# Robust Registration of Manuscript Images

Ryan Baumann
Center for Visualization & Virtual Environments
University of Kentucky
Lexington, KY, USA
ryan.baumann@uky.edu

W. Brent Seales
Center for Visualization & Virtual Environments
University of Kentucky
Lexington, KY, USA
seales@uky.edu

## ABSTRACT

In this paper we present an application of image registration techniques to the specific domain of manuscript images. We show the application of this technique to images of the Venetus A, a 10th century manuscript of Homer's *Iliad*. The same algorithm is used to register images of the MS across time (including photographs separated by over a century), as well as across imaging modalities.

## Categories and Subject Descriptors

I.4.3 [**Image Processing and Computer Vision**]: Enhancement—*Registration*

## General Terms

Algorithms

## Keywords

Image registration, manuscript restoration, image warping, multispectral imaging

## 1. INTRODUCTION

In 2007, a team of librarians, scholars, conservators, and photographers under the aegis of the Homer Multitext project traveled to Venice, Italy to undertake the photography and 3D scanning of the Venetus A manuscript at the Biblioteca Nazionale Marciana. Manoscritto marciano Gr.Z.454(=822): Homer I. Ilias, or Venetus A, is a 10th century Byzantine manuscript, perhaps the most important surviving copy of the *Iliad*, and the one on which modern editions are primarily based [7].

The manuscript also contains numerous layers of commentary (*scholia*) from sources which have since been lost, written in a minuscule script. These *scholia* are largely illegible in the only previous reproduction, a facsimile edition published in 1901. Because the highest quality images were desired while incurring the minimal amount of damage to the manuscript pages themselves, 3D scanning was included in the digitization process so that digital flattening techniques[3, 4, 5] could be applied to the newly-acquired images. Due to the size and condition of the text, a high-resolution (39.2 megapixel) digital back was used for photography, and separate macro (close-up) and ultraviolet light shots were taken of specific areas of text.

Because the 3D scanning system acquired untextured 3D models, a procedure to register the 2D photography to the 3D scans was performed periodically during acquisition. This also meant that both the 3D geometry and 2D textures had to be acquired in lock-step for each folio, such that the page was in the same position and physical configuration for each data set.

During one photography session, it was discovered that the sensor on the digital back had been accumulating foreign matter during battery changes, resulting in unacceptable quality for a number of images. Due to time constraints, it was decided that the images would be reshot, without the corresponding 3D geometry capture procedure for each page. This resulted in a number of folios which had two sets of data: a "dirty" image for which we had registered 3D geometry, and a "clean" image taken at a different time (with the page potentially deformed differently) with no corresponding 3D geometry, but to which we wished to apply the digital flattening algorithms.

This was the foundation of our image registration problem – how to obtain a high-quality deformation of the "clean" image such that the text was in the same position as the "dirty" image, allowing us to apply digital flattening using the acquired corresponding 3D geometry. Additionally, because the macro, UV illumination, and 1901 facsimile images were not registered, a robust registration technique would also allow for analysis and comparison of these images across resolution, modality, and time.

## 2. RELATED WORK

### 2.1 Image Registration

Image registration is the process of mapping a *sensed* image into the coordinate system of a *reference* image [17]. The sensed and reference images may differ by viewpoint, modality, illumination, or a number of other parameters. In the aforementioned case, the "clean" image is the sensed image and the "dirty" image is the reference.

Image registration is a long-studied problem with a wide variety of approaches and applications [2, 17]. The registration process itself can usually be broken down into a series of steps, each of which encompasses a variety of problems to be solved: feature detection, feature matching, transform modeling, as well as sampling and warping for the final transformation of the data [17]. For transform modeling, the typical division is between *rigid-body* (encompassing affine transforms such as rotation, translation, and scale) and *non-rigid* (also called non-linear or local) methods. The transform model seected for a given application should reflect some underlying understanding of how the data itself may have deformed between the sensed and reference images.

## 2.2 Registration of Manuscript Images

The Archimedes Palimpsest project has published a number of image sets, including images from Heiberg's 1906 photography of the palimpsest, registered to high-resolution multispectral images taken in August of 2007. According to Doug Emery, this was a completely manual registration process of scaling and warping [16]. Presumably this is an affine transform, and does not account for any non-linear physical distortions which may have occurred between images.

Multispectral imaging of texts and other objects also frequently demonstrate registration problems. Even when the object is imaged in other wavelengths while in the same position, using the same sensor, changes such as filters and lighting can cause angles of refraction to differ and result in an image which is not perfectly registered [13, page 214].

