Registration of Multi-view Images of Planar Surfaces

Radomír Vávra and Jiří Filip

Institute of Information Theory and Automation of the ASCR, The Czech Republic

Abstract. This paper presents a novel image-based registration method for high-resolution multi-view images of a planar material surface. Contrary to standard registration approaches, this method aligns images based on a true plane of the material's surface and not on a plane defined by registration marks. It combines the camera calibration and the iterative fitting of desired position and slant of the surface plane, image re-registration, and evaluation of the surface alignment. To optimize image compression performance, we use an error of a compression method as a function evaluating the registration quality. The proposed method shows encouraging results on example visualizations of view- and illumination-dependent textures. In addition to a standard multi-view data registration approach, it provides a better alignment of multi-view images and thus allows more detailed visualization using the same compressed parameterization size.

1 Introduction

Acquisition of a multi-view appearance is often used to achieve realistic visualization of textured objects. This paper is focused on visualization techniques which deal with multiple photos of the same planar surface acquired from different positions. This way a photo-realistic appearance of the surface can be captured, but the acquired photos cannot be directly used for rendering. They have to be mutually registered and rectified first.

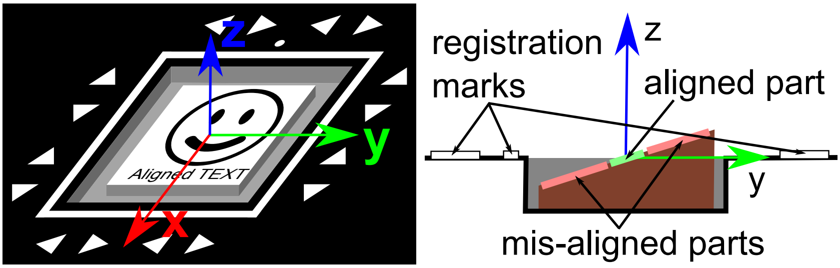
The most general function of multi-view photos of a planar surface is probably the *Bidirectional Texture Function* (BTF) proposed by Dana et al [1]. This seven-dimensional function $BTF(\lambda, x, y, \theta_i, \phi_i, \theta_v, \phi_v)$, describes reflectance properties of a material where λ is a wave length of incoming light or just a color channel; (x, y) are spatial coordinates on a surface of the material, and θ, ϕ are the elevation and azimuthal spherical angles of the vector of illumination- and view-directions (see [2]). A typical size of a BTF dataset containing thousands of images amounts to several gigabytes.

An another example of multi-view data is *Surface Light Field* [3]. It can be defined as a subset of a BTF with a fixed illumination direction.

Processing of acquired multi-view data consists of two steps: data registration and compression. Although the measured materials are planar, their rough structure often shows height variations causing significant variance of their appearance depending on illumination- and view-directions. The final appearance is

K.M. Lee et al. (Eds.): ACCV 2012, Part IV, LNCS 7727, pp. 497–509, 2013.

affected by self-occlusions, shadows, inter-reflections, and subsurface-scattering. This is the reason why the features of the material are non-stationary and cannot be directly used for reliable feature-based registration. Due to this, we use registration marks placed on a reference plane, which allows the measured sample to be easily replaced. However, the sample's orientation and shift with respect to the registration plane is unknown (see Fig. 1). Although one might use tilt/shift mechanical stage to fine-tune this misalignment manually, it is expensive and far less accurate than the proposed approach.



(a) (b)

Fig. 1. (a) A measured sample, the reference plane with the registration marks (white frame and triangles) and the world coordinate system. (b) a cross-section of (a), shows aligned and misaligned parts of the surface of the measured material when the positions of the reference plane and of the measured surface differ.

The registration based on the registration marks properly aligns only those parts of the measured sample which lie close to the plane specified by the marks (see Fig. 1-b). This is not an issue when a distance between the reference and the material's surface planes is small; the slant difference is also negligible. This might be less relevant when registered data are of low resolution or are used directly for a rendering. However, it has a significant impact on the resulting visual quality if compression methods for multi-view data are used (e.g., all the classes of global factorization methods based on PCA [4] or clustering [5], etc.).

The contribution of the paper – The main contribution of this paper is a novel technique for registration of multi-view images of planar surfaces that aligns measured planar surfaces regardless of their slight height and slant differences from the reference registration plane. Such data alignment allows us to better exploit the power of compression techniques and produce an image closer to the original image reconstruction using the same number of parameters.

