Pacific Graphics 2009
S. Lee, D. Lischinski, and Y. Yu
(Guest Editors)

Volume 28 (2009), Number 7

Image-to-Geometry Registration: a Mutual Information Method exploiting Illumination-related Geometric Properties

Massimiliano Corsini, Matteo Dellepiane, Federico Ponchio and Roberto Scopigno

Visual Computing Lab, ISTI - CNR
Via G. Moruzzi 1, 56124, Pisa, Italy
Email: {corsini, dellepiane, ponchio, scopigno}@isti.cnr.it

Abstract

This work concerns a novel study in the field of image-to-geometry registration. Our approach takes inspiration from medical imaging, in particular from multi-modal image registration. Most of the algorithms developed in this domain, where the images to register come from different sensors (CT, X-ray, PET), are based on Mutual Information*, a statistical measure of non-linear correlation between two data sources. The main idea is to use mutual information as a similarity measure between the image to be registered and renderings of the model geometry, in order to drive the registration in an iterative optimization framework. We demonstrate that some illumination-related geometric properties, such as surface normals, ambient occlusion and reflection directions can be used for this purpose. After a comprehensive analysis of such properties we propose a way to combine these sources of information in order to improve the performance of our automatic registration algorithm. The proposed approach can robustly cover a wide range of real cases and can be easily extended.*

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

1. Introduction

Many graphics applications related to color mapping and reflectance properties estimation require to register the images of an object on its virtual representation. Typically, the images are acquired during a photographic campaign, and the geometry of the model is acquired through 3D scanning. There are various ways to accurately recover intrinsic and extrinsic camera parameters, and, in the past years, several algorithms have been proposed for this task. Our approach was inspired from medical imaging, specifically from multi-modal image registration. One of the main problems 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* [PMV03a], 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 our context, we align a 3D model to a given image by using different renderings of the model and a grey-scale version of the input image.

One of the main issues regarding the use of mutual information for 2D/3D registration is to choose a rendering process that correlates well 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. Viola and Wells [VWMW97] proposed using surface normals and image brightness to correlate shading variations on the image with the model surface. Here, we extend their idea by using several types of renderings, such as ambient

occlusion, normal map, reflection map, silhouette map, and combined versions of them. We decide to exploit such maps because the properties of the geometric features used are related to the visual appearance of the model but generating them does not entail knowing the lighting environment of the scene. For example, the use of the ambient occlusion term is based on the assumption that the occluded parts of an object also remain darker in the image to align despite the lighting conditions and texture differences. In order to cover a wide range of real cases, where the validity and strength of these assumptions may vary, we integrated the information provided by these different maps.

The contributions of this paper can be summarized as:

- The outline of a simple and fast general method for image-geometry registration based on mutual information maximization using the renderings of several geometric properties of the model.
- The identification of a set of geometric properties which can be effectively used for this purpose: ambient occlusion terms, normals, reflection directions, and their combination.
- The comparison of the proposed approach with the one of Viola and Wells, which employs only surface normals for the registration.
- An empirical demonstration that this approach is able to cover a wide range of practical cases (i.e different materials and different lighting conditions).

2. Related Work

In this section we summarize the use of mutual information in multi-modal image registration and the issue of automatic texture-geometry registration. For a comprehensive overview of the general topics, please refer to the surveys cited.

Registration using Mutual Information. Image registration is a very popular research topic. Hundreds of different approaches and practical applications have been proposed. While a comprehensive survey of the issue is presented in [ZF03], we will focus on what we believe is 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 [VWMW97] and by Maes et al. [MCV*97]. Since then, several registration methods based on MI have been proposed (see [PMV03b] for a comprehensive overview). Most of these 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.

Another 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]. An interesting method for 3D object tracking allows almost real-time tracking of simple template-based objects has recently been proposed [LWIG97, PK08].

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.

Texture-Geometry Registration. The problem of automatically aligning a set of uncalibrated images to a 3D model is important both for Computer Graphics and Computer Vision. Robust semi-automatic approaches have been proposed [FDG*05] for general cases, but registration can be time consuming. An automatic planning of the number of images and camera positions could lead to good results [MK99] and reduce the importance of registration, but in most cases no information about camera positions is known in advance.

Automatic registration can be achieved by analyzing the image features [NK99] or using the reflectance value acquired during scanning [IOT*07]. Several papers rely on the analysis of the silhouette of the object [BLS92, LHS00]. However, in this case the entire object must be present in each image: this may be a limitation since the advances in acquisition devices mean that very large objects can be represented. In order to add highly detailed texture information to geometric models, it might be necessary to frame only portions of the object in each image.

