EUROGRAPHICS Workshops on Graphics and Cultural Heritage (2014)
R. Klein and P. Santos (Editors)

IBISA: Making Image-Based Identification of Ancient Coins Robust to Lighting Conditions

S. Marchand

Lab-STICC – CNRS, University of Brest, France

Abstract

The IBISA (Image-Based Identification/Search for Archaeology) system manages databases of digital images of archaeological objects, e.g. ancient coins, and allows the user to perform searches by examples. IBISA was designed to help the user decide, from their images, if two objects (coins) are either the same, come from the same matrix (die), share resemblance in style, or are completely different. The system searches for similarities in the databases using a registration method that must be resilient to the viewing conditions. Based on the Fourier transform, it cancels rigid transforms among images. Sub-pixel accuracy can be achieved with a very simple technique. However lighting conditions remain an issue. Fortunately, it is possible to extend this registration method to a light-independent model, considering the elevation or normal maps instead of intensity. The model is also useful for interactive visualization and museography. Although this model registration is now resilient to all viewing conditions, it is not practical in real scenarios where the target is a single image, from which a model can hardly be derived. Finally, a hybrid approach is investigated, with a target image but a model of the reference. It is more realistic, resilient to light conditions, gives excellent results with translations, but shows limitations for rotations.

Categories and Subject Descriptors (according to ACM CCS): I.3.8 [Computer Graphics]: Applications—I.4.3 [Image Processing and Computer Vision]: Enhancement—Registration I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Shading,Shape I.4.9 [Image Processing and Computer Vision]: Applications—

1. Introduction

IBISA (Image-Based Identification/Search for Archaeology) [MDV*09] is an ongoing research project aiming at allowing the user to perform searches by examples in databases of digital images of archaeological objects, e.g. ancient coins. These objects are only required to be quasi flat and produced from matrices, e.g. dies used for striking in the case of coins.

IBISA was designed to help the user decide, from their images, if two objects are either the same, come from the same matrix, share resemblance in style, or are completely different. It uses computer vision methods to make this decision while getting rid of the viewing conditions when searching for similarities in the databases. But before this computation, the effects of the viewing conditions have to be cancelled, such as the centering, orientation, and scale of the photographed objects. Thus the key part of the IBISA system is an intensity-based frequency-domain image registration method, based on phase correlation [RC96], and cancelling these rigid transforms. However, the lighting conditions could remain an issue.

We proposed in [Mar13] to use elevation instead of intensity for the registration. Although this "model registration" (in contrast to image registration) is resilient to the lighting conditions, it requires a shade-from-shading step to get the elevation maps (or at least the normal maps). This is possible when we get access to the object, but can hardly be done reliably when only one image is available, e.g. in a book, which is common in archaeology. Although we are obliged to consider the case of the target being an image (e.g. picture took during excavations), the reference could be the model of an object (e.g. acquired in the museum). This paper aims at investigating this "hybrid" (image-to-model) registration.

The remainder of this paper is organized as follows. Section 2 presents the image registration method, achieving subpixel accuracy. Section 3 introduces the model registration technique based on the previous method. It proves to be resilient to the lighting conditions, but is not practical when the target is a simple image. Thus Section 4 investigates the hybrid case, with a target image but a model of the reference. Finally, Section 5 gives future research directions.

2. Image Registration

The IBISA system initially considers gray-scale images, to get rid of any colorimetric issue. Let us denote by $g(p)$ the intensity of image g at point p. Although determining the similarity between two images g_1 and g_2 is a very complex problem [Gos12], we showed that using the classic inter-correlation factor yields very good results in practice for our purpose [MDV*09,Mar13], provided that the second (target) image has been previously aligned with the first (reference) image. Thus, the core of the IBISA system is a registration method, designed to get rid of viewing conditions. The images are shot with the photographic objective perpendicular to the surface of the (quasi planar) object. The problem is that translation, rotation, and scaling are likely to occur.

