Deformable Part Models for Automatically Georeferencing Historical Map Images

Nicholas Howe
Smith College, nhowe@smith.edu

Jerod Weinman
Grinnell College

John Gouwar
Grinnell College

Aabid Shamji
Grinnell College

Recommended Citation

Nicholas R. Howe, Jerod Weinman, John Gouwar, and Aabid Shamji. 2019. Deformable Part Models for Automatically Georeferencing Historical Map Images. In Proceedings of 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago, IL, USA, November 5–8, 2019 (SIGSPATIAL '19), 4 pages. https://doi.org/10.1145/3347146.3359367

Deformable Part Models for Automatically Georeferencing Historical Map Images

Nicholas R. Howe*
nhowe@smith.edu
Smith College
Northampton, MA

Jerod Weinman*
jerod@acm.org
Grinnell College
Grinnell, IA

John Gouwar
Aabid Shamji
Grinnell College
Grinnell, IA

ABSTRACT

Libraries are digitizing their collections of maps from all eras, generating increasingly large online collections of historical cartographic resources. Aligning such maps to a modern geographic coordinate system greatly increases their utility. This work presents a method for such automatic georeferencing, matching raster image content to GIS vector coordinate data. Given an approximate initial alignment that has already been projected from a spherical geographic coordinate system to a Cartesian map coordinate system, a probabilistic shape-matching scheme determines an optimized match between the GIS contours and ink in the binarized map image. Using an evaluation set of 20 historical maps from states and regions of the U.S., the method reduces average alignment RMSE by 12%.

CCS CONCEPTS

• **Information systems** → **Geographic information systems**; *Content analysis and feature selection*; *Image search*; • **Applied computing** → **Graphics recognition and interpretation**;

KEYWORDS

GIS, georeferencing, historical maps, vector-image alignment

1 INTRODUCTION

Blending views of maps across different eras or modalities creates an information-rich history of geography, politics, and power [17], but this requires precise *georeferencing* (alignment of map images to the Earth's geography) [10]. Because manual alignment takes painstaking effort, this work develops techniques for automatic georeferencing, building on prior work that approximately aligns maps based purely on their *toponyms* (place names) [21]. This work

---

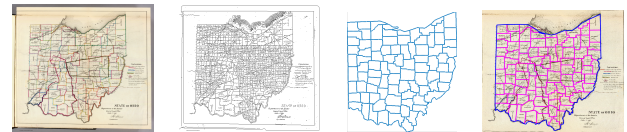*Both authors contributed equally to the paper

Figure 1: Map, skeleton, model and fit. (D0042-1070001 [2]).

complements the toponym approach with a geometric method that matches contours from GIS data—geographical and political boundaries, roadways, etc.—to historical map image contents, using an algorithm developed for word spotting in manuscripts [13].

Historical GIS data is uncommon. However, our model's flexibility allows it to align contemporary GIS data to historical maps. For empirical assessment, we have added extensive ground truth correspondences to an existing dataset of textually annotated historical maps. This allows us to precisely measure the improvement in georeferencing of the previous and newly proposed techniques.

2 RELATED WORK

Aligning data from multiple modalities arises in many fields but particularly in geospatial processing [10]. Many have used road lines to align vectors and geographical images [3, 4, 15]. We can further distinguish between orthoimagery and map images as the alignment target. In the latter category (to which this work belongs), Duan et al. [5] adjust an initial vector-image alignment by searching for consistent local control point deformations that improve the putative alignment to a binarized map image. Weinman's toponym-based methods [20, 21] robustly align with a full affine transform using only map region metadata and a gazetteer.

Although other recent methods have trained CNNs for alignment [18] and interpolation [22], our model requires no training data and is based solely on a geometric consistency score and a simple distance metric. Such deformable part models have successfully been applied to object recognition [6] and word spotting [12].