A recent work for 3D-3D and 2D-3D automatic registration [LSY*06] can be applied in a more general case, but under the assumption that the 3D scene contains clusters of vertical and horizontal lines. Similarly, other approaches such as [LS05] rely on orthogonality constraints, which may not always be present.

3. Background

3.1. Mutual Information

Two widely-used measures of similarity between two images A (the reference) and B (the image to register) are the Sum of Squared Differences (SSD) and Normalized Linear Correlation (NCC). SSD tends to zero whenever the corresponding values of A and B are very close. NCC is able to capture the linear dependence between the two images. For this reason, NCC is less sensitive than SSD to linear transformations of pixel values such as contrast and/or brightness changes. Mutual Information (MI) can be seen as an extension of NCC, i.e. a measure that is able to evaluate the *generic* dependency between two random variables, and it is

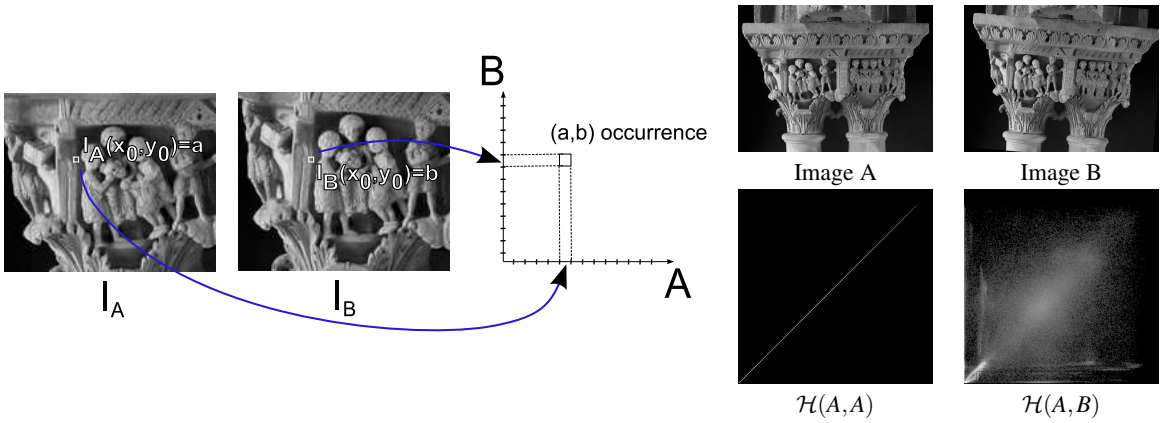


Figure 1: *Joint histogram. (Right) Construction. (Left) An example. The joint histogram of the image A with itself and the joint histogram of the same image with a rotated version of it.*

not limited to linear dependency. Since our aim is to align images whose visual appearance are very different, MI appears to be the most promising mathematical tool for measuring such similarities.

From an information theory viewpoint MI measures the information shared by two random variables A and B. An alternative definition is that MI is the amount of information about B that A contains. Mathematically, this can be expressed using entropy or joint probability. Following the latter interpretation, the mutual information $\mathcal{I}$ between two images I_A and I_B can be defined as:

$$\mathcal{I}(I_A,I_B) = -\sum_{(a,b)} \log p(a,b)\frac{p(a,b)}{p(a)p(b)} \qquad (1)$$

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 $N_A \times N_B$ bins; the occurrence (a,b) is associated with the bin (i,j) where $i = \lfloor a/N_A \rfloor$ and $j = \lfloor b/N_B \rfloor$ (see Figure 1).

For example, in Figure 1 the joint histogram of the perfectly aligned image pair is "very compact" with respect to the one corresponding to the rotated image. Since the two images are identical, the identity relationship between the pixels values is clearly visible in the joint histogram. In general, the higher the dispersion of the distribution of joint events, the lower the value of the MI. In the examples shown: $\mathcal{I}(A,A) = 6.307$ and $\mathcal{I}(A,B) = 1.109$.

4. Registration Algorithm

The image-geometry registration problem consists of determining the parameters of the camera model used to map the 3D model onto the image plane. In this work a perspective (or pinhole) camera model is assumed. According to this model, 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:

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

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 images and the horizontal and vertical scale factors are assumed to be known from the image resolution and the CCD dimensions.