Organization of the paper – The paper starts with description of related past research in Section 2, the brief description of a standard registration procedure is explained in Section 3. A basic overview of the proposed approach is presented in Section 4, while details of a camera calibration and an iterative registration procedure are given in Sections 5 and 6. Results of the presented method are shown in Section 7. Main conclusions and suggestions for future work are outlined in Section 8.

2 Related Work

The proposed approach is, to a great degree, related to methods for multi-view data registration and methods which reconstruct a 3D surface of a sample.

View-dependent image-based data are generally captured by setups based on gonioreflectometers realizing four mechanical degrees of freedom (Sattler et al [4], Holroyd et al [6]) and setups which reduce measurement times or complexity by using multiple lights or sensors simultaneously (Müller et al [7], Neubeck et al [8]). Sattler et al [4] measured BTF data and registered it using a projective transformation based on registration marks (see Section 3). Such registration is sufficient as long as the resolution of the captured images of the measured material is low enough and the plane which represents the surface of the material is close to the reference registration plane. Additionally, the problem of registration is not so pressing in this case because the authors applied compression to data of individual views separately and thus the multi-view correspondence does not affect performance of compression to the same degree as other compression methods do [9,10] .

Neubeck et al [8] were aware of a problem with BTF alignment. Their work is the most relevant to our paper as they propose to evaluate quality of BTF alignment using a function which computes average Euclidean distance between the intensities of those neighboring views that share the same lighting direction. They tested several plane heights and selected the one for which this function is minimal. In contrast, our technique allows us to not only compensate for height misalignment, but also for mutual rotation of the registration and sample planes, without need for repetitive measurement.

Müller et al [7] used a setup with no moving parts. Therefore, positions of the image sensors are known in advance and registration can be done in sub-pixel accuracy without the need for registration marks. In another paper, Müller et al [11] proposed an approach attempting to align individual BTF pixels based on optimization techniques reducing certain intra-variations in the data. This method rotates individual ABRDFs to achieve better global compression performance, and therefore it requires storage of an additional per-pixel rotation map. Nevertheless, in both cases the accuracy of the measurement or the fit depends on an initial position of a calibration plane and its difference from the plane representing the surface of a material, which can be compensated when our registration technique is used.

Ruiters and Klein [12] published a technique which represents the appearance of a material using a combination of surface depth-map and spatially-varying reflectance. The authors define a dense reference mesh and align its polygons to best fit the original data to estimate a depth-map of near-flat surfaces. This technique can deal with materials having variable surface-height; however, our method is easier to implement and it is computationally less expensive. We do not attempt to interpret surface depth (which might not even be possible for some translucent materials) but to find an alignment that maximizes the quality of registration of multi-view data.

Methods for simultaneous acquisition of shape and reflectance exist; e.g., Müeller's setup can be used for an acquisition of even non-planar objects [13]. Holroyd et al [6] used a system based on a spherical gantry, where each arm is fitted with a camera and a high frequency, spatially-modulated light sharing a common focal point and an optical axis. The proposed measurement method exploits multi-view stereo, phase-based profilometry, and light descattering to avoid 2D-3D data registration problems and leverage a restrictive assumption about BRDF as is often done by related methods. Weinmann et al [14] added multiple projectors into the setup [7] for a detailed 3D acquisition of an object. The projectors emit structured light used for unique identification of points on a surface of the object. Although such approaches allow us to find an exact position of a flat material surface as well, the required hardware would unnecessarily increase the financial cost of a setup with no advantage compared to the technique we have proposed.

Additionally, our method is robust, easy to implement, computationally efficient, and optimal in terms of the compression method used.

3 A Standard Image Registration Approach

This section outlines a principle of image registration. Given a set of photos of the same planar surface, registration applies a projective transformation to all the photos so that the features of the planar surface are aligned across all of the transformed images. In a standard registration approach depicted in Fig. 4-a, registration marks are placed around the photographed planar surface of a material. First, their 2D coordinates are found in all of the photos. Then, projective transformation matrices projecting these points to the desired target coordinates are computed. Finally, all the photos are transformed using these projective matrices. All the registered images have the same coordinate system.