3 METHODS

We seek to produce a mapping $(x', y') \leftrightarrow (\phi, \lambda)$ that associates a position on the map image with its equivalent geographic coordinate (latitude ϕ and longitude λ) as defined by the North American Datum of 1983 (NAD83). Between these coordinates we calculate other intermediate coordinates: projected Cartesian coordinates (u, v) and initial image coordinates (x, y) estimated via toponyms.

Further details of this work appear in a technical report [14].

3.1 Model Structure & Configuration

To allow for the deformations often observed in historical maps we adapt inkball models [12] to represent shape. These are deformable part models comprising a linked network of control points. Each point can move independently while seeking to retain the original configuration as much as possible. GIS features provide that initial shape: linked control points are placed along boundaries, roads, or other curvilinear features using rasterized GIS data projected at map resolution using the best available estimates. Points that neighbor each other in the rasterized shape image are neighbors in the model, represented as a graph with adjacency matrix N (generally a cyclic graph). The model configuration records the observed spatial offsets between neighbors $\mathbf{t}_{ij} \triangleq \mathbf{v}_i - \mathbf{v}_j$ where $\mathbf{v}_i$ and $\mathbf{v}_j$ are image coordinates along the rasterized GIS contour.

Previous work by Weinman [20] georeferences map images to improve toponym recognition. Candidate words, probabilities, and their image locations are used in a consensus-based estimation that results in coarse parameter estimates. To improve this initial toponym-based alignment, an Expectation-Maximization (EM) algorithm optimizes the overall probability model, selecting the projection family and further tuning parameter estimates [21]. This estimation process yields an affine transform between the projected (u, v) Cartesian map coordinates and the (x, y) image raster coordinates used for the inkball model configuration.

We also propose a novel alternative using inkball geometry alone. For a very crude projection, we set image coordinates $u = k_w \phi$ and $v = k_h \lambda$ (where k_w and k_h scale the structure to occupy 80% of the horizontal and vertical map dimensions) to build an initial model. From this projection we derive m tree-structured models from the original by selecting a seed node and adding nodes one at a time to generate a random spanning tree. Efficient methods then find the optimal fit of each tree-structured model to an image [7]. This gives m configurations for the model control points, which we combine by taking the median position of each. Finally we fit an affine transformation that brings the raw points in (u, v) space as close as possible (in a least-squares sense) to the tree-based models' collective prediction in (x, y) space. These affine-transformed points form the $\{\mathbf{v}_i\}$ that yield a model configuration.

3.2 Model Fitting

To fit an inkball model to a map, we search for a configuration that places control points near center lines of map features and preserves their position relative to neighbors. Deviations from either goal suffer a quadratic penalty. Several steps prepare the raster (RGB) map image for this process: (1) downsample by a factor of 8, (2) convert to grayscale, (3) binarize using Howe's method [11], (4) thin lines to single pixel width skeletons [9].

We model the twin goals as potential functions. Letting $\mathbf{x}_i$ be control point i's putative location on the image grid, the map feature potential is

$$\psi_i (\mathbf{x}_i) \triangleq \exp \left\{ -D (\mathbf{x}_i)^2 \right\}, \tag{1}$$

where D represents the minimum distance from the given point $\mathbf{x}_i$ to the ink, which can be computed efficiently by the distance transform [16]. Neighboring control points have interaction potential

$$\psi_{ij} (\mathbf{x}_i, \mathbf{x}_j) \triangleq \exp \left\{ - \left\| (\mathbf{x}_i - \mathbf{t}_{ij}) - \mathbf{x}_j \right\|^2 \right\}, \tag{2}$$

which increases with the squared distance of $\mathbf{x}_j$ from its expected location $\mathbf{x}_i - \mathbf{t}_{ij}$.