Let us now consider some geometric transformation

$$T: \quad \begin{array}{ll} p \mapsto \hat{p}, & \hat{p} = T(p), \\ g \mapsto \hat{g}, & \hat{g}(\hat{p}) = g(p). \end{array} \tag{1}$$

2.1. Translation

A translation of vector $t = (\Delta_x, \Delta_y)$ can be easily expressed using complex numbers with the Cartesian form (where x is the abscissa and y is the ordinate):

$$\begin{array}{rrcl} & \hat{p} & = & p+t \\ \text{i.e.} & \hat{g}(p) & = & g(p-t) \end{array} \tag{2}$$

where $\hat{p} = \hat{x} + j\hat{y}$, $p = x + jy$, and $t = \Delta_x + j\Delta_y$ ($j^2 = -1$). A consequence of (2) is that

$$\hat{G}(\omega) = G(\omega)e^{-j\omega t} \tag{3}$$

where $G = F(g)$, F denoting the Fourier transform. Then the phase correlation is defined by

$$\frac{G(\omega)\hat{G}^*(\omega)}{|G(\omega)||\hat{G}^*(\omega)|} = e^{-j\omega t} \tag{4}$$

which inverse Fourier transform is an impulse located at t.

In theory $|G| = |\hat{G}|$, thus in practice one can approximate $|\hat{G}|$ (which may depend on the lighting conditions, as in Section 4 where we render $\hat{g}$ from its model) by $|G|$ (known), thus $|G||\hat{G}^*| \approx |G|^2$.

In the discrete case, the impulse is in fact a cardinal sine (sinc) centered at t. Knowing this, for each dimension it is possible to estimate the translation with a sub-pixel accuracy (see also [FZB02]). The idea here is to consider (the inverse of) the ratio between the global maximum M of the inverse Fourier transform and the value m of its strongest neighbor. For one dimension, if $t = \lfloor t \rfloor + \delta$, with $|\delta| < 0.5$ (and without loss of generality let us consider the case where $\delta \geq 0$),

$$\begin{array}{rclcl} M & = & \mathrm{sinc}(\delta) & = & \frac{\sin(\pi\delta)}{\pi\delta} \\ m & = & \mathrm{sinc}(\delta-1) & = & \frac{\sin(\pi(\delta-1))}{\pi(\delta-1)} \end{array} \tag{5}$$

and since $\sin(\pi(\delta-1)) = -\sin(\pi\delta)$, we have

$$r = \frac{m}{M} = \frac{\delta}{1-\delta}. \tag{6}$$

In practice, r is measured, then δ can be estimated using

$$\hat{\delta} = \frac{r}{r+1} \tag{7}$$

yielding to an efficient estimate of t with a sub-pixel accuracy. For two dimensions, it is just a matter of applying the above procedure for each dimension, with horizontal or vertical neighbors, since $\mathrm{sinc}(x,y) = \mathrm{sinc}(x)\mathrm{sinc}(y)$.

2.2. Rigid Transforms

Without loss of generality, we can consider that any rigid transform can be represented as a translation followed by a rotation+scaling. Moreover, this rotation+scaling is equivalent to a translation in the log-polar representation. Finally, the full registration algorithm first estimates and inverts the rotation+scaling by finding a translation in the log-polar system (considering the magnitude spectra of the images to ignore the effects of the translation) then estimates and inverts the translation (now free from any rotation or scaling), see [RC96, MDV*09] for details about the method.

2.3. Performances

In order to test the performances of the registration method, we repeated 100 times the random selection of a reference image, the application of a rigid transform also chosen randomly (with uniform choices of the rotation angle in the $[-\pi, +\pi]$ interval, the scaling factor in the $[1/2, 2]$ interval, and sub-pixel translations), and finally the estimation of the parameters. With the classic registration method, the error distributions are roughly uniform, bounded by $1/2$ pixel for each coordinate of the translation as expected, by 0.006 radians for the angle, and by 0.02 for the scale. Using sub-pixel registration, the results greatly improve. The error distributions are zero-mean Gaussians, with standard deviations of $1/8$ pixel for each coordinate of the translation, and 0.002 for the angle and the scale. Moreover, the post-registration similarity factor is very high, close to 1.0 (the maximum).