Due to the complexity of finding an optimal solution for (2), the problem was divided into two sub-problems: finding an approximate solution, the *rough alignment*, and refining this solution to an optimal one, the *fine alignment*. The convergence properties and speed of the fine alignment algorithm determine the accuracy required for the solution provided by the rough alignment. For example, a very fast and robust fine alignment could help to solve the rough alignment with a simple greedy approach, by generating a large number of suitable initial views. Moreover, the larger the convergence range of the fine alignment, the "simpler" the

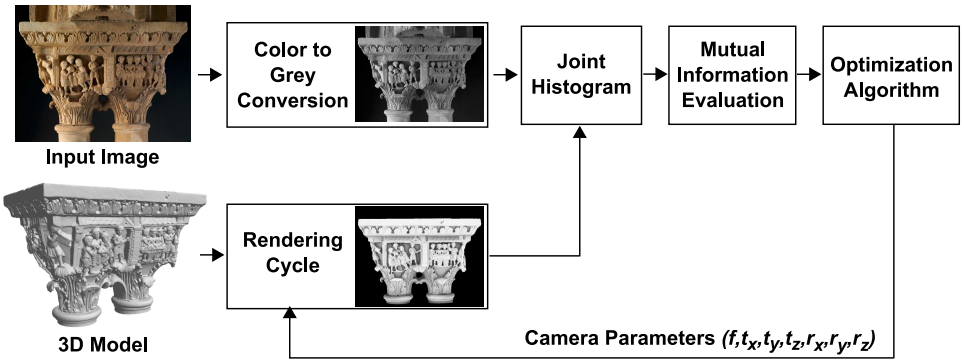


Figure 2: *Fine alignment algorithm overview.*

development of a global framework. For such reasons our goal in this work is to develop a fast and robust fine alignment algorithm with good convergence properties. We are currently still developing a global alignment framework by building on the proposed algorithm. A sketch of the proposed registration algorithm is given in Figure 2: we convert the input image into a greyscale version, generate a rendering of the 3D model given the current camera parameters, and we evaluate the mutual information. An iterative optimization algorithm updates the camera parameters and recalculates MI until the registration is achieved. More specifically, we minimize the opposite of the MI value.

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.

Viola and Wells [VWMW97] demonstrated a good correlation between surface normals and the variation of shading generated by the directional light source, regardless of light direction. In the next section we extend this idea by evaluating several illumination-related geometric features such as surface normals, geometrical occlusions, and reflections directions. In addition, we propose a method to integrate these different sources of information, in order to make the registration algorithm robust and applicable in a wide range of real cases.

The implementation details of the overall algorithm, in particular of the greyscale conversion and of the optimization framework, are given in Section 6.

5. Correlation Analysis of Illumination-related Geometric Properties

5.1. Mutual Information Analysis

In order to analyze the performance of the different geometric features, we evaluated the shape of the MI function in the neighborhood of the optimal solution. This analysis provides useful information about the convergence obtained using the various geometric properties. The ground truth was obtained using a semi-automatic tool called *TexAlign* (developed by Franken et al. [FDG*05]); the error in registration for high resolution images is estimated to be around two pixels. Our test images have a resolution of moreless 4M pixels, we scaled them to a width of about 800 pixels, so we expect our reference registration error to be around one pixel. Since the MI function around the aligned position is a function of seven camera parameters, we explore 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.

The normalization of the variables plays a key role in the context of the minimization: it is extremely important in order to ensure a good convergence for the minimization method. We define the *alignment error* as the mean square distance (in pixels) between the projections through $\mathcal{C}_r$ and $\mathcal{C}$ of all the vertices of the model. This measure is used to normalize the camera parameters: each parameter is scaled so that the increment of 1 unit produces an alignment error of 1 pixel. This normalization also improves the readability of the plots presented in the next Subsections. The X axis represents the distance in the scaled parameter space, while the Y axis represents the values of the mutual information. It is important to underline that these values are not normalized, and they depend on the specific image and 3D model. 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.

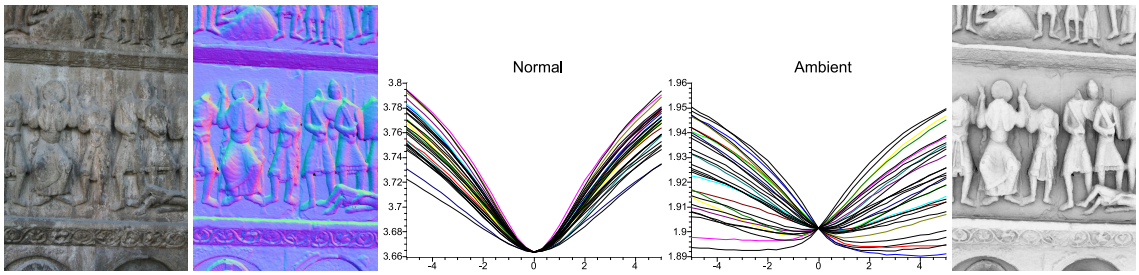


Figure 3: *Ambient vs Normal. As it is possible to notice the ambient occlusion fails to provide reliable functions in this case.*

5.2. The Role of Background

In MI based registration methods, only the "rendered" pixels are usually considered. However, we have noticed a significant positive influence of the background information on the overall performance of the algorithm. We thus decided to evaluate the joint histogram using the whole rendering viewport (and hence the whole image), so not only the part of the image that corresponds to the object's rendered pixels.

