IMAGE REGISTRATION FOR IMAGE QUALITY ASSESSMENT OF PROJECTION DISPLAYS

Ping Zhao, Marius Pedersen, Jon Yngve Hardeberg

Gjøvik University College
Gjøvik, Norway

Jean-Baptiste Thomas

Université de Bourgogne
Dijon, France

ABSTRACT

In the full reference metric based image quality assessment of projection displays, it is critical to achieve highly accurate and fully automatic image registration between the captured projection and its reference image in order to establish a sub-pixel level mapping. The preservation of geometrical order as well as the intensity and chromaticity relationships between two consecutive pixels must be maximized. The existing camera based image registration methods do not meet this requirement well. In this paper, we propose a marker-less and view independent method to use an uncalibrated camera to perform the task. The proposed method including three main components: feature extraction, feature expansion and geometric correction, can be implemented easily in a fully automatic fashion. The experiment results of both simulation and the one conducted in the field demonstrate that the proposed method is able to achieve image registration accuracy higher than 95% in a dark projection room and above 90% with ambient light lower than 30 Lux.

Index Terms— image quality, image registration, spatial distortion, projection display, full reference metric

1. INTRODUCTION

In the past decades, tremendous growth in the use of digital media indicates that our daily life has been greatly impacted by the rapid advancement of imaging technologies. Projection displays among various display technologies have unique advantages such as portability, high resolution, and flexibility on the specification. Recently, there has been an increased popularity of embedding mini projectors in smart phones and video cameras [1]. Furthermore, multiple projections can be tiled up to generate a large perceptually seamless picture with the help of a digital still camera [2]. It is cost effective for customers to visualize information in a very high resolution without issuing a customized manufacturing demand. Hence, Projection Display Image Quality (PDIQ) assessment becomes an increasingly interesting and essential topic among both of scientific research and industrial communities.

In full reference metric based image quality assessments [3, 4, 5], image quality is evaluated with respect to selected attributes. In a typical projection system, cameras are commonly used to acquire the scene, since they are capable to record all pixels in one shot. It is critical to achieve a highly accurate and fully automatic image registration between the captured projection and its reference image; then it is positive to apply metrics which require the distorted image and its reference image share the dimension and resolution. The preservation of geometrical order as well as the intensity and chromaticity relationships between two consecutive pixels on the screen must be maximized. However, existing camera based image registration methods do not meet this requirement well, because they either place simple assumptions on the projections and cameras in order to reduce the problem complexity, or they tend to modify the captured image quality implicitly. The captured images are expected to have various spatial distortions with respect to the relative positions and orientations of the projector, screen, and camera. The camera lens introduces additional spatial distortions. Hence, establishing a robust, accurate and reliable image registration for PDIQ assessment is a non-trivial task.

In this paper, we propose a marker-less and view independent method to use an uncalibrated camera to capture the projection scene and correct its nonlinear spatial distortions. The rest of this paper is organized as follows: first, in Section 2, we present the conventional image registration methods. Then, in Section 3, the proposed method is presented. In Section 4, experimental results are shown. At last, in Section 5, conclusions are drawn based on the data observations.

2. BACKGROUND

Cameras are conventionally calibrated off-line in order to eliminate lens distortion. The camera is typically modelled as a pinhole and the global homography [6, 7] are presented as a 4×3 extended nonsingular projective transformation matrix H. The intrinsic and extrinsic parametric coefficients [8, 9, 10, 11] are estimated by a least-square-fitting technique with respect to a large number of pair-wise feature observations. The captured features are assumed to distributed on a plane in the physical world. The pixel P_o in the reference image corresponds to the pixel P_s on the screen and the pixel $P_c = H \cdot P_s$ in the captured image where s stands for the scal-