Model fitting then proceeds via a form of message passing. Each control point maintains a record of what has been determined about its 2D location, represented as a grid of beliefs $b_i (\mathbf{x}_i)$ at the same resolution as the map image. Following an initialization step described below, the estimated locations are updated iteratively in turn to take into account information passed to it by its neighbors:

$$b_i^{(t+1)} (\mathbf{x}_i) \propto \psi_i (\mathbf{x}_i) \prod_{j:N_{ij}=1} \max_{\mathbf{x}_j} \psi_{ij} (\mathbf{x}_i, \mathbf{x}_j) \, b_j^{(t)} (\mathbf{x}_j) \tag{3}$$

These values are normalized so they sum to one over the entire grid. Intuitively, the belief update incorporates the local potential ψ_i, preferring locations near the map ink, but also accounts for the neighboring control point locations so as to optimize the interaction potential weighted by the current belief for those locations. We use a linear-time generalized distance transform [8].

After a chosen number of iteration rounds r, the final node positions are set to the most-preferred location:

$$\hat{\mathbf{x}}_i = \arg \max_{\mathbf{x}_i} b_i^{(r)}(\mathbf{x}_i). \tag{4}$$

3.2.1 Belief Initialization. The point configuration used to build the model provides a rough initial position estimate for control points. We use this to construct an initial belief $b_i^{(0)}$ based on the original local potential ψ_i of Equation (1) and a secondary potential ψ_i' preferring locations $\mathbf{x}_i$ near the estimated initial position $\mathbf{v}_i$,

$$\psi_i' (\mathbf{x}_i) \triangleq \exp \left\{ -\kappa \left\| \mathbf{x}_i - \mathbf{v}_i \right\|^2 \right\}, \tag{5}$$

where $\kappa = 0.01$ is an adjustable scale parameter. We combine these potentials to initialize the belief function,

$$b_i^{(0)} (\mathbf{x}_i) \propto \psi_i (\mathbf{x}_i) \psi_i' (\mathbf{x}_i). \tag{6}$$

After initialization, the model positions evolve toward ink, preserving shape, but retaining no further bias for the initial position.

3.2.2 Update & Termination. Control point beliefs are updated incrementally in a fixed order that proceeds initially clockwise around the outer perimeter. Although convergence is not guaranteed, in practice most estimates become stable after just a few iterations. We terminate the computation after four epochs, $r = 4 |C|$, and take each control point position as described in Equation (4).

3.2.3 Densification. Model alignment only fixes the 2D transformation $(x, y) \leftrightarrow (x', y')$ at the control points (where $\mathbf{v}_i \leftrightarrow \hat{\mathbf{x}}_i$). In order to fix the map coordinates of any locations that are not in the model, we must derive a dense 2D correspondence from the sparse control point data. We explore two methods:

Affine : Fit an affine transform to the control point matches using robust linear least squares.

Robust TPS : Robustly fit a thin plate spline to the control point matches [19]. This allows for localized warping.

The affine transform must optimize across all control points, whereas the thin plate spline only enforces local consistency and allows more variation between separated regions. While they differ in flexibility, both of these transforms imply some form of regularization and may not be able to fully express the actual relationship between GIS coordinates and the map's content.
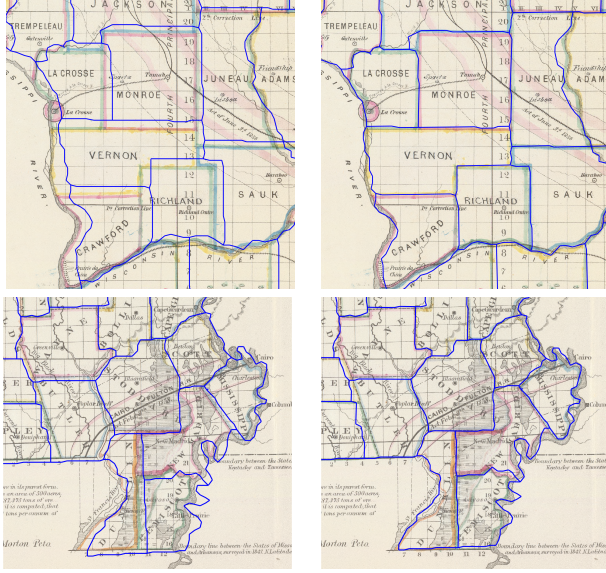
Figure 2: Sample results: starting toponym (left) and final TPS alignments (right). (D0042-1070005, D0042-1070009 [2])

4 EXPERIMENTS

This section presents the results of testing the described methods on a collection of 20 historical maps. For comparability, we use the same set as in prior work [21]. The main results use toponym-based configurations from county or state boundaries, with supplemental experiments using roadways and tree-derived configurations.