3. Model Registration

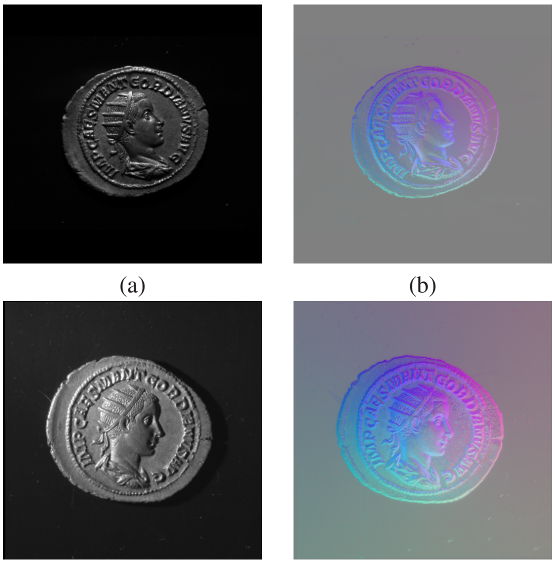
The problem is that the lighting conditions may have a great effect on the similarity factor [Mar13], as seen on Fig. 2 when the light source turns around the object (reference at 0 deg.). To get rid of these conditions, we can consider the elevation or the normals instead of the intensity of the pixels. The same registration method as in Section 2 can still be used, but this time $g(p)$ denotes either the elevation of the surface of the object at point p or the intensity rendered using some model involving the normal vector at this point.

3.1. Model Estimation

Getting the elevation information is possible from several images with fixed object and camera but different light source positions, using a principle close to Reflectance Transformation Imaging (RTI), see [MVSL05]: For each pixel, from the peak in the (interpolated) luminance distribution, we can deduce the normal vector, see [Mar13]. And from the normal map, we can integrate the elevation. In fact, the last step is trickier (see [SSa*12] for a survey). The integration may also introduce arbitrary constants. Fortunately, adding a constant or multiplying by a constant the data should have no impact on the similarity factor. In practice, for now we prefer to consider an articial rendering consisting in taking the peak luminance for each pixel.

3.2. Performances

For the experiments, we used one ancient Roman coin and took 24 pictures with the light source moving around it, using two different cameras (Nikon CoolPix 995 and Canon EOS 5D), with different settings regarding the centering, orientation, and zoom. For both image series, we used the above algorithm to get the normal and peak luminance maps. Then we used the registration method of Section 2, with the peak luminance instead of the intensity of the pixels. The method managed to align the two models, and the final similarity factor was close to 1.0, meaning that the system detected that the coin was the same. The classic image registration failed, the viewing conditions being quite different (see Fig. 1).



(a) (b) (c) (d)

Figure 1: *Picture of the first shooting (left) and estimated normals (right) on top, and similar data for the second shooting (bottom). Whereas the classic image registration fails (because the lighting conditions vary), the model registration succeeds in detecting (b)-(d) as similar. The hybrid registration succeeds with (a)-(b), but fails with (a)-(d) because of the rigid transform, which is not a translation...*

4. Hybrid Registration

The model registration of Section 3 is clearly resilient to the lighting conditions. However, it is not realistic in practice: although a model can be used for the reference, a single image is the only information available for the target. In these conditions, an image-to-model (so-called "hybrid") registration has to be designed. Let us then go back to the fundamentals of Section 2, considering that g (target) is an image, whereas $\hat{g}$ (reference) is also an image but rendered from its model depending on the lighting conditions.