We argue that including the background pixels *implicitly* accounts for the silhouette of the 3D object. This is particularly like to be true when the statistics of the image background is very different from the one of the object of interest. The explicit contribution of the shape silhouette in our framework is evaluated in the experimental results section, since many methods for 2D/3D registration rely on silhouette matching.

5.3. Surface Normals

For Lambertian material the reflection depends on the incident angle of the light. In a standard rendering this is computed at each point by the scalar product of light direction and surface normal. As already mentioned, the work of Viola and Wells [VWMW97] is based on this observation and it works reasonably well under the assumption of directional lighting.

Since the normal map has three components (i.e. the coordinates of the normal), we evaluate the MI of each component separately and sum up these values as the final MI value. More specifically, only the x and y components are used since the third one (z) is redundant.

5.4. Ambient Occlusion

Ambient occlusion is used as an illumination-related geometric property since the occluded parts of the model will receive little light irrespectively by the lighting environment of the scene. Ambient occlusion is also view-independent, since its values for each point of the surface depend only on the surrounding geometry. Hence, it can be pre-calculated

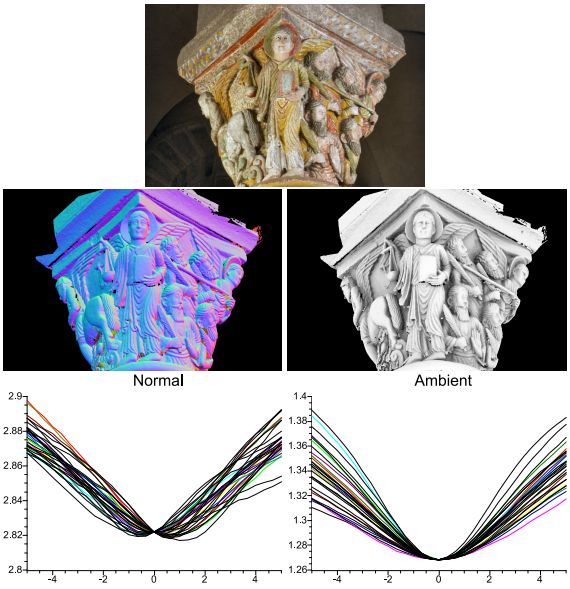


Figure 4: *Ambient vs Normals. (Top) Original image. (Middle) Normal and ambient occlusion maps of the aligned model. (Bottom) Corresponding MI functions plots.*

and stored in order to speed-up the rendering. To our knowledge, this kind of information has never been used in registration algorithms before. There is only another work that demonstrates a correlation between geometry and texture surfaces [MKC*06]; in this work ambient occlusion and other geometry-correlated measures are used for texture transfer.

In our tests, the ambient occlusion term is precalculated with the hardware accelerated technique described in [PG04], and stored as per-vertex color. The ambient occlusion values are mapped such that black corresponds to a completely occluded vertex, while pure white corresponds to a completely exposed surface. The rendering is obtained using OpenGL with Gouraud shading to interpolate the ambient values between vertices.

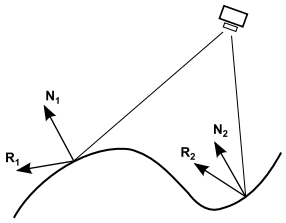


Figure 5: *Reflection map. The resulting map is view-dependent: unlike the normal vector, the reflection vector changes with the position of the camera. This effect becomes more pronounced for large fields of view.*

The utility of ambient occlusion is related to the geometric shape of an object and to the lighting of the original scene; in the case of uniform illumination or complex geometry, it plays a predominant role in the final visual appearance of the object. This is shown by the tests presented in Figure 3 and 4, which show the MI plots between the original image and the normals and ambient occlusion renderings of the registered 3D models. In the first example (Figure 3), regarding a portion of a stone portal, a local minimum is clearly visible in the normals plot. Due to the color and illumination in the image, ambient occlusion has little influence on the lighting of the surface, and the corresponding plot exhibits no minimum at the aligned position. Conversely, the second example (Figure 4) shows that the ambient occlusion plot produces a strong minimum, while normals plot have shallower and shifted minima which could cause the minimization algorithm to stop in an incorrect local minima. This is related to the more complex geometry of the stone capital and uniform illumination, which generates local self-shadowing as the predominant visual effect.