A projective transformation, also called a *homography*, is a 2D coordinate transformation preserving straight lines (see [15] for a 2D coordinate transformation survey). Given a photo of a planar surface we want to transform together with an orthonormal coordinate system (u, v) of the photo, $\mathbf{m} = [u, v, 1]^T$ denotes an augmented point in this system in the planar surface. An augmented 2D point $\mathbf{m}' = [u', v', 1]^T$ in a new orthonormal coordinate system (u', v') into which we transform can be computed as $s\mathbf{m}' = \mathbf{H}\mathbf{m}$, where $\mathbf{H}$ is the 3×3 homography matrix and s is an arbitrary scalar. The homography can be computed if we know coordinates of at least four corresponding points in the source and target images and it is defined uniquely up to a scale factor. If there are more than four such points and they are not perfectly corresponding, the homography has to be computed in a least-square sense (e.g., [16]).

4 An Overview of the Proposed Registration Method

When a standard registration approach is used, the features which do not lie in the registered plane will not be aligned after application of the projective

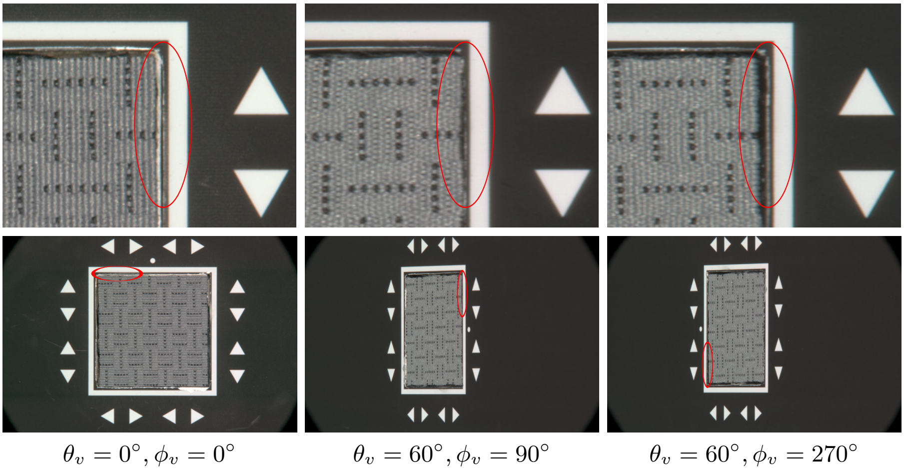


Fig. 2. An example of incorrect registration of a material's surface for images taken from different views, while the reference plane (the registration marks) is registered correctly (top row)

transformation as it is shown in Fig. 2. The same pixel will correspond to different physical points on the surface of the material (see Fig. 3-a,c). Therefore, we have to estimate the plane which represents the surface of the registered sample for appropriate registration (Fig. 3-b,d). Unfortunately, this plane (i.e., its offset and orientation with respect to the world coordinate system) cannot be determined accurately enough from the specimen of the measured sample, or directly from the acquired photos.

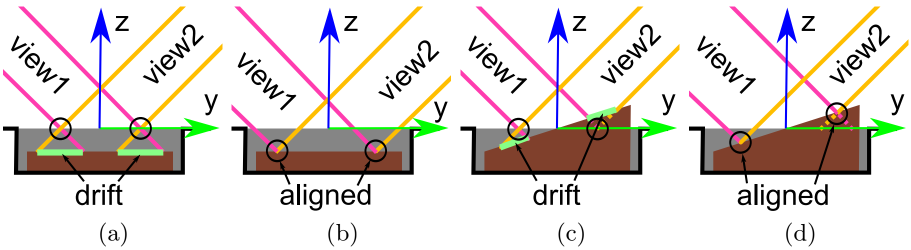


Fig. 3. When the plane defined by the registration marks is chosen for registration (a,c), the same pixel will correspond to different points on the surface of the measured sample, i.e., resulting in a small drift. When the appropriate plane on the surface of the material is chosen, the drift disappears (b,d).