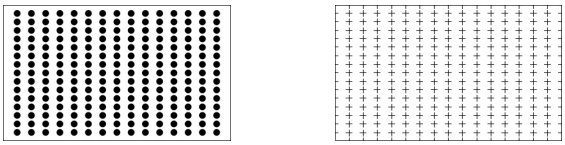


Fig. 1. Projected grayscale pattern images in a sequence

ing factor for the homogeneous coordinates of pixel P_c and it correlates to the camera settings like capturing resolution, focal length and aperture. a_{ij} are collectively called intrinsic parameters of the camera, while r_{ij} and t_i defining the rotation and translation of the view transformation respectively are collectively called extrinsic parameters, and

$$H = s \cdot \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ 0 & a_{11} & a_{12} \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_{00} & r_{01} & r_{02} & t_0 \\ r_{10} & r_{11} & r_{12} & t_1 \\ r_{20} & r_{21} & r_{22} & t_3 \end{bmatrix} \quad (1)$$

The pincushion, barrel and mustache distortions of the camera lens are corrected by estimating the coefficients of inward or outward displacements of feature points from their ideal locations and inversing the transform. The features are detected from a series of captured patterns whose physical dimension and appearance are known in prior. Many formulations [12, 13, 14, 15] are introduced accordingly. The geometric distortions are then corrected by down sampling the captured image and register it with the reference image by inversing the perspective transformations like translation, scaling, rotation, skewing and shearing etc.

In a summary, the conventional image registrations require the camera lens distortion to be identified manually in prior, and the camera settings must be fixed for the use in the future. In real practice of projection displays, an identical projection may appear to have different types of distortions with slightly varied camera positions and/or orientations. The camera is likely to be relocated in the field in order to obtain a view dependent optimized image quality. In such cases, the camera settings will be adjusted and the camera must be recalibrated accordingly. The process of conventional camera calibration involving both offline and online procedures is non-trivial, so it makes camera based PDIQ assessment inflexible and impractical.

3. PROPOSED METHOD

The proposed method has three major components: feature extraction, feature expansion and geometric correction.

3.1. Feature Extraction

Two gray patterns (Figure 1) are generated and projected in full screen size in order to estimate spatial distortions. The dot pattern contains round solid black dots evenly distributed in a $M_d \times N_d$ grid layout. The cross pattern includes crosses that share the center locations and radius with the dots in the dot pattern. Lets denote the captured dot pattern as I_d, then a contour map C can be generated as

$$C = M_c \left(G_a \left(I_d \right) - G_b \left(I_d \right) \right) \quad (2)$$

where the Gaussian filter G_a with kernel size a (empirically $a < 5$) is adopted to reduce screen-door effect [16, 17]. The kernel size a should be kept as small as possible to preserve the details in the captured image. The Gaussian filter G_b with kernel size b (empirically $b > 41$) is adopted to spread energy from highly illuminated pixels to their neighborhoods. The median filter M_c with kernel size c (empirially $c = 3$) is adopted to remove the salt-and-pepper noises, and smooth the detected object contours. False contours might be visible in the generated map C. In this context, we assume the largest most visible continuous object in the scene is the projection on the screen and then the false detections of object contours can be eliminated by applying the binary thresholding:

$$I_b = \begin{cases} 1 & C_i > (1-p) \cdot L_{min} + p \cdot L_{max} \\ 0 & otherwise \end{cases} \quad (3)$$

where C_i denotes the ith pixel in contour map C, L_{min} and L_{max} denote the minimum and maximum gray values in the image respectively, and $p \in [0, 1]$ is a constant threshold. Smaller value of p forces the detected projection boundaries to be compressed and vice versa. The value of p has an inverse effect on the detected dot contours. Hence, the pixels corresponding to positive thresholding are kept and the rests are removed. The algorithm [18] proposed by Suzuki et al. is adopted to determine the contours and their hierarchy relationships in the binary image I_b. The outermost and longest object contour corresponds to the projection area, while the innermost and shortest contours correspond to the dots. The rest of contours are therefore discarded. For each identified contour, a moving window (empirically size equals to 5) is placed along its pixels and a dynamic threshold with respect to local statistics in the corresponding area of the original captured image I_d is calculated. Then, the contour pixel at the window center is shifted toward its neighborhood either horizontally or vertically to achieve the goal of local optimization. The local threshold is determined as

