Eurographics Italian Chapter Conference (2010)
E. Puppo, A. Brogni, and L. De Floriani (Editors)

Improving 2D-3D registration by mutual information using gradient maps

Gianpaolo Palma[†1,2], Massimiliano Corsini[2], Matteo Dellepiane[2], Roberto Scopigno [2]

[1]Department of Computer Science, University of Pisa, Italy
[2]Visual Computing Lab, ISTI-CNR, Pisa, Italy

Abstract

In this paper we propose an extension for the algorithms of image-to-geometry registration by Mutual Information(MI) to improve the performance and the quality of the alignment. Proposed for the registration of multi modal medical images, in the last years MI has been adapted to align a 3D model to a given image by using different renderings of the model and a gray-scale version of the input image. A key aspect is the choice of the rendering process to correlate the 3D model to the image without taking into account the texture data and the lighting conditions. Even if several rendering types for the 3D model have been analyzed, in some cases the alignment fails for two main reasons: the peculiar reflection behavior of the object that we are not able to reproduce in the rendering of the 3D model without knowing the material characteristics of the object and the lighting conditions of the acquisition environment; the characteristics of the image background, especially non uniform background, that can degrade the convergence of the registration. To improve the quality of the registration in these cases we propose to compute the MI between the gradient map of the 3D rendering and the gradient map of the image in order to maximize the shared data between them.

Categories and Subject Descriptors (according to ACM CCS): I.2.10 [Vision and Scene Understanding]: Intensity, color, photometry, thresholding—I.3.7 [Three Dimensional Graphics and Realism]: Color, shading, shadowing and texture—I.4.8 [Scene Analysis]: Shading—I.4.1 [Digitization and Image Capture]: Imaging Geometry—I.4.3 [Enhancement]: Registration—

1. Introduction

The geometric registration or alignment of a set of images of an object over its 3D model is a important task for all the applications related to color mapping and reflectance properties estimation. The main purpose is to align one or more images of the same object taken at different times and from different viewpoints during a photographic campaign to the geometry of the object acquired through 3D scanning. For each image, the outputs are the intrinsic and extrinsic camera parameters that describe how the 3D points are projected on the image plane. In the past years, several algorithms have been proposed to estimate accurately these parameters.

Several proposed approaches are inspired from medical image processing, specifically from multimodal image registration. The main problem in medical imaging is the registration of images coming from different sensors, such as magnetic resonance (MR), computerized tomography (CT), PET, x-rays, and so on. Most of the algorithms developed in this field are based on Mutual Information, a statistical measure of dependency between two data sources. This measure can be employed efficiently for both 2D/2D and 2D/3D registration, by setting up an optimization framework where the parameters of the geometric transformation associated with the registration are calculated by maximizing the mutual information. In the image-to-geometry registration context, the 3D model is aligned to a given image by using different renderings of the model and a gray-scale version of the input image.

The main issue regarding the use of mutual information for 2D/3D registration is the choice of a rendering process

† gianpaolo.palma@isti.cnr.it

that correlates the 3D model with the images to align. The main problem is that the input images contain texture and unknown lighting conditions: this could make their visual appearance very different from a rendering of the geometry. To solve this problem, Viola and Wells [VW97] proposed using surface normals and image brightness to correlate shading variations on the image with the model surface. Corsini et al. [CDPS09] extend this idea by using several types of renderings, such as ambient occlusion, normal map, reflection map, silhouette map, and combined versions of them. These type of renderings are based on geometric proprieties related to the visual appearance of the model but generating them does not entail knowing the lighting environment of the scene.

We propose a further extension of the approach in [CDPS09] where we maximize the mutual information between the gradient map of the rendering of the 3D model and the gradient map of the image.

2. Related Work

Image registration is a very popular research topic. Hundreds of different approaches and practical applications have been proposed. We will focus on one of the most promising groups of methods for multi-modal registration: the ones based on Mutual Information (MI). Two of the first methods of this kind were developed by Viola and Wells [VW97] and by Maes et al. [MCV*97]. The Viola's alignment approach uses the mutual information between the surface normal and the image brightness to correlate the shading variations of the image with the surface of the model. Leventon et al. [LWG97] extended this alignment framework to use multiple views of the object when a single image does not provide enough information. Since then, several registration methods based on MI have been proposed (see [PMV03] for a comprehensive overview).