We propose to find the position and slant of an ideal registration plane as follows. First, the reference plane defined by the registration marks is taken. As we expect that the estimated plane which represents the surface of the material is close to the reference plane, a new hypothetical position of the estimated plane can be generated using a slight modification of the reference plane position by shifting it in a direction of its normal vector and/or by tilting it. Finally, the best estimation of the position and slant of the plane can be found by repeatedly alternating adjustment of the position and slant, registration of photos and evaluation of alignment for the surface features. A principle of the proposed registration method is expanded upon in Fig. 4-b. The method consists of two main parts discussed in more detail in the next two sections. The first one is the camera intrinsic and extrinsic parameters estimation (Section 5) and the second is the iterative fitting of the estimated plane position (Section 6).

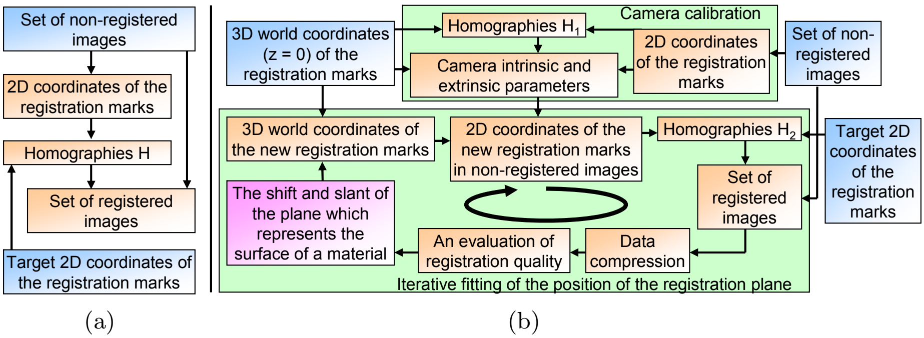


Fig. 4. A standard image registration approach (a) and a scheme of the proposed method (b)

5 Camera Calibration

The calibration of a camera is a process where we look for a 3×3 matrix $\mathbf{A}$ of the camera's intrinsic parameters, and the camera's extrinsic parameters which consist of a 3×3 rotation matrix $\mathbf{R}$ and a translation vector $\mathbf{t}$. While intrinsic parameters $\mathbf{A}$ do not change as long as the internal setup of the camera does not change (e.g., focal length), the extrinsic parameters change when the camera moves, i.e., all the photos of the measured planar sample should have the same camera's intrinsic parameters but the corresponding extrinsic parameters can be different. Using the camera parameters we can project an augmented 3D point $M = [x, y, z, 1]^T$ from the world coordinate system into an image by

$$s\mathbf{m} = \mathbf{A}\left[\mathbf{R}\,\mathbf{t}\right] M, \tag{1}$$

where $\mathbf{m} = [u, v, 1]^T$ denotes an augmented 2D point and s is an arbitrary scalar. The usual pinhole camera model is assumed.

In a case where we work with extensive view- and illumination-dependent data (e.g., BTF), the procedure of calibrating the camera and iteratively fitting the position of the estimated plane should start with selection of their representative subset. Although the registration plane can be determined more precisely if all the photos are used, the estimation process would take a very long time if there were hundreds or even thousands of images. Therefore, we recommend working with images of one surface light field only; i.e., images where illumination directions are fixed while view angles are changed.

Next, the registration marks are found for all images in the subset. Without loss of generality, we define the world coordinate system so that the reference registration marks plane is on $z = 0$ (see Fig. 1). Spatial coordinates of the marks (x, y) should correspond to their real positions in a natural system of units, i.e., millimeters. Now, projective transformation matrices $\mathbf{H}_1$ projecting points from the reference plane to photos of the plane are computed based on the coordinates of the registration marks. From 1, we have

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{r}_1 \, \mathbf{r}_2 \, \mathbf{r}_3 \, \mathbf{t} \end{bmatrix} \begin{bmatrix} x \\ y \\ 0 \\ 1 \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{r}_1 \, \mathbf{r}_2 \, \mathbf{t} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{H}_1 \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \tag{2}$$

where $\mathbf{r}_i$ are column vectors of the rotation matrix $\mathbf{R} = [\mathbf{r}_1 \, \mathbf{r}_2 \, \mathbf{r}_3]$. Using the knowledge that column vectors of $\mathbf{R}$ are orthonormal and the matrix $\mathbf{A}$ is upper triangular, the camera intrinsic and extrinsic parameters can be derived from homographies $\mathbf{H}_1$. For a detailed explanation we refer the reader to Zhang's paper [17]. For camera calibration we have used the Camera Calibration Toolbox for Matlab[1] which implements Zhang's work.

