intrinsic and extrinsic settings are estimated in a linear hierarchical framework. This optimization is performed iteratively until the error cannot be reduced any further or is sufficiently small.

The matching of real and synthetic view is guided by an optical flow-based optimization. The optical flow constraint relates the spatial image gradients and intensity differences between the images to a dense 2D displacement field. Since this constraint is underdetermined with two unknown displacement components for each equation, additional information is required to solve for the displacement field. Instead of adding smoothness priors we exploit an explicit model of the displacement field being a linearized function of the intrinsic and extrinsic parameter updates. This model contains information about the shape of the calibration target, its position and orientation, as well as the camera settings. Since the model is valid for the entire view, its combination with the optical flow constraint results in a highly overdetermined set of linear equations which can be solved efficiently following a least squares approach. Problems of limited allowed mismatch between the images and linearization errors are addressed by conducting a Gauss-Newton optimization in scale space. Thus, highly accurate results are obtained even for rough initial alignment of the images.

In order to enhance the precision of the calibration and to cover the entire viewing area of the cameras, multiple images for each camera may be used with the calibration target being placed at different locations. For each additional capture, 6 additional unknown extrinsic parameters for the object position and orientation have to be estimated. The intrinsic parameters, however, are constant for all views of a single camera so that they can be robustly estimated jointly over all available views. Since the total number of unknowns can be great for setups with many cameras and many views, the resulting linear system of equations may be very large. Due to the sparse structure of this matrix, however, efficient solution algorithms can be used so that only small sub-matrices need to be inverted and stored enabling optimization of large camera arrays in reasonable time.

## 5 Color Adaption and Image Rectification

When the cameras have been calibrated, the actual image data can be recorded and further processed. To improve visual homogeneity of the data, we propose a simple color adaption algorithm. To receive a smooth rotational image sequence, the images are rectified as described in 5.2.

### 5.1 Color Adaption

Without inference of inter-camera image correspondences, color adaption is applied globally. The adaption is performed between one pair of images and propagated through the camera array starting from a reference camera which the user can set up manually. A cumulative histogram for each color channel is computed in both images. Certain values of each histogram, e.g. at 20, 40, 60 and 80 per cent of maximum image intensity are considered and a smooth gradation function mapping these values from the source histogram to the corresponding values in the target histogram is computed. Figure 3 exemplifies this mapping for one color channel. The mapping function is a smooth gradation curve which does not change the intensity values for black and white but the distribution in between.

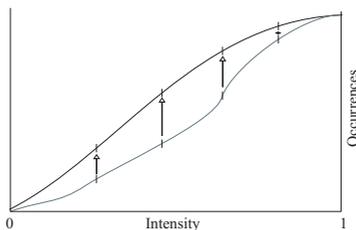


Figure 3: Example of cumulative histograms of source (gray) and target (black) image.

### 5.2 Image Rectification

Camera calibration provides a camera position  $\mathbf{c}_i$ , an intrinsic parameter matrix  $K_i$ , a rotation matrix  $R_i$  and an estimation of radial lens distortion for each camera. These (apart from lens distortion) form the  $3 \times 4$ -projection matrix

$$P_i = K_i R_i [I_3 | -\mathbf{c}_i] \quad (2)$$

where  $I_3$  is the  $3 \times 3$  identity matrix. With  $\mathbf{c}_i$  representing the camera center in the world coordinate frame, the mapping  $\mathbf{x}'$  of any point  $\mathbf{x}$  in the world frame onto the image plane of camera  $i$  is given by

$$\phi_2(\mathbf{x}') = P_i \phi_3(\mathbf{x}) \quad (3)$$

where  $\phi_k$  is the homogeneous representation of a point in  $k$ -dimensional space such that

$$\phi_k(\mathbf{x}) = \begin{bmatrix} w\mathbf{x} \\ w \end{bmatrix} \quad (4)$$

has  $k + 1$  dimensions for a  $k$ -vector  $\mathbf{x}$ .

To rectify the recorded image data, we render a new view for each camera, thereby virtually changing the viewing axis of the camera such that the axes of all cameras precisely meet in one point  $\mathbf{z}$  in the world coordinate frame. The principal axis of camera  $i$  is given by

$$\mathbf{v}_i = \det(M_i) \mathbf{m}_i^3 \quad (5)$$

where  $M_i$  is a  $3 \times 3$ -matrix containing the first three columns of  $P_i$  and  $\mathbf{m}_i^{3T}$  is the third row of  $M_i$ .

To find the center point  $\mathbf{z}$ , we first estimate the plane which is closest to the centers and principal axes of all cameras in the world coordinate frame. This principal plane is fit to the center and one point on the principal axis,  $\mathbf{c}_i + \lambda \mathbf{v}_i$  for each camera where we choose  $\lambda$  to be 1m in regard of the radius of our camera setup. The principal axes of all cameras are projected onto this plane and  $\mathbf{z}$  is found as the point minimizing the squared distances to these projections.