4.1 Data

U.S. boundaries for shape models come from the U.S. Census Bureau Cartographic Boundary Files.[1] Highway paths come from the U.S. Geological Survey National Transportation Dataset (NTD).[2]

Our benchmark consists of map images previously used for testing toponym text recognition [21]. The 20 maps used in this benchmark data set range from years 1866–1927 and are drawn from seven atlases [1]. For this work, we add ground truth correspondences between geographical coordinates (latitude and longitude) and pixel coordinates (row and column) in the map images.[3] Additional map details appear in the technical report [14].

4.2 Results

Alignments found by the proposed technique are more accurate than those of prior work. Figure 2 shows initial alignments with slight systematic errors that are reduced by our model.

The degree of improvement depends upon the densification method. Although both the thin plate spline (TPS) and affine methods work from the same control point correspondences, their performance varies. TPS fits can reduce error in local areas where the global, rigid affine fit cannot. Conversely, when some control points

Table 1: Alignment results of individual maps, grouped by atlas. GT is best (least-squares) affine fit to the ground truth correspondences after projection. SAC is the initial toponym-based MLESAC affine fit [20], and EM is the adjusted affine fit [21]. Others are the model described in this work with two densification (cf. Section 3.2.3) strategies and GIS layers indicated.

| Map | RMSE (px) | | | | | | |
| Id. | GT | SAC | EM | State, County | | State | |
				Aff	TPS	Aff	TPS
a	40.7	95.5	68.4	*66.5	***61.3**	-	-
b	84.2	206.1	100.3	94.6	93.3	**90.8**	91.3
c	47.7	297.0	99.1	89.8	81.8	**63.8**	64.8
d	45.0	79.4	52.6	49.0	**33.8**	83.9	48.7
e	25.4	106.6	50.4	38.7	**32.5**	56.5	36.2
f	85.0	238.1	97.1	74.8	74.5	**69.8**	94.2
g	36.5	101.8	49.7	37.2	**34.0**	39.7	36.8
h	33.4	66.9	40.1	33.5	**29.4**	36.3	38.3
i	41.1	69.3	45.5	43.5	**42.8**	47.6	45.5
j	16.9	62.6	55.5	**63.6	****56.7**	40.91	35.57
k	39.0	54.5	44.6	40.0	41.9	**39.3**	40.3
l	45.3	54.5	46.8	46.1	50.5	**45.2**	48.8
m	37.8	39.7	**38.1**	41.7	53.1	39.0	41.8
n	48.1	59.7	52.0	64.1	51.4	51.1	**50.5**
o	27.0	71.2	38.1	60.7	43.1	**34.7**	37.6
p	48.0	50.5	51.5	52.8	59.8	**50.1**	51.0
q	26.1	33.6	28.8	**27.5**	27.7	27.7	27.7
r	26.0	34.2	28.7	35.5	**26.5**	68.2	27.4
s	25.1	43.4	34.9	27.7	**26.2**	31.5	31.7
t	27.4	61.1	58.0	29.0	**27.8**	32.7	32.1
Avg	40.28	91.29	54.01	50.82	**47.4**	-	-

*Meta-data indicates sub-state regional map; model uses only GIS county boundaries
**Meta-data indicates multi-state map; model uses only GIS state boundaries

are mismatched, the global affine fit is more likely to successfully ignore outliers than a local method like TPS (cf. Section 4.2.2).

Table 1 shows detailed results. The overall outcomes shown for the affine and TPS densifications are significantly better than Weinman's toponym-based fits [21] for RMSE in both kilometers and pixels ($p < 0.011$).[4] Overall results show Weinman's adjustments [21] reduce RMSE (in pixels) by 41% over the prior technique [20], and the approach in this work nets a further 12% reduction.