There are four keys issues in the use of the MI: preprocessing, measure, transformation and optimization. The preprocessing entails any image processing to prepare and improve the image for registration (low-pass filtering to remove the noise, extraction of region of interest, image resampling). In the registration procedure several definitions of the mutual information measure can be used. There exist measures based on the conditional and joint entropy, where we can choose different definitions of entropy, and measures based on the Kullback-Leibler distance between two distributions. Furthermore, several adaptations of mutual information have been proposed: normalization with respect to the overlapping part of the image (Normalized Mutual Information [SHH99], Entropy Correlation Coefficient [MCV*97]) and inclusion of spatial information. A method of incorporating spatial information is to combine mutual information with the gradient, as in [PMV00] where the MI measure seeks to align gradient vectors of large magnitude as well as of similar orientation. Another important key issue is how to model the transformation between the images. Most of proposed studies regard simple geometric transformations such as 2D roto-translations or affine transformations. This means that some issues related to the camera model registration are not addressed. Moreover, the resolution of medical data is often quite poor, so using MI in a general case is difficult if no specific adjustments are made. Last key issue in the use of MI is the choice of the optimization strategy to achieve the maximization; the pros and cons of several methods are presented in [MVS99].

Several applications of the registration by MI have been presented in the last years. An interesting method for 3D object tracking has recently been proposed in [PK08] to allow almost real-time tracking of simple template-based objects. Regarding more complex texture registration tasks, a system has been developed to improve texture registration by exploiting 2D-2D and 2D-3D MI maximization [CS07]. However, the optimization is only introduced in 2D-2D registration, while for 2D-3D alignment Viola and Wells's approach is used. Viola and Wells's method was also implemented in [NSI99], where a 3D model with reflectance values (acquired using 3D Scanning) was used. Recently a new solution was proposed in [ZCS09] for the automatic 2D-3D registration. The method projects the surfaces of the 3D model to the 2D normal image space to extract both local geodesic feature descriptors and global spatial information for estimating initial correspondences for 2D-2D and 2D-3D registration. Then the 2D-3D registration is further refined using MI.

3. Algorithm

Mutual Information measures the information shared by two random variables A and B. Mathematically, this can be expressed using entropy or joint probability. Following this interpretation, the Mutual Information $\mathcal{MI}$ between two images I_A and I_B can be defined as:

$$\mathcal{MI}(I_A, I_B) = \sum_{(a,b)} p(a,b) \log\left(\frac{p(a,b)}{p(a)p(b)}\right) \tag{1}$$

where $p(a,b)$ is the joint probability of the event (a,b), $p(a)$ is the probability that a pixel of I_A gets value a and $p(b)$ is the probability that a pixel of I_B gets value b. The joint probability distribution can be estimated easily by evaluating the joint histogram ($\mathcal{H}$) of the two images and then dividing the number of occurrences of each entry by the total number of pixels. A joint histogram is a bi-dimensional histogram made up of $n \times n$ bins; the occurrence (a,b) is associated with the bin (i,j) where $i = \lfloor a/m \rfloor$ and $j = \lfloor b/m \rfloor$ and m is the width of the bin. We use a joint histogram of 256×256 bins.