# 3. IMAGE REGISTRATION ALGORITHM

## 3.1 Feature Detection and Matching

The primary unit we are interested in for registration of manuscript images is the text, at the sub-character level. That is, not only should lines match to lines, and characters to characters, but the paleographical features which make up the individual characters should also be matched and maintained. Areas of the manuscript image without writing (such as large unoccupied margins) are not of particular concern for our application.

For feature detection, we use an implementation of the scale-invariant feature transform (SIFT) [12]. Because this is a scale-invariant feature detector, it is robust across changes in region of interest (macro shots) and resolution (facsimile photography). In order to remain orientation-invariant, SIFT utilizes a $4 \times 4$ descriptor across 8 orientations, resulting in a 128-dimensional feature space. Because feature matches are computed by Euclidean distance in this space, we use a $k$d-tree data structure to quickly locate nearest neighbors for a given feature point [1, 11]. The SIFT algorithm's use of local image gradients for feature descriptors performs well at detecting unique features and matches in areas of high-contrast transition, such as between ink and folio.

Due to practical memory limitations, our implementation of the algorithm breaks the high-resolution images into an overlapping grid of sub-image pairs, such that the grid overlap is large enough to capture the majority of feature matches which would otherwise cross a grid boundary. In order to also reduce the necessary overlap size, a preliminary registration pass is performed which uses a RANSAC-pruned [9] sparse set of very strong feature correspondences in downsampled versions of the input to perform a global rigid transform between the sensed and reference images. This allows us to detect keypoints and correspondences using only the memory resources required for a single sub-image pair at a time as opposed to two 39.2 megapixel images. The preliminary registration pass also reduces the amount of computation necessary for registration of images of specific areas of the text, such as macro shots. The results from this sub-image pair processing are maintained in an array which defines matches for the full-resolution images, and are used to determine the mapping between them.

**Figure 1: The fore edge of Venetus A**

## 3.2 Image Warping

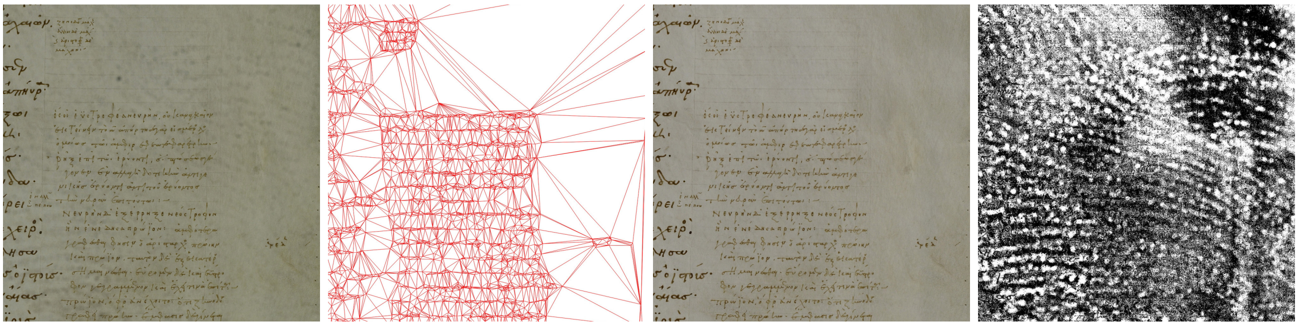
Because the pages of the manuscript are quite cockled and thin (Figure 1), they deform differently each time it is opened to a specific page. In the photographic image, the projective transform of this irregular wrinkling between sets of images of the same folio can be considered as a non-linear warping of the image.

We use a triangle-based deformation to model this warping. In order to construct the triangle mesh, matched feature points from the reference image are inserted into a Delaunay triangulation (an example is shown in Figure 2(b)) [8]. Point correspondences are used to determine the deformation of the mesh in the sensed image. The properties of the Delaunay triangulation method (maximizing the minimum angle) help to avoid long slivers in the final triangulation mesh which are undesirable from a sampling and warping perspective.

If an incorrect feature match is used in constructing this mesh, the corresponding warped triangulation may be a degenerate triangulation containing foldovers, violating our rubber-sheeting model. We avoid such degenerate triangulations by using a greedy algorithm which checks the corresponding mesh at the time of Delaunay triangulation point insertion and removes points which result in invalid joint triangulations. This differs from foldover-free approaches which use multiple time-dependent triangulations [10] or exponential refinement [15].