5.5. Reflection Map

The direction of the mirror reflection is another geometric property which is related to the lighting effects (Figure 5). We can evaluate the reflection vectors with a simple shader knowing the viewpoint and the position and normal at each point of the model. We expect a specular material to exhibit highlights that are statistically related to these vectors, hence, we analyzed also the contribution of a rendered reflection map. Our experiments demonstrated that the reflection map performs similarly or slightly better than the normal map, depending on the object's surface reflectivity. In Figure 6 using the reflection map provides slightly better results than using the normals, especially regarding the presence of a good local minimum. This is related to the strong specularity of the material of the object.

5.6. Information Integration

The previous analysis underlines that it is possible to find cases where the performances of a specific type of rendering are better than the others. The ideal solution would be to combine all these contributions in order so as to obtain good results in the general case. This integration is difficult for several reasons. The simple sum of the MI contributions could introduce errors due to the fact that is difficult to normalize MI, and its range of values depends on the object and on the rendering type. Another problem is that objects made of specular material could present a weaker ambient occlusion term, due to the reflected light.
One idea could be to detect the highlights and assign a higher weight to the reflection map in the corresponding parts of the image, but this can worsen the registration performances for certain types of objects, or in particular lighting conditions. For these reasons, the integration of all the contribution is still under investigation, but in the following we demonstrate that the combination of ambient and normal could be done elegantly and providing effective results. Since the ambient and reflection map are integrated in the same way, for the reasons mentioned above, we might expect in certain cases the performance of this measure to be slightly worse than the ambient and normal combination.

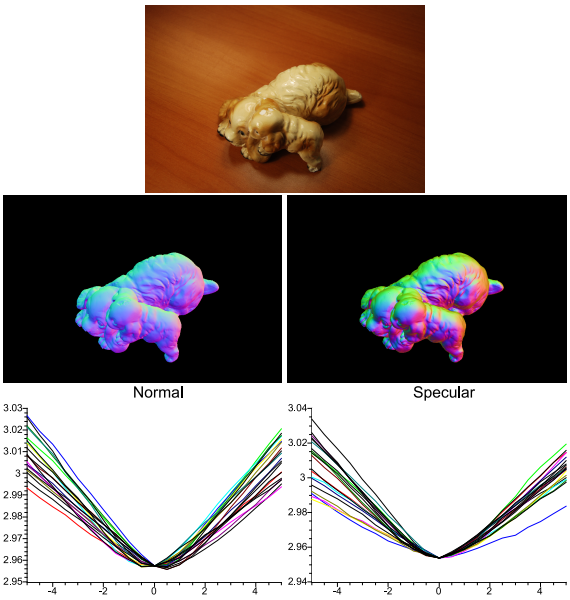
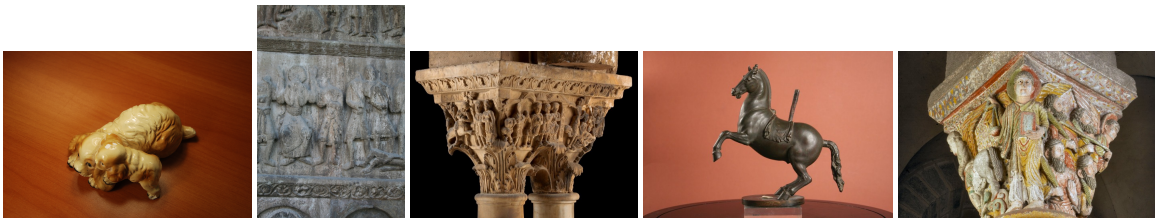


Figure 6: *Reflections vs Normals. (Top) Original image. (Middle) Normal and reflection maps of the aligned model. (Bottom) Corresponding MI functions plots.*

We combine the information provided by the ambient and normal map linearly into a single image 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):

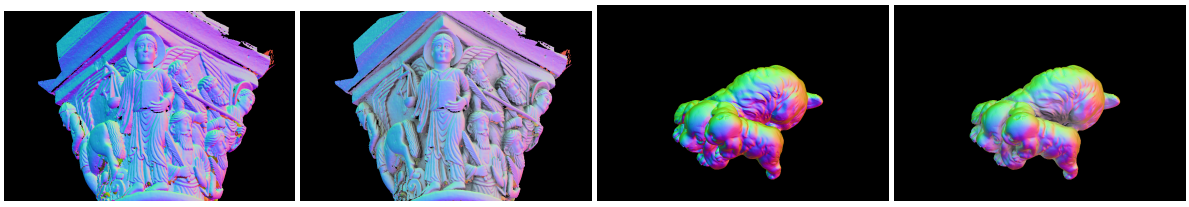
$$C_{\text{comb}} = (1 - C_A)C_A + C_A C_N \tag{3}$$