The image-to-geometry registration problem consists of determining the parameters of the camera model used to project the 3D model onto the image plane. We assume a

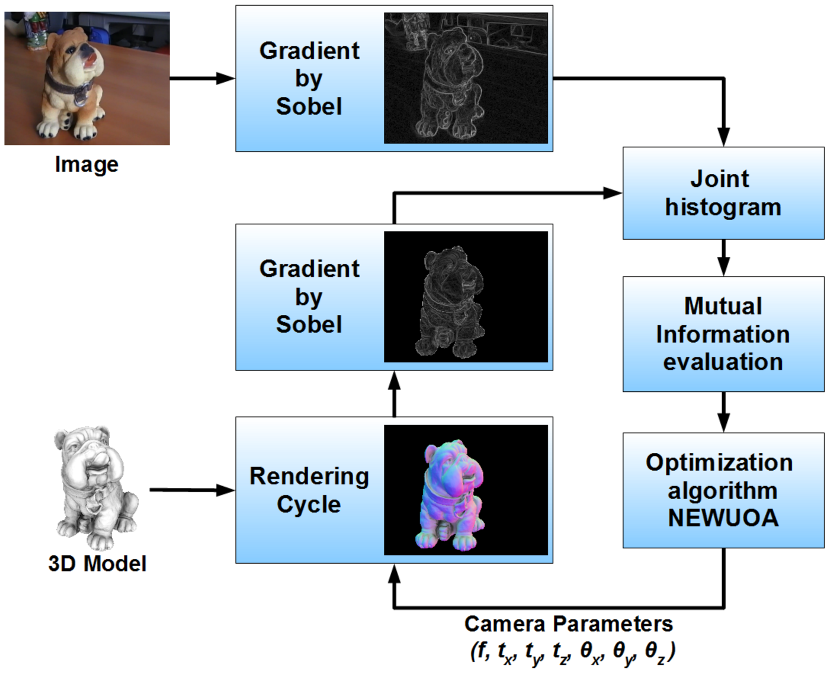

Figure 1: *Algorithm overview*

perspective (or pinhole) camera model where the transformation is described by the projection (intrinsic) parameters plus the position and orientation of the camera in the space (extrinsic parameters).

In this context the registration can be formalized as an optimization problem in a 7D space:

$$\begin{aligned} \mathcal{C}^* &= \arg\max_{\mathcal{C}\in\mathbb{R}^7} \mathcal{MI}(I_A, I_B(\mathcal{C})) \\ \mathcal{C} &= (t_x, t_y, t_z, \theta_x, \theta_y, \theta_z, f) \end{aligned} \tag{2}$$

where f is the focal length, (t_x, t_y, t_z) and $(\theta_x, \theta_y, \theta_z)$ define the position and orientation of the camera, I_A is the pre-processed image to align and I_B is a rendering of the 3D model. Hence, I_B depends on the camera parameters $(\mathcal{C})$. The intrinsic camera parameters, except for the focal length, are assumed as being pre-determined. More specifically, the skew factor is assumed to be zero, the principal point is set as the center of the image and the horizontal and vertical scale factors are assumed to be known from the image resolution and the CCD dimensions.

A sketch of the proposed registration algorithm is given in Figure 1. We generate a rendering of the 3D model with some illumination related properties given the current camera parameters, we compute the gradient map of the rendering and the gradient map of the image and then we evaluate the mutual information of these gradient maps. An iterative optimization algorithm updates the camera parameters and recalculates MI until the registration is achieved. The image gradient is computed by applying the Sobel operator to the images' CIE luminance. More specifically, we minimize the opposite of the MI value. In the computation of the joint histogram we use all the pixels in the rendering viewport but we assign a lower weight to the pixels on the background according the 3D rendering.

The lack of a-priori knowledge about lighting, color and material reflectance information from the model prevents from generating realistic renderings. However, the goal of the rendering cycle is not to generate a photorealistic rendering but to synthesize an image which has a high correlation with the input picture under a wide range of lighting conditions and material appearances. On the other hand, the goal of the gradient is to maximize the shared data between the images discarding all the effects, like specular reflection and subsurface scattering, which we don't take into account in the rendering of the 3D model, and decreasing the influence of the image background, especially non uniform background, on the convergence of the optimization algorithm toward the best camera parameters.

For the rendering of the 3D model we combine the information provided by the ambient occlusion and the normal map, as suggested in [CDPS09]. The ambient occlusion is precalculated and stored in the 3D model as per-vertex color.