6 Iterative Fitting of the Position of the Registration Plane

We can define an almost arbitrary position of the expected plane which represents the surface of a material in the world coordinate system just by setting new z-coordinates of the registration marks. If we project the estimated 3D coordinates of the new registration marks back to non-registered photos, we obtain new 2D coordinates of the registration marks in the coordinate systems of the photos. We can now register the images by specification of target coordinates of the registration marks and by computation of homography matrices $\mathbf{H}_2$ for registration (see Fig. 4-b). The latter should not be confused with the homography matrices $\mathbf{H}_1$ mentioned above, which project points in the reference registration marks plane of the world coordinate system to the non-registered images. In contrast, these new homography matrices $\mathbf{H}_2$ project points from the non-registered photos to the registered images as homography matrices $\mathbf{H}$ in a standard registration approach (Fig. 4-a).

Therefore, we suggest a novel iterative method for the position and slant of the registration plane estimation, image registration and alignment of surface features evaluation depicted in Fig. 4-b. As we look for optimal vertical shift and slant of the material surface plane, three parameters have to be found: a z-coordinate of an auxiliary point $P = [0, 0, z]^T$ which lies in the surface

[1] `http://www.vision.caltech.edu/bouguetj/calib_doc/`

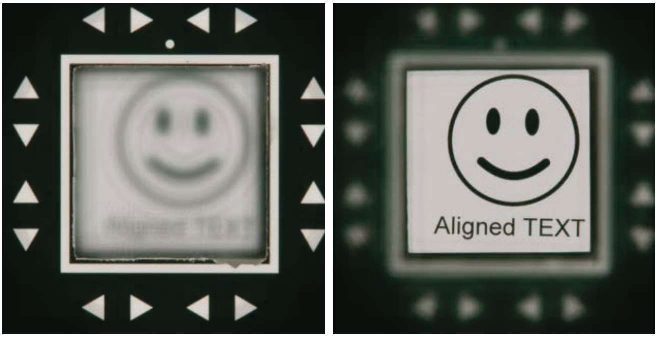
plane, an elevation θ of the normal vector of the surface, and an azimuthal angle φ of the normal vector. As a search state space is three-dimensional and the image registration, compression and evaluation function execution can be computationally demanding, at least the local minimum can be found quickly by alternating between estimation of individual parameters. As there can be significant variations in height on the material's surface, there may be more than one good surface plane position.

Our goal is to provide as accurate a visualization of the measured material sample as possible. As the visualization quality relies mostly on visual quality after data compression, evaluation functions which estimate alignment of surface features should reflect properties of the selected compression method. Therefore, the error of a compression technique will be used as an objective quality measure. As we work with only a subset of all of the photos, the compression as well as its error evaluation can be done quickly enough to be practical. An ideal position of the plane is the one where the compression (i.e., rendering) error is minimal. One iteration of reference plane modification, data registration, and visual quality evaluation after the compression takes about one second depending on counts of the registered pixels and images. In order to avoid local minima the search space was sampled uniformly, alternating between estimation of the three parameters (height z, plane normal's elevation θ, and azimuth φ) with the following step sizes: z=0.1 mm (range [-2,2]mm), $\theta=0.1^o$ (range $[0,3]^o$), $\varphi=10^o$ (range $[0,360]^o$), and then refined near a global minimum (step sizes: z=0.01 mm, $\theta=0.01^o$, $\varphi=1^o$). Typically, around 800 iterations are necessary to find a proper orientation and height of the registration plane.

7 Results

This section illustrates performance of the method on two registration experiments using artificial and real data. In the experiments we used the PCA compression of all registered images [9] and applied its data reconstruction error as a registration performance evaluation function in the proposed method. All pixels selected from individual BTF images are ordered into vectors and centered using the mean BTF image vector. All these vectors form a matrix $\mathbf{B}$, whose PCA is computed. The individual eigenvalues from the resulting diagonal matrix weight the importance of the resulting eigenvectors. A limited set k of eigenvectors is used to reconstruct the original n images, where $k << n$. The PCA-based methods are the most common in multi-view data compression; however, any other global BTF data compression technique would also benefit from the proposed algorithm.