To perform the warping, the containing triangle for every pixel in the reference is found and its barycentric coordinates are computed. These barycentric coordinates are then used in the matching triangle in the sensed image to find the corresponding pixel coordinates. The barycentric coordinates act to define the points within each triangle in relation to that triangle's centroid.

The computation of barycentric coordinates can be im-



(a) Close-up of reference image (b) Delaunay triangulation of features (c) Close-up of registered sensed image (d) Thresholded absolute difference of images (a) and (c)

**Figure 2: Image registration results**

proved by first computing a barycentric inversion matrix for each triangle with vertices $v$

$$inv = corners^{-1} = \begin{bmatrix} v1_x & v2_x & v3_x \\ v1_y & v2_y & v3_y \\ 1 & 1 & 1 \end{bmatrix}^{-1} \quad (1)$$

For a given pixel $x, y$, its barycentric coordinates $\lambda_{1\ldots3}$ in relation to this triangle are computed by

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = inv * \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2)$$

The original uninverted *corners* matrix for the corresponding triangle is used to convert these barycentric coordinates into warped $x', y'$ coordinates in that triangle

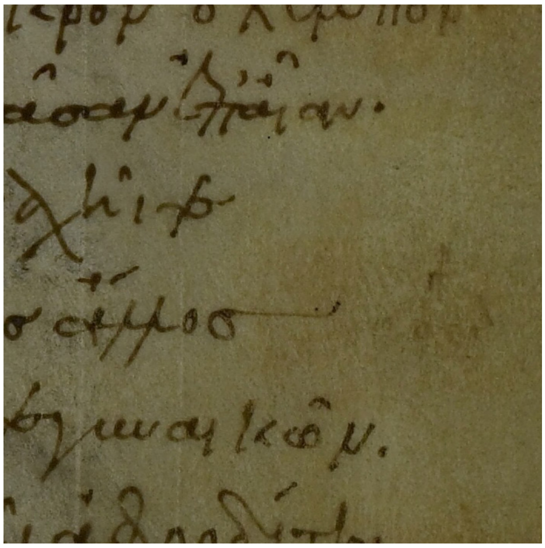
$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = corners * \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} \quad (3)$$

Thus giving an efficient method of using barycentric coordinates to perform the triangulation warp. All warping for the digital flattening process is done using UV texture coordinates for the 3D geometry mesh, allowing the original sensed image to be used (this is useful if the same image with different processing is to be placed on the mesh).

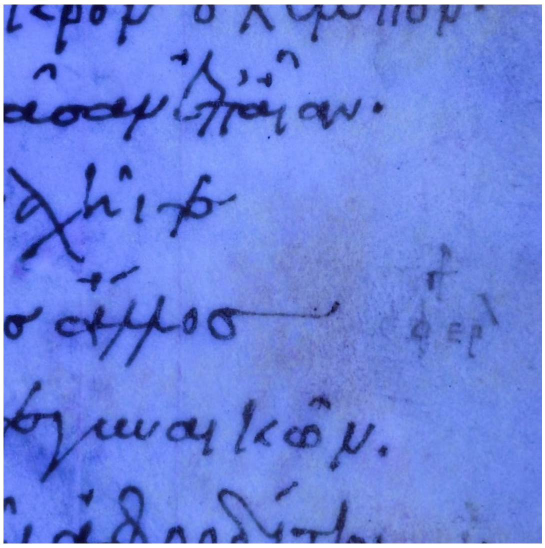
## 4. RESULTS

Results of this registration process for our primary case of "clean" and "dirty" images are shown in Figure 2. As seen in Figure 2(d), the majority of the difference is in foreign matter (i.e. defocused fingerprint residue on the image sensor) and illumination changes. This allows us to accurately apply digital flattening techniques to the text of the "clean" image as desired.

Due to the use of a robust feature detection and matching algorithm as the basis for a non-linear mapping, the registration process also works quite well for images which differ in region of interest and imaging modality (provided there is some portion of mutual information between images). Figure 3 illustrates registration of a macro photograph taken under UV illumination to a photograph of a full folio taken with normal lighting. In this case, seeing the images in the same frame of reference clearly reveals the advantages of UV photography for this area of damaged text.

