

UC Santa Barbara

GIS Core Curriculum for Technical Programs (1997-1999)

Title

Unit 11: Registration and Conflation

Permalink

<https://escholarship.org/uc/item/8cr9r7m2>

Authors

Unit 11, CCTP
Sutton, Paul

Publication Date

1998

Peer reviewed

UNIT 11: REGISTRATION AND CONFLATION

Written by Paul Sutton, University of California, Santa Barbara

Context

Registration and Conflation are two related procedures in which two or more geographic datasets are combined, compared, or merged. *Registration* is the means by which one or more of the datasets are mathematically manipulated to a common coordinate system. *Conflation* is a less well defined term that is associated with the many issues and problems associated with integrating two or more datasets that cover the same portion of the earth.

Example Application

The U.S. census bureau is interested in determining if nighttime satellite imagery derived from the Defense Meteorological Satellite Programs Operational Linescan System (DMSP OLS) is a good proxy measure of urban extent. You are provided with a DMSP OLS nighttime image of the continental U.S. and a vector coverage of metropolitan statistical areas (MSAs) as defined by the census. You will perform the following:

1. Identify the projections of both the vector and raster datasets.
 - a. If one or more is not projected you will rubber-sheet them together and if both are not projected rubbersheet the co-registered datasets to a projected dataset of the continental United States.
 - b. If both of the datasets are in a projection, reproject one into the others projection

Note: this is the preferred route particularly for areal extents of this size.

2. Next, you will do a raster to vector data model conversion of the image to get polygons that represent the urban clusters in the image. This will most likely involve some evaluation of threshold values of the DMSP pixels and perhaps something like the REGIONGROUP command of Arc/INFOs GRID module. This

will be followed by something like the GRIDPOLY command.

- 3. Next you will intersect the polygon coverage derived from the DMSP image to the vector coverage of the MSAs and evaluate the overlap. Various mathematical comparisons can be made here.**

Learning Outcomes

The following list describes the expected skills which students should master for each level of training, i.e. Awareness/Competency/Mastery.

Awareness:

A student should know that geographic datasets come in various data models and structures that do not necessarily lend themselves to comparison and or combination.

Competency:

A student should be able to re-project and/or rubber-sheet any spatially referenced dataset to another one. S/he should also know how to decide which method (rubber-sheeting or re-projecting) is appropriate for these given datasets by examining their metadata. This module covers a broad array of issues including data model conversions, scale aggregations, and other GIS manipulation and analysis functions.

Mastery:

Conflation is such a large topic that mastery can only be achieved through experience. Conflation could be described as all the issues associated with the benefits a GIS creates by allowing for the interface of co-spatially extent geographic datasets. In any case, mastery clearly requires that the student have an appreciation of the issues associated with the various data model/structures used by GISs. It also requires that the student understand the mathematical differences between rubber-sheeting and re-projecting.

Preparatory Units

Recommended:

Unit 7 - Using and Interpreting Metadata

Unit 9 - Converting Digital Spatial Data

Unit 10 - Projecting Data

Complementary:

Awareness

Registration: Inherent in any digital spatial data is a mathematical structure that locates where the data is relative to itself. For a point coverage these may be X and Y coordinates that may or may not be geo-referenced (See Figure 1 below).

Figure 1

It is read by a GIS which represents this data could be as simple as one line of numbers separated by commas:

1, 4. 4, 6. 1, 2, 6.6, 5.8, .. 11, 2. 7, 1. 6, 12, 4.8, 1.7

The display and manipulation of a file like the one above is what the GIS handles. Image files based on pixels or raster cells rarely even have coordinates, they are simply a series of numbers representing the values of the pixels. A header file says something like there are 20 rows and 30 columns of data in this image the following strings of numbers is the values of these pixels starting at the upper right of the image and reading across etc.

Suppose the two point coverages above are UFO sightings on the surface of the earth. Assume the first coverage merely contains the relative locations of the sightings with respect to one another on an arbitrary coordinate system. Suppose the second coverage is actually geo-referenced in the sense that the coordinates of the 12 points are the actual longitude and latitude of the sightings. A brief inspection of the relationship should produce the simple transformation equations between one coverage and the other:

$$X \rightarrow X : 10 * X ; Y \rightarrow Y : (10 * Y) 30$$

Registration is an issue because these transformation equations are rarely this simple to obtain; however, the identification of these transformation functions is the crucial element of the registration procedure. These transformation equations can be applied to point coverages (as in the above example), line coverages, polygon coverages, and images. In the case of line and polygon coverages the transformation equations are applied to the points that define the lines and/or polygons and the topology of the coverage is used to rebuild the coverage in the new transformed space. The figure below shows a slightly more complex transformation of several points from one coordinate system to another. This kind of transformation can be accomplished by identifying the transformation equations and sending the coordinates through these equations to produce new coordinates.