In the first experiment a flat paper printout was used, positioned approximately one millimeter below the reference registration plane. We took 80 different views on the plane which uniformly covered a hemisphere of viewing directions. An illumination direction was fixed in a direction opposite to the reference plane's normal vector. When a standard registration approach was used, only the reference plane features were aligned, while the misalignment in individual images caused the mean image of all the registered images to be blurred in the area of the measured sample (see Fig. 5-a). In contrast, when the proposed approach was applied we obtained the mean image shown in Fig. 5-b, where the desired surface was aligned well but the registration marks were blurred. The estimated surface plane's deviation is 1.24 millimeters bellow the reference plane, its normal vector elevation is 0.29° and its azimuthal angle is 176°.



standard registration proposed approach

Fig. 5. An extreme example of registration based on the reference plane (left) and based on the plane which represents the true surface of the sample found using the proposed method (right) when the heights of the reference plane and the plane of the surface differ considerably. Mean images of the 80 registered images are depicted here.

In the second experiment, five BTF samples were registered using standard and proposed approaches. The samples *wood01*, *fabric01*, *fabric02*, *fabric03*, and *leather01* were taken from the UTIA BTF database[2]. The results of our method are shown in Fig. 7 and mark a considerable improvement against the standard registration approach without alignment. The compression of data registered in a standard way (Fig. 7-b) leads to data visualization that is less sharp in comparison with the non-compressed aligned data (Fig. 7-a) considered the ground-truth. The compression after application of the proposed data registration method leads to considerably sharper images (Fig. 7-c). Note that in both cases the same compressed parametric representation is used (50 PCA components allowing real-time rendering). Registration of such a BTF sample comprising 6561 images typically takes around five hours on Intel Xeon 2.7 GHz using our Matlab implementation using six cores. However, due to the massive size of datasets (415 GB) much of this time is consumed by disk data transfer operations. Note that a smaller visible area of non-aligned datasets (Fig. 7-b) is due to cropping of visual artifacts at borders of individual misaligned images. As the proposed alignment method re-projects original locations of registration marks, the registered images are slightly shifted and scaled. Therefore, their fair pixel-wise comparison (e.g., using RMSE, SSIM) with the original image is impossible.

[2] `http://btf.utia.cas.cz`

Table 1. The estimated values of shift (z) and slant (θ,φ) for the tested samples

sample	height z [mm]	elevation θ	azimuth φ
wood01	-0.49	1.15°	101°
fabric01	-0.20	0.15°	128°
fabric02	0.12	0.20°	260°
fabric03	-0.30	0.43°	323°
leather01	0.10	0.05°	354°

As there is no robust texture-similarity measure available, we performed a psychophysical experiment with 5 naive subjects comparing Fig. 7-a with Fig. 7-b and c in a random order. The (c) was always perceived as more visually similar to the (a) than to the (b).

Tab. 1 shows estimated values of vertical position and orientation of the estimated plane with respect to the reference plane. The images show that the more the sample's plane deviated from the reference plane, the higher visual improvement was achieved as, e.g., for the sample of *wood01* in Fig. 7. From the values shown it is apparent that even when the sample is aligned with the registration plane as much as possible, the estimated differences are still relatively high. Finally, we remark that the visual effects of such misalignment are more pronounced if the resolution of captured images is higher.

Speed of the algorithm depends on the size of user-defined patches on the planar surface that are used for registration quality evaluation, as well as on a number of multi-view images. Fig. 6 shows execution times for a single iteration of the algorithm depending on the number of pixels and images processed. While the speed increases almost linearly with the number of pixels, it depends on the number of images n with $O(n^3)$ due to PCA compression used.

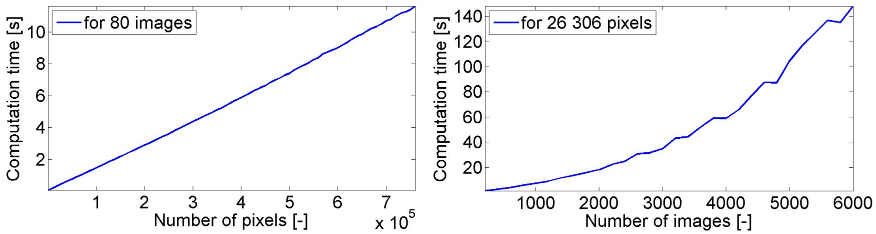


Fig. 6. Computational time of one iteration of the algorithm depends on the number of pixels used for quality evaluation (left), and on the number of processed images (right)

The proposed method is very robust. Its only obvious limitation is that it cannot guarantee a correct alignment for surfaces having wide height variations or several possible alignment heights (see, e.g., material *fabric03* in Fig. 7). However, even in such a case the material will be aligned to minimize the compression/rendering error. Additionally, only pixels which belong to the required height can be selected by a user-defined mask and can be taken into account during the registration to achieve even better alignment for such materials.