The resulting rendering strategy integrates the characteristics of both ambient and normal (ambient prevails in darker



Dogs | Bas-relief | Capital1 | Horse | Capital2

Figure 7: *The images used for testing (together with 3D scanned models).*



Normal | Combined of 1st Type (Normal + Ambient) | Reflection | Combined of 2nd Type (Reflection + Ambient)

Figure 8: *Two examples of combined map.*

areas while normals in exposed areas) and should correlate well with a wider range of cases. The same approach is used to integrate the ambient with the reflection map. Figure 8 shows two examples of combined maps. In order to test the proposed renderings map in an extensive way, we applied the registration algorithm to several cases in order to cover different issues related to automatic alignment, such as object material, shape, color, complex visible silhouette, and so on. The test cases are presented in Figure 7.

6. Implementation details

The first step of the alignment algorithm, color-to-grey conversion of the input image, is performed using the formula:

$$Y = 0.3R + 0.59G + 0.11B \tag{4}$$

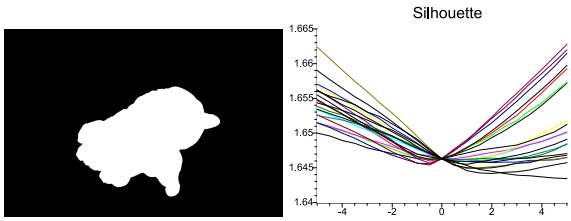
This is the CIE luminance as described in the ITU-R Recommendation BT. 601-4 used in the definition of the NTSC television standard [Poy03]. This conversion provides similar results to other color transformations, such as RGB to HSL or RGB to HSV. Hence, it has been selected for its simplicity. The rendering cycle is implemented using OpenGL with a GLSL shader for each rendering type.

The mutual information is calculated on the CPU, by evaluating the Equation (1). The joint probability distribution $p(a,b)$ is estimated directly from the joint histogram, by dividing the occurrence of the (a,b) event by the total number of pixels. A joint histogram with 128×128 bins was used.

The choice of the optimization algorithm depends on many factors, a very important one is the smoothness of the function to optimize. Algorithms based on estimating derivatives perform poorly on the non-smooth MI functions, especially near the minimum. Other algorithms, based on pattern search, require many iterations to converge, due to the number of dimensions of the problem. Most image registration techniques try to produce smooth MI function using techniques such as Parzen-window [Par62] or antialiasing rendering. We chose a different approach using the recent optimization algorithm NEWUOA, described in [Pow04]. 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. This algorithm entails tuning just one parameter, RHOBEG, which is related to the initial trust region (approximatively 1/10th of the expected initial error), which was set to 2.0 in all of our tests.

7. Experimental Results

In this section we provide several experimental results in order to evaluate the overall performances of the proposed algorithm. First, since many image-geometry registration algorithms are based on silhouette information, we discuss how it contributes to our framework. Then, we evaluate the performance of the proposed algorithm by estimating the range of convergence for each rendering type. Finally, we give some details about the computational times.



Silhouette Map MI function plots

Figure 9: *Silhouette map performances. The silhouette information fails to register the model. The use of the internal visual information is decisive in this case.*

7.1. Notes about silhouette information

One of the main remark regarding the proposed algorithm could be that the main contribution to its convergence comes from the silhouette information, with a minor role of the other geometric features. For certain objects, the presence of an articulated silhouette plays an important role in the convergence range, but it is also important to underline that the use of the geometric features is fundamental in many cases. A first demonstration of this statement is shown in Figure 9 where the MI function plot of the image in Figure 6 is obtained using a "silhouette map". This map is generated by the rendering cycle assigning black color to the background and white color to the 3D object. The plot shows no local minimum, and this is confirmed by the poor alignment performance of the silhouette map (Table 1). Moreover, in certain cases the silhouette of the object is not present, as in Figure 3, where the registration is obtained using normal maps.

Finally, the contribution of the silhouette in our framework is achieved without any image pre-processing, while the state-of-the-art silhouette-based methods often need to extract the contour of the object in the image to register, or to acquire the image of the object on a uniform background.

7.2. Overall performances

In order to test the convergence properties of our algorithm we applied 300 random perturbations of the camera parameters of the aligned images. The parameters were perturbed simultaneously, 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, defined as a final error of less than 4 pixels to take into account the registration error of the ground truth as well. This perturbation can be quite severe; for example, the amount of angular perturbation associated with an error of 20 pixels depends on the shape of the model and the position in the image, but it can be more than 20 degrees for certain models (e.g. CAPITAL1).