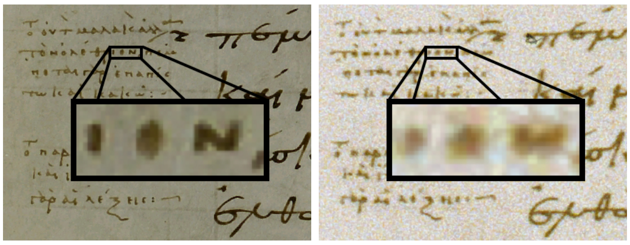


(a) Normal lighting



(b) UV lighting

**Figure 3: Registered ultraviolet macro photograph of Venetus A folio 322v**



(a) 2007 Photograph (b) 1901 Facsimile

**Figure 4: Registered 1901 facsimile of Venetus A folio 24r, with detail inset**

Because of various features of the 1901 facsimile (extreme amounts of noise, poor lens quality leading to radial blurring), it is more difficult to register to modern photography than images taken with similar acquisition setups. The registration algorithm described performs well here in a global sense, but due to a sparser set of good feature matches the non-linear local registration could be refined. Figure 4 shows a cropped example of the registration's performance on an area with *scholia* which is practically unreadable in the 1901 facsimile. Global registration between old and new also reveals areas where writing has been damaged over time.

# 5. CONCLUSIONS AND FUTURE WORK

High-resolution, multispectral digital imaging of important documents is emerging as a standard practice for enabling scholarly analysis of difficult or damaged texts. As imaging techniques improve, documents are revisited and re-imaged, and registration of these images into the same frame of reference for direct comparison can be a powerful tool.

We have presented the design and implementation of a registration algorithm suitable to these kinds of problems, as well as applications of this algorithm to a specific manuscript and imaging project. Although born out of a need for registration between highly similar images, the design has proven applicable for non-linear registration between a variety of image types. In summary, we believe robust, automated registration procedures such as the one described will prove practical and useful to text imaging projects which seek to combine the full range of data available and present it coherently for scholarly study and reference.

We envision several opportunities for future work. Similarity or correlation based measures could be used to refine local registration in the warped image. Multi-resolution registered images could be used as the basis for super-resolution techniques [6], for example of old photographs or facsimiles. Registered images across time or modality could also be used as the basis for change detection algorithms [14]. Integrating such a system so that it could be used quickly and easily during acquisition could help inform the imaging process (such as targeting damaged areas for multispectral or macro imaging) as well as conservation approaches. Finally, because registration warping will always introduce some form of resampling, exposing registration information as machine-actionable, standardized metadata which can be used in conjunction with the raw captured data would be desirable for publication and preservation.

# 6. ACKNOWLEDGMENTS

We would like to thank the Center for Hellenic Studies and the Homer Multitext Project for their collaboration in imaging the Venetus A.

This material is based in part upon work supported by the National Science Foundation under Grant Numbers IIS-0535003, and IIS-0736476. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

# 7. REFERENCES

[1] J. Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 1975.

[2] L. Brown. A survey of image registration techniques. *ACM Computing Surveys (CSUR)*, 1992.

[3] M. Brown and W. Seales. Document restoration using 3d shape. *Proceedings of ICCV'01*, 2001.

[4] M. Brown and W. Seales. Image restoration of arbitrarily warped documents. *Pattern Analysis and Machine Intelligence*, 2004.

[5] M. Brown, M. Sun, R. Yang, L. Yun, and W. Seales. Restoring 2d content from distorted documents. *Pattern Analysis and Machine Intelligence*, 2007.

[6] D. Capel and A. Zisserman. Super-resolution enhancement of text image sequences. *International Conference on Pattern Recognition*, 2000.

[7] C. Dué and M. Ebbott. The Homer multitext project: an introduction. *Digital Humanities*, 2008.

[8] L. P. Chew. Constrained delaunay triangulations. *Algorithmica*, 1989.

[9] M. Fischler and R. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 1981.

[10] K. Fujimura and M. Makarov. Foldover-free image warping. *Graphical Models and Image Processing*, 1998.

[11] P. Grother, G. Candela, and J. Blue. Fast implementations of nearest neighbor classifiers. *Pattern Recognition*, 1997.

[12] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004.

[13] R. Netz and W. Noel. *The Archimedes Codex*. Da Capo Press, 2007.

[14] R. Radke, S. Andra, O. Al-Kofahi, and B. Roysam. Image change detection algorithms: a systematic survey. *Image Processing*, 2005.

[15] A. Saalfeld. Joint triangulations and triangulation maps. *Proceedings of the third annual symposium on Computational Geometry*, 1987.

[16] M. Toth and D. Emery. Infinite possibilities: Archimedes on the web.

[17] B. Zitová and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 2003.