$$T = L \cdot \left(1 - k \cdot \frac{\sigma}{L/2} \right) \quad (4)$$

where σ denotes the standard deviation of gray values within the local window, L denotes the maximum gray scale level (256 for 8 bit image) and the constant k (empirically $k \in [0.1, 0.3]$) indicates the confidence of the image quality of the captured image I_d. In the case of good image quality, the value of k can be scaled down to 0. Otherwise, it should be scaled up. Eventually, we adopt the algorithm [19] proposed by Fitzgibbon et al. to fit dot contours into ellipses with respect to the least square error minimization, so the actual center, size and orientation of each ellipse can be estimated simultaneously. The estimated ellipse centers will be slightly shifted according to the "cornerSubPix" algorithm provided by OpenCV library [20]. Such algorithm incorporates the detected cross centers from the cross pattern image to locally optimize the ellipse centers since the dots and crosses share the same center location and radius.

3.2. Feature Expansion

The detected dot grid needs to be expanded to cover the entire projection area, so that the image registration can be independent from the geometry and content of projected images. In this case, we fit the coordinates of all dot centers in the same row or in the same column into a parametric natural cubic spline function as sample points, and estimate the parametric coefficients accordingly. Once the spline functions are determined, we can generate a smooth parametric cubic spline passing through each set of the feature points. In turn, each spline is extrapolated to intersect with the detected contour of the projection area to generate a pair of new feature points. Thus, in total, $2(M_d + N_d)$ new feature points are generated. The four extreme corners of projection contour can be determined by applying the split-and-merge algorithm [21] proposed by Heckbert et al. iteratively to eliminate pixels until only four corners are left. These corners are used as feature points as well. Eventually, we will have $(M_d + 2) \cdot (N_d + 2)$ feature points covering the entire projection area. The reason to employ the natural cubic spline is to take the advantage of its unique mathematical properties. A typical parametric formulation can be presented as

$$x\left(p_x\right) = \sum_{k=0}^{3} \alpha_{ik} \left(p_x - c_i\right)^k, y\left(p_y\right) = \sum_{k=0}^{3} \alpha_{jk} \left(p_y - c_j\right)^k \quad (5)$$

where $p_x \in [0, M_d + 1]$ and $p_y \in [0, N_d + 1]$ are the two parametric coefficients for the spatial coordinates of a point on the ith and jth spline section respectively; α_{ik} and β_{jk} are the local polynomial regression of kth parametric coefficient of the ith and jth spline sections respectively; $c_i \in [0, M_d + 1]$ represents the parametric coordinate of ith sample point on spline $x\left(p_x\right)$, and $c_j \in [0, N_d + 1]$ represents the parametric coordinates of jth sample point on spline $y\left(p_y\right)$. Each feature point in the expanded dot grid represents one sample point for the corresponding spline. The coefficients are estimated to make sure that around each key point the two consecutive spline sections share the same first and second derivatives; so the whole spline curve is differentiate and continuous below third polynomial order at all possible locations. In addition, the estimated splines are adapted to local variance within each spline section. Higher order spline may not be employed to avoid adaptation to the errors inherited from the capturing process or from the calculations above.

3.3. Geometric Correction

We create a sub-pixel level mapping between the captured image and its reference, and the captured image can be undistorted by down sampling with respect to a specified interpolation method. The basic idea is to register pixels between the Cartesian coordinate system in the camera space and a distortion independent coordinate system defined by the expanded feature grid. Suppose the reference image resolution is given as $N_x \times N_y$ in pixels and the capturing resolution as $M_x \times M_y$ in pixels. Any pixel $P_o = (x, y)$ where $x \in [0, N_x - 1]$ and $y \in [0, N_y - 1]$ in the reference image corresponds to pixel $P_c = (u, v)$ where $u \in [0, M_x - 1]$ and $v \in [0, M_y - 1]$ in the captured image and the pixel p_u in the undistorted image. Their coordinates are defined in the Cartesian coordinate systems and their pixel correspondences in the distortion independent space are