If the model consists of normal vectors, several rendering methods are available, e.g. Lambertian or Ward's. Since our objects are metallic coins, the latter can be more realistic choice, adding a specular contribution to the reflectance function. When Ward's model is used in our experiments, we set its parameters to the gold material (see [NDM05]).

4.1. Lighting Conditions

For the sake of simplicity, let us first assume Lambertian rendering with a constant light vector $\mathbf{L}$, yielding

$$\mathbf{L} \cdot \mathbf{n} = \hat{g} \tag{8}$$

$$\text{i.e. } \; l_x n_x + l_y n_y + l_z n_z = \hat{g} \tag{9}$$

with $\mathbf{L} = (l_x, l_y, l_z)$ being the lighting conditions and $\mathbf{n} = (n_x, n_y, n_z)$ the model consisting of the normals, respectively. Moreover, since $\mathbf{L}$ is constant, applying the Fourier transform to (9) yields

$$l_x N_x + l_y N_y + l_z N_z = \hat{G} \tag{10}$$

where $N_{x,y,z} = F(n_{x,y,z})$. Put in another way, we have

$$\mathbf{L} \cdot \mathbf{N} = \hat{G} \tag{11}$$

Note that, if $\mathbf{n}$ (the model) and $\hat{g}$ (the image) are known, $\mathbf{L}$ (the viewing conditions) can be estimated in the least squares manner, in the spatial (n, $\hat{g}$) or spectral ($\mathbf{N}$, $\hat{G}$) domains.

4.2. Viewing Conditions

In the case of a translation, from (11) and (4) we obtain

$$l_x \frac{G^* \cdot N_x}{|G|^2} + l_y \frac{G^* \cdot N_y}{|G|^2} + l_z \frac{G^* \cdot N_z}{|G|^2} = e^{-j\omega t} \tag{12}$$

and applying the inverse Fourier transform (F^{-1})

$$l_x c_x + l_y c_y + l_z c_z = \delta_t \tag{13}$$

where δ_t is again a Dirac function located at t and c_x, c_y, c_z are known, and given by

$$c_{x,y,z} = F^{-1}\left(\frac{G^* \cdot N_{x,y,z}}{|G|^2}\right). \tag{14}$$

This might be the reason why the effects of the lighting conditions on the estimation of the translation are quite low. The generalization to any rigid transform is more complicated in theory (and also fails in practice, see below). Thus, in practice, we propose an iterative algorithm for image-to-model registration: starting with random initial lighting conditions,

1. render $\hat{g}$ with the current conditions,
2. estimate the viewing conditions (align g on $\hat{g}$),
3. estimate the lighting conditions, using the aligned version of g and the model of $\hat{g}$,
4. update the lighting conditions with the estimation.

4.3. Performances

We ran the same experiments as in Section 2, but this time with the target image rendered from random lighting conditions prior to the random rigid transform. When this transform is limited to a translation, we obtain excellent results within only 2 iterations of the preceding algorithm: the error distributions are not Gaussian anymore, but still bounded by approx. 0.12 pixel (similar to Section 2). The post-registration similarity factor stays close to 1.0 (distribution mode above 0.99). This works with several rendering models: Lambertian, Ward's, and even Polynomial Texture Mapping (PTM), see [MGW01], if the model is changed from normal vectors to polynomial coefficients. And with PTM, the rendered image is really close to the original photographs. However, with a more general rigid transform than a simple translation, the performances are really degraded.