Five real objects were used in the convergence tests (see Figure 7). The images of the objects came from different digital cameras. The corresponding 3D models were generated via 3D scanning using a Konica Minolta VI910 laser scanner. Such objects cover several cases; CAPITAL1 is a diffuse object with complex geometry and uniform color, CAPITAL2 is similar to CAPITAL1 but it is colored, DOGS and HORSE are two specular objects (HORSE presents an articulated silhouette w.r.t DOGS) and the BAS-RELIEF is a diffuse object with no silhouette information at all, since only a part of the whole object is depicted in the image. Six rendering types were evaluated: ambient, normal, reflection, silhouette, ambient+normal and ambient+reflection. Table 1 summarizes the convergence results obtained.

Although the performances of ambient rendering for the DOGS image is poor the use of an integrated ambient+normal map gives results that are similar to the normal ones. This shows that the use of this combined measure doesn't affect performances even when ambient or normal maps fail. In this case the role of the silhouette is marginal, as apparent in Figure 9. But even with a low contribution of silhouette information, the convergence rate is good, and gives satisfying results even with perturbations of 30 pixels. In the case of the BAS-RELIEF, the ambient performed poorly due to a local minimum close to the right one which influences the algorithm. On the other hand, the other applicable renderings (even the combined ones) result in good performances. The CAPITAL1 test is another example where, even though the silhouette fails, the other renderings gave very good results. On the contrary, the results obtained with the HORSE test are probably positively influenced by the complex silhouette of the model.
Finally, the CAPITAL2 test showed very good performances for the ambient rendering, while the other individual contributions are not reliable. In this case too, the performances of the combined renderings were very similar to the best single one.

In conclusion, the rate of convergence obtained by the combined renderings was generally very good: the quality of results was always very similar to the best single rendering approach. This indicate that the combined renderings could be an effective solution to for registering images to 3D models in a variety of different cases in terms of geometry, material and lighting conditions. Moreover, good percentages of convergence were obtained with perturbations of 50 pixels or more, which is an important value considering that the viewport resolution used was 800×600.

7.3. Computational time

The registration algorithm has been tested on a Intel Core Duo 2300Mhz with 4GB of RAM and an NVidia 8600 GTS 512MB. The convergence is usually reached in around 100 iterations at 25fps, requiring about 4 seconds, with a 1 million triangle model and a viewport of 800×600 pixels. A GPU implementation could improve performance by computing the histogram directly on the graphics hardware, thus

Test	Map	Convergence (% of success)				
		Initial registration errors (pixels)				
		10	20	30	40	50
DOGS	Ambient	33	23	4	0	0
	Normal	100	94	57	19	18
	Reflection	100	100	68	5	15
	Silhouette	0	4	0	0	0
	Amb+Norm	100	85	52	13	7
	Amb+Refl	100	76	51	5	0
BAS-RELIEF	Ambient	0	0	0	0	0
	Normal	100	100	100	70	53
	Reflection	100	100	100	95	65
	Silhouette	*	*	*	*	*
	Amb+Norm	100	100	100	94	82
	Amb+Refl	100	100	100	95	94
CAPITAL1	Ambient	100	100	100	100	100
	Normal	100	100	100	84	100
	Reflection	100	100	81	73	50
	Silhouette	0	10	23	0	9
	Amb+Norm	100	100	100	89	90
	Amb+Refl	100	100	100	100	81
HORSE	Ambient	100	100	91	50	20
	Normal	100	100	100	78	66
	Reflection	100	100	100	83	85
	Silhouette	100	100	90	75	71
	Amb+Norm	100	100	100	78	62
	Amb+Refl	100	100	95	92	66
CAPITAL2	Ambient	100	100	100	100	100
	Normal	0	11	11	9	11
	Reflection	33	25	15	9	6
	Silhouette	0	0	0	0	0
	Amb+Norm	100	93	89	54	76
	Amb+Refl	100	94	94	66	52

Table 1: *Convergence tests.*

avoiding having to transfer data to the CPU. A GPU calculation of mutual information is given in [SB07].

8. Conclusions and future work

In this paper we have analyzed the potentials of illumination-related geometric properties in a mutual information framework for image-geometry registration. The presented study demonstrates that different sources of information, such as normals, ambient occlusion, and reflection directions can be usefully employed for 2D/3D registration. In particular, we have proposed a way to combine the normal map with the ambient occlusion map to achieve a robust fine alignment algorithm, regardless of the color, material and lighting environment of the object.

Very good results were obtained, as shown in the experimental results section: for perturbations of up to 50 pixels the convergence was obtained within few seconds. This algorithm could be a reliable component in the context of a completely automatic global registration algorithm. Our ongoing preliminary results in this direction are encouraging.