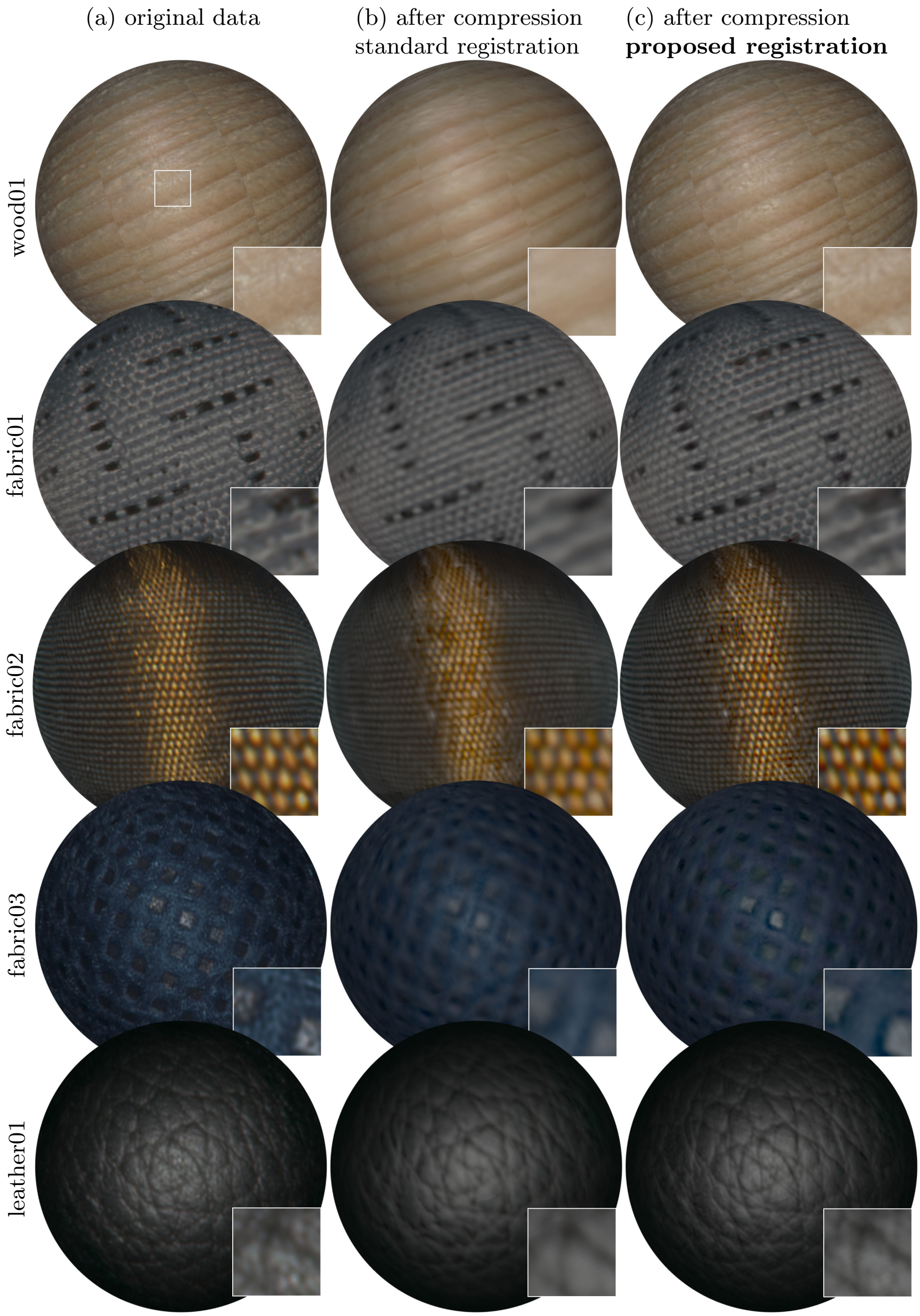


Fig. 7. A comparison of BTF visualization: (a) rendering using all 6561 images, (b) and (c) rendering from a compressed representation using only 50 eigen-images without and with application of the proposed alignment method