$$Q_o = \left(\frac{x \cdot (M_d + 1)}{(N_x - 1)}, \frac{y \cdot (N_d + 1)}{(N_y - 1)}\right) \quad (6)$$

$$Q_c = \left(\frac{u \cdot (M_d + 1)}{(M_x - 1)}, \frac{v \cdot (N_d + 1)}{(M_y - 1)}\right) \quad (7)$$

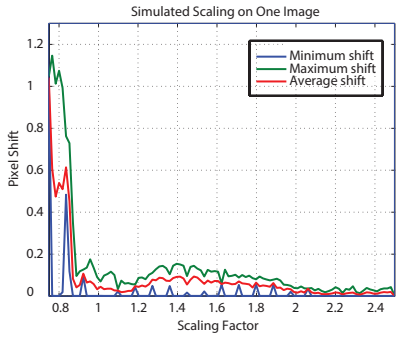
respectively. The correspondence between P_o and Q_o, as well as P_c and Q_c are established with respect to the expanded feature grid which is generated based on cubic splines. Since the undistorted image is expected to exactly registered with the reference image, then the coordinates of P_o is equal to P_u in the distortion independent space. Special attention must be paid to the screen-door effect [16, 17]. The geometric correction may introduce wave-like artifacts. A trade off has to be made between blurring the captured image to register the geometry with the reference image, or distorting the reference image to register it with the captured image. In this paper, we adopt the former approach to make sure that existing full reference metric based image quality metrics can be incorporated without any modification.

4. EXPERIMENT

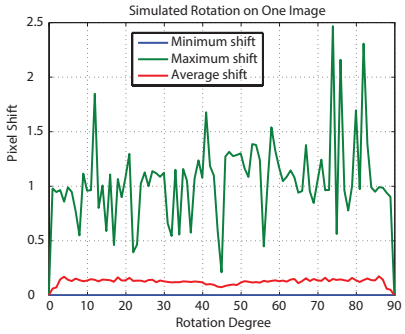
The experiment is performed in a controlled lab environment. A portable LCD projector SONY APL-AW15 (1280x768) is placed in front of a planar screen, and a DSLR Nikon D200 (3872x2592) is used to take the shots. All 24 images from Kodak Photo CD PCD0992 [22] are adopted for the test. We evaluate the proposed method against the pictures either generated by simulation tools or the ones taken in the field.

4.1. Simulation

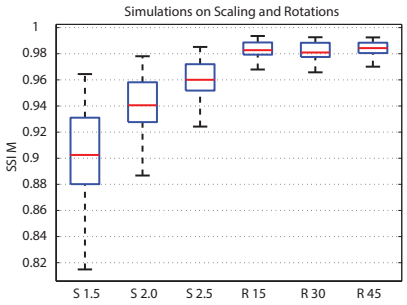
The reference images and pattern images are scaled, rotated and translated respectively at a series of levels to simulate a specific type of spatial distortion. The output images have the same resolution as the captured images. Since the actual distortions are known in prior for the simulation, the image registration accuracy can be evaluated with respect to the maximum absolute displacements of pixels from their ideal locations. In the cases scaling factors are greater than 1 (Figure 2a), the maximum displacements are below 0.2 pixel. These small errors are largely negligible if the capturing resolution is at least two times higher than the reference resolution (very typical). The lowest displacements are given for special rotation angles as expected (Figure 2b). In all other cases, the absolute displacements are between 0.5 and 2 pixels, and they correspond to the mis-adjustments of the contour fine-tune algorithm (Formula 4) due to blurring edges in the captured images. The proposed method is completely independent from