During the rendering the value of ambient occlusion is interpolated by Gouraud shading among the triangle vertices. The final color C is obtained by weighting the normal map C_N with the value C_A of the ambient occlusion map (that is normalized between 0.0 and 1.0):

$$\begin{aligned} C_x &= (1-C_A)C_A + C_A C_{Nx} \\ C_y &= (1-C_A)C_A + C_A C_{Ny} \\ C_z &= \sqrt{1-(C_x^2+C_y^2)} \end{aligned} \quad (3)$$

For the iterative optimization we use the algorithm NEWUOA [Pow08]. This algorithm iteratively minimizes a function $F(x), x \in R^n$, by approximating it with a quadric Q. A trust region procedure adjusts the variables looking for the minimum of Q, while new values of the function improve the approximation.

Test	Map	Convergence (% of success) Initial registration errors (pixels) 10	20	30	40	50
HORSE	Norm+Amb	100	95	84	75	34
	Gradient	100	100	91	83	75
SHEPHERD	Norm+Amb	57	70	70	51	46
	Gradient	100	95	88	70	55
DOG	Norm+Amb	18	7	9	3	1
	Gradient	80	88	60	22	12
OMOTONDO	Norm+Amb	50	22	16	8	5
	Gradient	100	49	35	12	4
GARGOYLE	Norm+Amb	100	91	36	10	4
	Gradient	100	98	94	88	86

Table 1: *Convergence tests.*

4. Results

In this section we provide several experimental results in order to evaluate the improving obtained by the proposed algorithm. In particular we compare the results obtained by the computation of the MI on the gradient maps with the results obtained by the framework proposed in [CDPS09], where the MI is computed directly on the model rendering (normals + ambient occlusion) and on the image without computation of the gradient. In our experiment we used five objects with different reflection behaviors (see Figure 2). The photos were acquired with a digital camera with the exception of the DOG example that is a deinterlaced frame of a video acquired with a camcorder. All the photos were scaled to a width of 800 pixel to have a comparable registration error. The corresponding 3D models were generated by 3D scanning using a Konica Minolta VI910 laser scanner.

In order to evaluate the performance for each example we show the shape of the MI function (Figure 2) and the convergence properties of the algorithm (Table 1). To draw the shape of the MI function we evaluated the function in the neighborhood of the optimal solution. The optimal solution was obtained using a semi-automatic tool called TexAlign [FDG*05], based on the selection of 2D-3D correspondences to use in the Tsai's calibration method [Tsa87]. The error in the optimal solution is estimated to be about one pixel. Since the MI function around the aligned position is a function of seven camera parameters, we explored the overall shape around the aligned position with a number of 1D sections, 30 in our case, calculated in random directions in the 7D space; where the MI has a local minimum every section should exhibit the same minimum. In the Figure 2 we show a comparison between the MI function graphs of our algorithm (central column) and of the method proposed in [CDPS09] (left column).

The quality of the MI function is defined by its shape: the important factors are the existence of a well defined minimum and a smooth shape, which permits a wider range of convergence. Analyzing the graphs in Figure 2 we can conclude that the use of the gradient allows to generate a smoother function with better convergence properties near the minimum due to a higher curvature. Especially for the examples with non uniform background (DOG, OMOTONDO, GARGOYLE) the improving is more evident.

In order to test the convergence properties of our algorithm we applied 300 random perturbations to the camera parameters of the aligned images. The parameters were perturbed simultaneously and the maximum allowable registration errors with respect to the reference registration was 50 pixels. For each set of perturbations we measured the percentage of success in convergence of the MI registration algorithms, defined as the number of times that the final registration error is less than 2 pixels with respect to the ground truth obtained by the TexAlign tool. From the data in the Table 1 we can note that the convergence percentage obtained with the gradient maps is higher. Generally for large perturbations, like 40 or 50 pixels, the difference between the convergence rates becomes more marked. Especially in the DOG example, where we use a camcorder, we can note the big improvement introduced by the use of the gradient maps that allow to decrease the influence of the background and of the characteristics of the image acquisition system which can present some image degrading factors, like noise and lens distortions. Other general improvements are obtained in the SHEPHERD example, where the image is acquired with a spotlight, and in the OMOTONDO example, while the HORSE and GARGOYLE examples show the most evident improvement in the convergence rate for large perturbations.