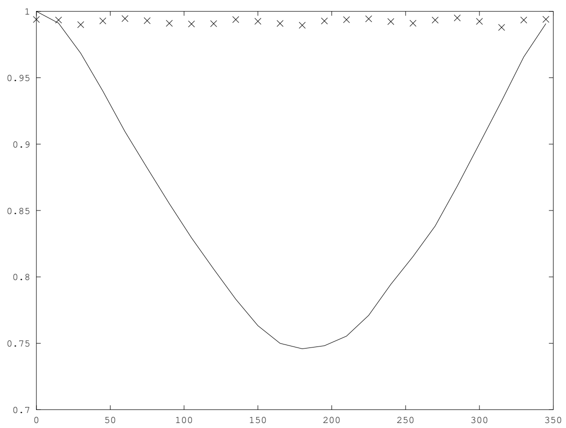


Figure 2: *With classic image registration (line), the similarity factor varies as a function of the light source angle (degrees). With hybrid registration (×), it stays close to* 1.0*.*

5. Conclusion and Future Work

The IBISA project aims at identifying archaeological objects from their images, in a manner resilient to viewing conditions. The key part of the system is an image registration method, now extended to sub-pixel accuracy in a simple way, and able to cancel any rigid transform in a very efficient way, provided that the lighting conditions are constant.

To get rid of the lighting conditions, a radical approach can be to consider a shape-from-shading algorithm to estimate the normal maps from the intensity of the pixels, then use an artificial rendering from this model. Using the registration method with this approach leads to promising results. However, in practice the target is often a single image, from which such model can hardly be derived.

For this reason, a hybrid registration method has been proposed, associated with an algorithm where the viewing and lighting conditions are estimated in sequence. We obtained promising results with different models, provided that the rigid transform is limited to a translation. The goal is now in theory to generalize this hybrid method to any rigid transform, and in practice to build the acquisition device (RTI dome) to get more coin models, for extensive testing.

References

[FZB02] FOROOSH H., ZERUBIA J. B., BERTHOD M.: Extension of phase correlation to subpixel registration. *IEEE Transactions on Image Processing 11*, 3 (March 2002), 188–200. 2

[Gos12] GOSHTASBY A. A.: *Image Registration: Principles, Tools and Methods*. Springer-Verlag London Limited, 2012, ch. Similarity and Dissimilarity Measures, pp. 7–66. 2

[Mar13] MARCHAND S.: IBISA 3D: Image-based identification/search for archaeology using a three-dimensional coin model. In *Archaeology in the Digital Era. Papers from the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA), Southampton, 26–29 March 2012* (2013), Amsterdam University Press, pp. 188–195. URL: http://dare.uva.nl/document/516092. 1, 2, 3

[MDV*09] MARCHAND S., DESBARATS P., VIALARD A., BECHTEL F., AMARA A. B., CICUTTINI B., BOST J.-P., BRESSON A., KONUK K., BEURIVÉ A.: IBISA: Image-Based Identification / Search for Archaeology. In *Proceedings of the International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST)* (St. Julians, Malta, September 2009), The Eurographics Association, pp. 57–60. VAST-STAR, Short and Project Proceedings. 1, 2

[MGW01] MALZBENDER T., GELB D., WOLTERS H.: Polynomial texture maps. In *Proceedings of ACM SIGGRAPH* (Los Angeles, California, USA, August 2001), Computer Graphics Proceedings, Annual Conference Series, pp. 519–528. 4

[MVSL05] MUDGE M., VOUTAZ J.-P., SCHROER C., LUM M.: Reflection transformation imaging and virtual representations of coins from the hospice of the grand st. bernard. In *Proceedings VAST* (Pisa, Italy, November 2005), The Eurographics Association, pp. 29–29. 3

[NDM05] NGAN A., DURAND F., MATUSIK W.: Experimental analysis of BRDF models. In *Proceedings of Eurographics Symposium on Rendering* (Konstanz, Germany, June/July 2005), Bala K., Dupré P., (Eds.), pp. 117–126. 3

[RC96] REDDY B. S., CHATTERJI B. N.: An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Transactions on Image Processing 5*, 8 (August 1996), 1266–1271. 1, 2

[SSa*12] SARACCHINI R. F. V., STOLFI J., AO H. C. G. L., ATKINSON G. A., SMITH M. L.: A robust multi-scale integration method to obtain the depth from gradient maps. *Computer Vision and Image Understanding 116*, 8 (August 2012), 882–895. Elsevier. 3