(a) Simulated scaling on a single image



(b) Simulated rotation on a single image



(c) Simulated scaling and rotation on all images

Fig. 2. Simulated scaling and rotation on a single image

spatial translations. We also scale and rotate all 24 testing images in a similar fashion and apply SSIM metric [23] (kernel size 5) to measure the structure similarity between images in luminance channel. This is largely ignored by conventional image registration evaluations. The metric incorporates the visibility of structure errors; it concerns the displayed image content and is able to detect complicated image quality issues like artefacts for example. Figure 2c ("S" for scaling, "R" for rotation) and its corresponding statistics table (Table 1) illustrates that the mean of similarity increases rapidly and the variance becomes smaller and more stable. However, the image rotation has limited influence on the proposed method and since the structure similarity are always above 98%.

4.2. Captured Images

In a controlled lab environment, we use the camera to take pictures of each of the projected reference image at 25 random locations and orientations in the field under low light (0 Lux) and dimmed light (30 Lux) conditions respectively, since the light condition has great impact on the visual experience [24]. Then we apply SSIM metric to the undistorted images and their references due to the lack of ground truth for the projections. The minimum structure similarity is higher than 92% for all cases under the low light condition (Figure 3). The mean is larger than the one under high light condition while the corresponding variance is much smaller (Table 2).

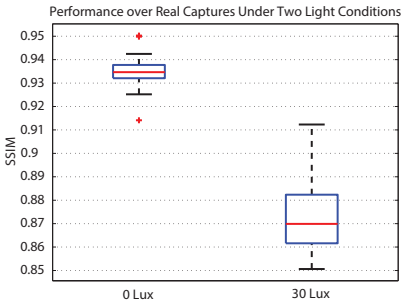


Fig. 3. SSIM for real captures under two light conditions

Table 1. Statistics of SSIM for simulation results

	S 1.5	S 2.0	S 2.5	R 15	R 30	R 45
min	0.815	0.887	0.924	0.968	0.966	0.970
mean	0.902	0.940	0.960	0.983	0.982	0.984
max	0.964	0.978	0.985	0.994	0.993	0.993
95% int.	0.017	0.010	0.007	0.003	0.003	0.003

Table 2. Statistics of SSIM for real captures

	0 Lux	30 Lux
min	0.914	0.851
mean	0.935	0.872
max	0.950	0.909
95% int.	0.002	0.004

5. CONCLUSION

In this paper, we propose a marker-less view independent method to use an un-calibrated camera to achieve a sub-pixel-level registration between the captured projections and their reference images. The preservation of geometrical order as well as the intensity and chromaticity relationships between two consecutive pixels on the display are maximized. The experiments results against distortion simulations and real captured images prove that the registration accuracy is considerably high under typical light conditions of projection systems. By incorporating this method, we can apply all existing full reference image quality metrics to captured projections without any modification to metrics. In the future, we will integrate the method into an unified full reference metric based image quality assessment framework for projection displays and study the perceptual properties of displayed images with respect to the correlations between human observations and metrics results despite of the actual projection geometries.

6. REFERENCES

[1] C. H. Son and Y. H. Ha, "Color Correction of Images Projected on a Colored Screen for Mobile Beam Projector," *Imaging Science and Technology*, vol. 52, no. 3, pp. 1–11, 2008.

[2] R. Raskar, M. S. Brown, R. Yang, W. C. Chen, G. Welch, H. Towles, B. Seales, and H. Fuchs, "Multi-projector Displays Using Camera-based Registration," in *IEEE Visualization*, San Francisco, USA, 1999, pp. 161–168.

[3] M. Pedersen, N. Bonnier, J. Y. Hardeberg, F. Albregtsen, and B. N., "Attributes of Image Quality for Color Prints," *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 011016–1–011016–13, Jan. 2010.

[4] M. Pedersen, N. Bonnier, J. Y. Hardeberg, and F. Albregtsen, "Image Qualitymetrics for The Evaluation of Print Quality," *Proceedings of the SPIE*, vol. 7867, pp. 19, Jan. 2011.