5. Conclusion

In this paper we have proposed an improvement of the image-to-geometry registration by Mutual Information that allows to increase the performance and the quality of the registration. The algorithm is based on the computation of the MI between the gradient of the image and the gradient

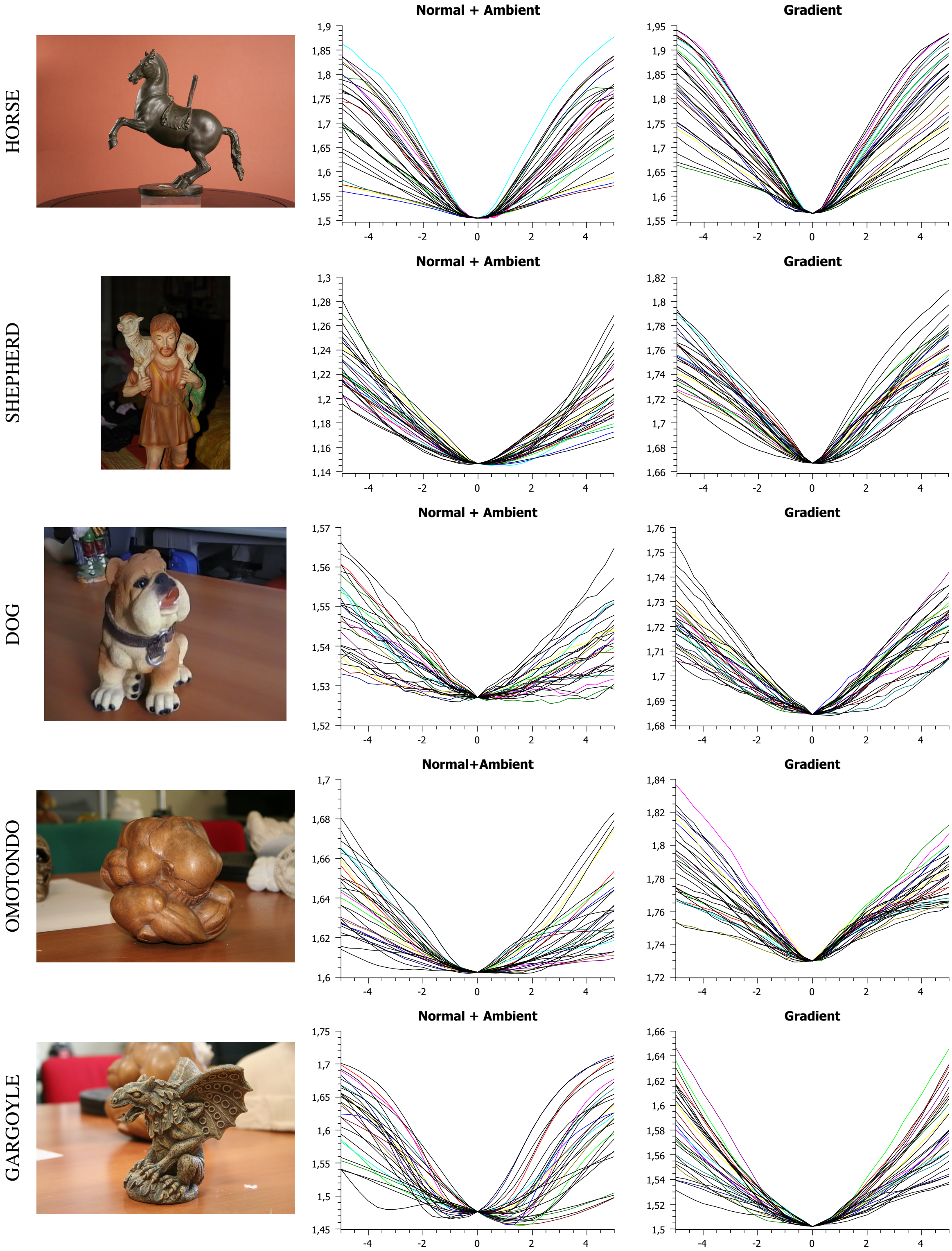

Figure 2: *Images used for the testing and MI function plots: (Central Column) MI function graphs for Normal+Ambient Occlusion rendering; (Right Column) MI function graphs for Gradient Map.*

of the rendering of the model with a combination of normals and ambient occlusion.