4.2.1 Model Sensitivity. Geographical features change over time and modern data may not match historical map features. Attempts to find non-existent matches can lead to incorrect alignments.

For example, maps from atlas D0117 do not include county boundaries; Figure 3 shows how control points on a county boundary model match a road instead, curving the straight county lines southward. Table 2 compares models using state boundaries and major roads to county boundaries on the D0117 road atlas maps. Despite the maps' deviations from modern road data, the resulting TPS fits improve significantly ($p < 0.035$), reducing RMSE (in kilometers) by 27% to a rate even lower than the best affine fit to ground truth.

Figure 3: Importance of the right model: TPS fit with county boundaries (left) versus roads (right). (D0117-5755025 [2] original map ©Rand McNally 1927)

Figure 4: Outlier result: TPS fit with toponym (left) and tree-based (right) configurations. (D0042-1070004 [2])

Table 2: Alignment results using state boundaries and roads.

Map Id.	GT	EM	State,County		State		State,Roads	
			Aff	TPS	Aff	TPS	Aff	TPS
l	3.04	3.14	3.13	3.25	3.02	3.52	3.00	**2.25**
m	3.18	3.20	3.55	4.66	3.31	3.66	3.51	**2.57**
n	5.83	6.26	7.16	6.63	6.14	6.11	8.10	**5.02**
o	3.12	4.41	6.98	5.27	4.00	4.27	4.09	**3.99**
p	6.88	7.31	7.99	8.37	7.12	7.34	7.34	**7.00**
Avg.	4.41	4.86	5.76	5.64	4.72	4.98	5.21	**4.17**

4.2.2 Tree-based Configurations. Table 3 shows the performance of models built from raw latitude/longitude data using $m = 5$ tree fits (cf. Section 3.1) rather than toponym configurations. The technique relies on a crude scaling heuristic, so it fails on about half the attempts. Nevertheless it produces reasonable results on the rest, and for certain maps it achieves the best results of any method. Figure 4 shows the most salient case: a dearth of toponyms in Michigan's upper peninsula leads the toponym configuration astray, but the tree-based configuration matches well using the shape.

5 CONCLUSION

The methods proposed herein significantly improve the accuracy of map alignment as compared to prior work. They are most effective when the models used correspond to the actual map contents. Future work should explore techniques for post-alignment verification so that bad matches can be identified and improved.

REFERENCES

[1] L. Boateng Asante, D. Cambronero Sanchez, R. Chande, D. Gumm, and J. Weinman. 2017. *A Data Set of Annotated Historical Maps*. Technical Report. Grinnell College. https://doi.org/11084/19349
[2] Cartography Associates. [n. d.]. David Rumsey Map Collection. ([n. d.]). http://www.davidrumsey.com
[3] C.-C. Chen, C. A. Knoblock, and C. Shahabi. 2008. Automatically and accurately conflating raster maps with orthoimagery. *GeoInformatica* 12, 3 (2008), 377–410.
[4] Y.-Y. Chiang, C. A. Knoblock, C. Shahabi, and C.-C. Chen. 2009. Automatic and accurate extraction of road intersections from raster maps. *GeoInformatica* 13, 2

Table 3: Alignment results comparing model configurations: toponym-based versus tree-based. Results shown only for tree-based fits that perform no worse than SAC.

Map Id.	GT	SAC	EM	Toponym		Tree	
				Aff	TPS	Aff	TPS
b	10.40	27.41	12.51	11.61	**11.27**	15.41	12.95
c	5.78	41.85	12.56	11.74	**11.29**	24.00	20.87
d	4.02	6.96	4.72	4.42	**2.99**	3.76	3.07
e	2.50	10.59	5.18	4.47	**2.82**	3.29	2.87
g	3.71	10.31	5.13	3.81	**3.61**	4.79	3.96
h	3.41	6.66	4.11	3.44	**3.03**	5.66	5.70
i	4.05	6.91	4.50	**4.29**	4.30	6.66	6.61
k	6.70	9.31	7.48	**6.71**	6.95	6.70	6.83
q	2.27	2.97	2.51	**2.40**	2.41	2.72	2.44
r	2.08	2.77	2.30	2.48	**2.11**	2.37	2.29
s	2.15	3.74	2.98	2.37	**2.28**	3.34	2.51
t	2.51	5.49	5.45	2.68	**2.64**	4.48	3.28