[5] T. Eerola, J. K. Kamarainen, L. Lensu, and H. Kalviainen, "Framework for Applying Full Reference Digital Image Quality Measures to Printed Images," in *Scandinavian Conference on Image Analysis*, 2009, vol. 5575, pp. 99–108.

[6] O. Bimber, D. Iwai, G. Wetzstein, and A. Grundhöfer, "The Visual Computing of Projector-Camera Systems," *Computer Graphics Forum*, vol. 27, no. 8, pp. 2219–2245, 2008.

[7] A. Agarwal, C. V. Jawahar, and P. J. Narayanan, "A Survey of Planar Homography Estimation Techniques," Tech. Rep., Centre for Visual Information Technology, 2005.

[8] Z. Y. Zhang, "Flexible Camera Calibration By Viewing a Plane From Unknown Orientations," in *International Conference on Computer Vision*, 1999, pp. 666–673.

[9] P. F. Sturm and S. J. Maybank, "On Plane-Based Camera Calibration: A General Algorithm, Singularities, Applications," in *Computer Vision and Pattern Recognition*, 1999, pp. 1432–1437.

[10] Z. Y. Zhang, "A Flexible New Technique for Camera Calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[11] K. L. Low and A. Ilie, "Computing a View Frustum to Maximize an Object's Image Area," *Journal of Graphics Tools*, vol. 8, no. 1, pp. 3–15, 2003.

[12] R. Strand and E. Hayman, "Correcting Radial Distortion by Circle Fitting," in *Procedings of the British Machine Vision Conference*. 2005, pp. 9.1–9.10, British Machine Vision Association.

[13] T. Thormahlen, H. Broszio, and I. Wassermann, "Robust Line-Based Calibration of Lens Distortion From A Single View," in *International Conference on Computer Vision / Computer Graphics Collaboration Techniques and Applications*, Nice, France, 2003, pp. 105–112.

[14] F. Bukhari and M. N. Dailey, "Robust Radial Distortion from a Single Image," *International Symposium on Visual Computing*, vol. 6454, pp. 11–20, 2010.

[15] G. Q. Wei and S. D. Ma, "Implicit and Explicit Camera Calibration : Theorv and Experiments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 5, pp. 469–480, 1994.

[16] H. Arora and A. Namboodiri, "Projected Pixel Localization and Artifact Removal in Captured Images," in *TENCON 2008 - 2008 IEEE Region 10 Conference*. Nov. 2008, pp. 1–5, Ieee.

[17] L. Zhang and S. Nayar, "Projection Defocus Analysis for Scene Capture and Image Display," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 907–915, 2006.

[18] S. Suzuki and K. Abe, "Topological Structural Analysis of Digitized Binary Images by Border Following," *Computer Vision, Graphics, and Image Processing*, vol. 30, no. 1, pp. 32–46, 1985.

[19] A. Fitzgibbon and R. Fisher, "A Buyer's Guide to Conic Fitting," *Procedings of the British Machine Vision Conference 1995*, pp. 51.1–51.10, 1995.

[20] G. Bradski and A. Kaehler, *Learning OpenCV*, OReilly Media, Inc., Sebastopol, CA, USA, first edit edition, 2008.

[21] P. S. Heckbert and M. Garland, "Survey of Polygonal Surface Simplification Algorithms," Tech. Rep. May, School of Computer Science, CarnegieMellon University, 1997.

[22] R. Franzen, "PhotoCD PCD0992," 1999.

[23] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P Simoncelli, "Image Quality Assessment: from Error Visibility to Structural Similarity," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 13, no. 4, pp. 600–12, Apr. 2004.

[24] P. Zhao, M. Pedersen, J. Y. Hardeberg, and J.-B. Thomas, "Camera-based Measurement of Relative Image Contrast in Projection Displays," in *4th European Workshop on Visual Information Processing (EUVIP)*, Paris, France, 2013, pp. 112–117, IEEE.