8 Conclusions and Future Work

In this paper we focus on the correct registration of multi-view images of planar material surfaces. Our approach exploits the fact that the reference registration plane and measured sample plane may be misaligned. When this misalignment is found and compensated from the measured dataset during the registration stage, a better material features alignment is achieved. Quality of the registration is verified by a reconstruction error of the data compression method. Consequently, the proposed approach allows more efficient application of multi-view data compression approaches, i.e., producing sharper images using the same size of compressed parametric representation. The proposed method is robust, easy to implement, and computationally efficient.

Acknowledgement. This work has been supported by GAČR grants 103/11/0335 and 102/08/0593, EC Marie Curie ERG 239294, and CESNET grant 409. We would like to thank Škoda-Auto a.s. for providing samples for measurement.

References

1. Dana, K., van Ginneken, B., Nayar, S., Koenderink, J.: Reflectance and texture of real-world surfaces. ACM Transactions on Graphics 18, 1–34 (1999)
2. Filip, J., Haindl, M.: Bidirectional texture function modeling: A state of the art survey. IEEE TPAMI 31, 1921–1940 (2009)
3. Gortler, S.J., Grzeszczuk, R., Szeliski, R., Cohen, M.F.: The lumigraph. In: Proceedings of SIGGRAPH 1996, pp. 43–54 (1996)
4. Sattler, M., Sarlette, R., Klein, R.: Efficient and realistic visualization of cloth. In: Proceedings of the 14th Eurographics workshop on Rendering, pp. 167–177 (2003)
5. Havran, V., Filip, J., Myszkowski, K.: Bidirectional texture function compression based on multi-level vector quantization. Computer Graphics Forum 29, 175–190 (2010)
6. Holroyd, M., Lawrence, J., Zickler, T.: A coaxial optical scanner for synchronous acquisition of 3D geometry and surface reflectance. ACM Trans. Graph. 29, 99:1–99:12 (2010)
7. Müller, G., Meseth, J., Sattler, M., Sarlette, R., Klein, R.: Acquisition, synthesis and rendering of bidirectional texture functions. Computer Graphics Forum 24, 83–109 (2005)
8. Neubeck, A., Zalesny, A., Gool, L.V.: Viewpoint consistent texture synthesis. In: Proceedings of the 3D Data Processing, Visualization, and Transmission, 2nd International Symposium, pp. 388–395 (2004)
9. Koudelka, M., Magda, S., Belhumeur, P., Kriegman, D.: Acquisition, compression, and synthesis of bidirectional texture functions. In: Texture 2003, pp. 47–52 (2003)
10. Müller, G., Meseth, J., Klein, R.: Compression and real-time rendering of measured BTFs using local PCA. In: Vision, Modeling and Visualisation, pp. 271–280 (2003)
11. Müller, G., Sarlette, R., Klein, R.: Data-driven local coordinate systems for image-based rendering. Computer Graphics Forum 25 (2006)
12. Ruiters, R., Klein, R.: Heightfield and spatially varying BRDF reconstruction for materials with interreflections. Computer Graphics Forum 28, 513–522 (2009)

13. Müller, G., Bendels, G.H., Klein, R.: Rapid synchronous acquisition of geometry and BTF for cultural heritage artefacts. In: The 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST), pp. 13–20 (2005)
14. Weinmann, M., Schwartz, C., Ruiters, R., Klein, R.: A multi-camera, multi-projector super-resolution framework for structured light. In: International Conference on 3DIMPVT, pp. 397–404 (2011)
15. Szeliski, R.: Image alignment and stitching: A tutorial. Found. Trends. Comput. Graph. Vis. 2, 1–104 (2006)
16. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2000) ISBN: 0521623049
17. Zhang, Z.: Flexible camera calibration by viewing a plane from unknown orientations. In: ICCV, vol. 1, pp. 666–673 (1999)