Good results were obtained as shown in results section: a better convergence rate even with a big perturbation and a better shape for the MI function that helps the optimization algorithm NEWOUA to converge towards the right camera parameters. These results are encouraging for the development of an automatic global registration algorithm based on our technique.

6. Acknowledgment

This work was partially funded by EU IST IP 3DCOFORM (IST-2008-231809) and Tuscany Regional Project STArT.

References

[CDPS09] CORSINI M., DELLEPIANE M., PONCHIO F., SCOPIGNO R.: Image-to-geometry registration: a mutual information method exploiting illumination-related geometric properties. *Computer Graphics Forum 28*, 7 (2009), 1755–1764. 2, 3, 4

[CS07] CLEJU I., SAUPE D.: Stochastic optimization of multiple texture registration using mutual information. In *Pattern Recognition, 29th DAGM Symposium,* (2007), vol. 4713, pp. 517–526. 2

[FDG*05] FRANKEN T., DELLEPIANE M., GANOVELLI F., CIGNONI P., MONTANI C., SCOPIGNO R.: Minimizing user intervention in registering 2d images to 3d models. *The Visual Computer 21*, 8-10 (sep 2005), 619–628. 4

[LWG97] LEVENTON M. E., WELLS W. M., GRIMSON W. E. L.: Multiple view 2D-3D mutual information registration. In *Image Understanding Workshop* (1997), pp. 625–630. 2

[MCV*97] MAES F., COLLIGNON A., VANDERMEULEN D., MARCHAL G., SUETENS P.: Multimodality image registration by maximization of mutual information. *IEEE Transactions of Medical Imaging 16*, 2 (Apr. 1997), 187–198. 2

[MVS99] MAES F., VANDERMEULEN D., SUETENS P.: Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. *Medical Image Analysis 3*, 4 (1999), 373?386. 2

[NSI99] NISHINO K., SATO Y., IKEUCHI K.: Appearance compression and synthesis based on 3D model for mixed reality. In *ICCV* (1999), pp. 38–45. 2

[PK08] PANIN G., KNOLL A.: Mutual information-based 3d object tracking. *Int. J. Comput. Vision 78*, 1 (2008), 107–118. 2

[PMV00] PLUIM J. P. W., MAINTZ J. B. A., VIERGEVER M. A.: Image registration by maximization of combined mutual information and gradient information. *IEEE Trans. Medical Imaging 19*, 8 (Aug. 2000), 809–814. 2

[PMV03] PLUIM J. P. W., MAINTZ J. B. A., VIERGEVER M. A.: Mutual information based registration of medical images: A survey. *IEEE Transactions of Medical Imaging 22*, 8 (2003), 986–1004. 2

[Pow08] POWELL M. J. D.: Developments of NEWUOA for minimization without derivatives. *IMA Journal of Numerical Analysis 28*, 4 (Oct. 2008), 649–664. 4

[SHH99] STUDHOLME C., HILL D. L. G., HAWKES D. J.: An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition 32*, 1 (Jan. 1999), 71–86. 2

[Tsa87] TSAI R. Y.: A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation 3* (1987), 323–344. 4

[VW97] VIOLA P. A., WELLS W. M.: Alignment by maximization of mutual information. *International Journal of Computer Vision 24*, 2 (Sept. 1997), 137–154. 2

[ZCS09] ZHENG H., CLEJU I., SAUPE D.: Highly-automatic MI based multiple 2D/3D image registration using self-initialized geodesic feature correspondences. In *ACCV (3)* (2009), Zha H., ichiro Taniguchi R., Maybank S. J., (Eds.), vol. 5996 of *Lecture Notes in Computer Science*, Springer, pp. 426–435. 2