To change the viewing direction of camera  $i$ , we first compute a new rotation matrix

$$R_i^* = [\mathbf{a}_i^x \ \mathbf{a}_i^y \ \mathbf{a}_i^z]^{-1} \quad (6)$$

with

$$\begin{aligned} \mathbf{a}_i^z &= \frac{\mathbf{z} - \mathbf{c}_i}{\|\mathbf{z} - \mathbf{c}_i\|} \\ \mathbf{a}_i^y &= \frac{\mathbf{n} - (\mathbf{n}^T \mathbf{a}_i^z) \mathbf{a}_i^z}{\|\mathbf{n} - (\mathbf{n}^T \mathbf{a}_i^z) \mathbf{a}_i^z\|} \\ \mathbf{a}_i^x &= \frac{\mathbf{a}_i^y \times \mathbf{a}_i^z}{\|\mathbf{a}_i^y \times \mathbf{a}_i^z\|} \end{aligned} \quad (7)$$

where  $\mathbf{n}$  is the normal of the principal plane obtained above. With this rotation matrix, we can get a new projection matrix for each camera by

$$P_i^* = K_i R_i^* [I_3 \mid -\mathbf{c}_i] \quad (8)$$

We transform this projection matrix into the canonical coordinate frame induced by the original camera  $i$  by

$$\hat{P}_i = P_i^* \begin{bmatrix} P_i \\ 0001 \end{bmatrix}^{-1} \quad (9)$$

and finally obtain a  $3 \times 3$ -matrix homography between the projected images of both camera matrices by

$$H_i = A_i - \mathbf{a}_i \mathbf{v}_0^T \quad (10)$$

where

$$\hat{P}_i = [A_i | \mathbf{a}_i] \quad (11)$$

and

$$\mathbf{v}_0 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (12)$$

is the viewing axis of the camera corresponding to the canonical coordinate frame.

Before applying these homographies to the image sequence, we compensate the radial lens distortion obtained in 4.2 in each image.

## 6 Results and Outlook



Figure 4: Samples of a recorded sequence (top) and corresponding images of the rectified and color-adapted sequence (bottom).

We have proposed a complete setup for the acquisition of 3D image data and shown its application to the recording of a smooth rotational image sequence.

Figure 4 shows samples of an image sequence recorded with the low-cost cameras introduced in section 3. All images have been cropped to the same portrait section. It is noticeable that the subject is well-placed in the center in the rectified sequence and that global color adaption has evened out most color discrepancies. However, since the subject does not expose the same colors and illumination from every viewing angle, a global approach to color adaption is limited to minimizing an imprecise error function.

The number of possible applications and extensions of this setup is manifold. Approaches to correspondence analysis are being tested at the moment and will offer new possibilities, e.g. for a correspondence-guided local color adaption. To reduce the number of cameras and generate new views, methods for image interpolation such as in [9] are applied to the data obtained from our setup. Different strategies for the camera setup will be researched as well, especially non-planar layouts are expected to yield advantages for 3D-based operations. A multi-view consistent segmentation of the subject from the background is also being researched at the moment.

## References

- [1] P. Baker and Y. Aloimonos. Complete calibration of a multi-camera network. In *OMNIVIS '00: Proc. of the IEEE Workshop on Omnidirectional Vision*, page 134, Washington, DC, USA, 2000.
- [2] P. Eisert. Model-based camera calibration using analysis by synthesis techniques. In *Proc. International Workshop on Vision, Modeling, and Visualization*, pages 307–314, Erlangen, Germany, Nov. 2002.
- [3] U. Fecker, M. Barkowsky, and A. Kaup. Improving the prediction efficiency for multi-view video coding using histogram matching. In *Picture Coding Symposium (PCS 2006)*, 2006.
- [4] I. Feldmann, M. Müller, F. Zilly, R. Tanger, K. Müller, A. Smolic, P. Kauff, and T. Wiegand. HHI test material for 3D video. ISO/IEC JTC1/SC29/WG11 MPEG2008/M15413, April 2008.
- [5] R. Hartley. Theory and practice of projective rectification. *International Journal of Computer Vision*, 35(2):115–127, November 1999.
- [6] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.
- [7] A. Ilie and G. Welch. Ensuring color consistency across multiple cameras. In *ICCV, 2005*, pages 1268–1275, 2005.
- [8] D. Liebowitz and A. Zisserman. Metric rectification for perspective images of planes. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 482–488, Washington, DC, USA, 1998.
- [9] D. Schneider, A. Hilsman, and P. Eisert. Patch-based reconstruction and rendering of human heads. In *Proc. International Conference on Image Processing (ICIP)*, Hong Kong, Oct. 2010.
- [10] T. Svoboda, D. Martinec, and T. Pajdla. A convenient multicamera self-calibration for virtual environments. *Presence: Teleoper. Virtual Environ.*, 14(4):407–422, 2005.