(2009), 121–157.
[5] W. Duan, Y.-Y. Chiang, C. A. Knoblock, V. Jain, D. Feldman, J. H. Uhl, and S. Leyk. 2017. Automatic Alignment of Geographic Features in Contemporary Vector Data and Historical Maps. In *Proc. Workshop on Artificial Intelligence and Deep Learning for Geographic Knowledge Discovery (GeoAI '17)*. 45–54.
[6] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. 2010. Object Detection with Discriminatively Trained Part Based Models. *IEEE Trans. PAMI* 32, 9 (2010), 1627–1645.
[7] P. Felzenszwalb and D. Huttenlocher. 2005. Pictorial Structures for Object Recognition. *Intl. J. Comp. Vis.* 61, 1 (2005), 55–79.
[8] P. F. Felzenszwalb and D. P. Huttenlocher. 2012. Distance Transforms of Sampled Functions. *Theory of Computing* 8, 1 (2012), 415–428.
[9] Z. Guo and R. Hall. 1989. Parallel thinning with two-subiteration algorithms. *Comm. of the ACM* 32, 3 (1989), 359–373.
[10] A. Hackeloeer, K. Klasing, J. M. Krisp, and L. Meng. 2014. Georeferencing: a review of methods and applications. *Annals of GIS* 20, 1 (2014), 61–69.
[11] N. Howe. 2013. Document Binarization with Automatic Parameter Tuning. *Intl. J. Doc. Analysis & Recog.* 13, 3 (September 2013), 247–258. DOI: 10.1007/s10032-012-0192-x.
[12] N. Howe. 2013. Part-Structured Inkball Models for One-Shot Handwritten Word Spotting. In *Proc. ICDAR*.
[13] N. Howe and J. W. Chung. 2019. Symmetric Inkball Alignment with Loopy Belief Propagation. In *Proc. ICDAR*.
[14] N. Howe, J. Weinman, J. Gouwar, and A. Shamji. 2019. *Details of Deformable Part Models for Automatically Georeferencing Historical Map Images*. Technical Report. Smith College. https://doi.org/10.35482/csc.001.2019
[15] Y. Li. 2009. *An automated system for image-to-vector georeferencing*. Ph.D. Dissertation. University of Texas, Dallas, TX.
[16] C. R. Maurer, R. Qi, and V. Raghavan. 2003. A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions. *IEEE Trans. PAMI* 25, 2 (2003), 265–270.
[17] D. Rumsey and M. Williams. 2002. Historical Maps in GIS. In *Past Time, Past Place: GIS for History*, A. K. Knowles (Ed.). ESRI Press, Redlands, California, 1–18.
[18] M. Simonovsky, B. Gutiérrez-Becker, D. Mateus, N. Navab, and N. Komodakis. 2016. A deep metric for multimodal registration. In *Proc. MICCAI*. 10–18.
[19] R. B. Tennakoon, A. Bab-Hadiashar, D. Suter, and Z. Cao. 2013. Robust data modelling using thin plate splines. In *Proc. Intl. Conf. DICTA*.
[20] J. Weinman. 2013. Toponym Recognition in Historical Maps by Gazetteer Alignment. In *Proc. ICDAR*. 1044–1048.
[21] J. Weinman. 2017. Geographic and Style Models for Historical Map Alignment and Toponym Recognition. In *Proc. ICDAR*.
[22] I. Yilmaz and M. Gullu. 2012. Georeferencing of historical maps using back propagation artificial neural network. *Exp. Techniques* 36, 5 (2012), 15–19.

ACKNOWLEDGMENTS