Several interesting research directions can be pointed out. In particular, a better integration of the illumination-related geometric properties to improve the applicability of the algorithm, the correlation analysis of other renderings (e.g. precalculated radiosity map), the use of multi-resolution and mutual information GPU implementation in order to speed up the algorithm.

Acknowledgements

This work was funded by EU IST IP 3DCOFORM project. We would also like to thank the anonymous reviewers for their useful comments.

References

[BLS92] BRUNIE L., LAVALLÉE S., SZELISKI R.: Using force fields derived from 3d distance maps for inferring the attitude of a 3d rigid object. In *ECCV '92: Proceedings of the Second European Conference on Computer Vision* (London, UK, 1992), Springer-Verlag, pp. 670–675.

[CS07] CLEJU I., SAUPE D.: Stochastic optimization of multiple texture registration using mutual information. pp. 517–526.

[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.

[IOT*07] IKEUCHI K., OISHI T., TAKAMATSU J., SAGAWA R., NAKAZAWA A., KURAZUME R., NISHINO K., KAMAKURA M., OKAMOTO Y.: The great buddha project: Digitally archiving, restoring, and analyzing cultural heritage objects. *Int. J. Comput. Vision 75*, 1 (2007), 189–208.

[LHS00] LENSCH H. P. A., HEIDRICH W., SEIDEL H.-P.: Automated texture registration and stitching for real world models. In *PG '00: Proceedings of the 8th Pacific Conference on Computer Graphics and Applications* (Washington, DC, USA, 2000), IEEE Computer Society, p. 317.

[LS05] LIU L., STAMOS I.: Automatic 3d to 2d registration for the photorealistic rendering of urban scenes. *cvpr 2* (2005), 137–143.

[LSY*06] LIU L., STAMOS I., YU G., WOLBERG G., ZOKAI S.: Multiview geometry for texture mapping 2d images onto 3d range data. *Computer Vision and Pattern Recognition 02* (2006), 2293–2300.

[LWIG97] LEVENTON M. E., WELLS W., III, GRIMSON W.: Multiple view 2d-3d mutual information registration. In *In Proc. Image Understanding Workshop* (1997), pp. 625–630.

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

[MK99] MATSUSHITA K., KANEKO T.: Efficient and handy texture mapping on 3d surfaces. *Computer Graphics Forum 18*, 3 (1999), 349–358.

[MKC*06] MERTENS T., KAUTZ J., CHEN J., BEKAERT P., DURAND F.: Texture Transfer Using Geometry Correlation. Akenine-Möller T., Heidrich W., (Eds.), Eurographics Association, pp. 273–284.

[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), 373Ű386.

[NK99] NEUGEBAUER P. J., KLEIN K.: Texturing 3d models of real world objects from multiple unregistered photographic views. *Computer Graphics Forum 18*, 3 (1999), 245–256.

[NSI99] NISHINO K., SATO Y., IKEUCHI K.: Appearance compression and synthesis based on 3d model for mixed reality. vol. 1, pp. 38–45 vol.1.

[Par62] PARZEN E.: On the estimation of a probability density function and the mode. *Annals of Mathematical Stathistics 33* (1962), 1065–1076.

[PG04] PHARR M., GREEN S.: Ambient occlusion. In *GPU Gems*, Fernando R., (Ed.). Addison-Wesley, 2004, pp. 279–292.

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

[PMV03a] PLUIM J., MAINTZ J., VIERGEVER M.: Mutual-information-based registration of medical images: a survey. *Medical Imaging, IEEE Transactions on 22*, 8 (Aug. 2003), 986–1004.

[PMV03b] PLUIM J. P. W., MAINTZ J. B. A., VIERGEVER M. A.: Mutual-information-based registration of medical images: a survey. *Medical Imaging, IEEE Transactions on 22*, 8 (2003), 986–1004.

[Pow04] POWELL M. J. D.: *The NEWUOA software for unconstrained optimization without derivatives*. Tech. rep., Department of Applied Mathematics and Theoretical Physics, Cambridge, England, 2004.

[Poy03] POYNTON C.: *Digital Video and HDTV Algorithms and Interfaces*. Morgan Kaufmann, 2003.

[SB07] SHAMS R., BARNES N.: Speeding up mutual information computation using nvidia cuda hardware. In *DICTA '07* (Washington, DC, USA, 2007), IEEE Computer Society, pp. 555–560.

[VWMW97] VIOLA P., WILLIAM M. WELLS I.: Alignment by maximization of mutual information. *Int. J. Computer Vision 24*, 2 (1997), 137–154.

[ZF03] ZITOVA B., FLUSSER J.: Image registration methods: a survey. *Image and Vision Computing 21*, 11 (October 2003), 977–1000.