Figure 2

In the case of images another issue comes into play because images are not really point data in the strict sense. The pixels of an image have a value that is representative of an area the size of a pixel. Determining the values of the pixels in the transformed image (from reprojection or rubbersheeting) raises the issue of resampling. The registration procedure usually involves the stretching, bending, or warping of the original space to match a geo-referenced space. The figure below shows a raster or pixel based image of the same area in the

two "projections" shown above. Note how the raster data structure influences the shape of the representation of the data. When points, lines, and polygons are transformed via reprojection or rubbersheeting the attributes are carried directly to the new representation. With pixels or raster cells the values have to be interpolated as described in the figure below. Be aware that the measurement scale (Nominal, Ordinal, Interval, or Ratio) of your image will have an influence on which of these resampling methods you will want to use.

[Figure 3](#)

Competency

Ground Control Points and Identifying the Transformation Equations: The means by which the transformation equations are obtained for rubbersheeting is basically an ordinary least squares regression on the coordinates of "Ground Control Points" that are selected from each image. The figure below shows the regression results and plots of the X_{old} vs X_{new} and Y_{old} vs. Y_{new} coordinates of the simple point coverages shown in Fig 1.

[Figure 4](#)

The plots above are perfect fits. This rarely happens because of error in the selection of ground control points, limitations imposed by the spatial resolution of the data, and other error. Also, the relationship between the coordinates is often non-linear, and sometimes uses both the old X and old Y coordinate.

Selection of ground control points can be quite difficult. Ground control points are points that are in the same location in both datasets. Usually they are chosen interactively in which both datasets are displayed and the user clicks on a location in one image and then the same corresponding location in the second image. Typically ground control points are easily identifiable features such as major road intersections, unique land-water boundaries, etc. The software package usually automatically stores the coordinates of the points chosen and writes them to a separate file. The selection of ground control points must be done carefully and the number of points selected influences the types of curve fitting you can perform. (i.e. you need at least two points for a line, three points for a simple quadratic, etc.) Typically you want ground control points to be spatially scattered widely over the datasets you are co-registering and you want a lot more points than is minimally necessary to fit the polynomial you think you need to use.

The key thing to understand about rubber-sheeting and selecting ground control points is the mathematical limitations of Ordinary Least Square Polynomial fitting. If you select N ground control points you will have N pairs of (X, Y) coordinates: e.g. if $N = 3$ youd have: (X_1, Y_1) , (X_2, Y_2) , (X_3, Y_3) and their corresponding coordinates in the second dataset: (X_1, Y_1) , (X_2, Y_2) , (X_3, Y_3) . The ordinary least squares procedure is limited to finding polynomial equations of various orders. As the order increases the number of parameters to be found (and consequently the number of control points needed to estimate them) increases dramatically.

The following shows the types of polynomials that are solved for and their corresponding order (The letters A thru T represent the parameters to be estimated):

First Order:

$X = A * X + B * Y + C$ This is a classic 6 parameter

$Y = D * X + E * Y + F$ Affine transformation

Second Order:

$X = A * X + B * X^2 + C * Y + D * Y^2 + E * XY + F$

$Y = G * X + H * X^2 + I * Y + J * Y^2 + K * XY + L$

Third Order:

$X = A * X + B * X^2 + C * X^3 + D * Y + E * Y^2 + F * Y^3 + G * XY^2 + H * X^2Y + I * XY + J$

$Y = K * X + L * X^2 + M * X^3 + N * Y + O * Y^2 + P * Y^3 + Q * XY^2 + R * X^2Y + S * XY + T$

It should be noted that rubber-sheeting rarely works beyond second order anyway. In my experience the first and second order transformations have accounted for 100% of those that I have performed. Also, many of the parameters such as the coefficient for the $X * Y$ term are often zero. The important point to note is that you do not see any transcendental functions like sine, cosine, or logarithm in the above polynomial equations that are fitted.

Rubber-sheeting is sort of "mathematically ad-hoc". If there is a systematic variation between the two datasets you are registering that varies according to some transcendental function, rubber-sheeting will not capture it.

Rubber-sheeting vs. Re-projection: It is imperative to be aware of the difference between rubber-sheeting and re-projection. The quality of the registration of two different datasets overlapping the same geographic space will be very poor if rubber-sheeting is used when re-projection was appropriate. Map projections such as the Mercator, Robinson, Lambert, etc. usually involve transcendental mathematical functions such as: Sine, Cosine, Tangent, and Logarithm. The rubber-sheeting process cannot capture these kinds of mathematical relationships between datasets. Rubber-sheeting can only capture the aforementioned polynomial transformations, and these are generally better for smaller areas than for large ones. Therefore, if your data is projected you should reproject the data rather than rubber-sheet it to conflate it with other geo-referenced data.

Mastery

Conflation:

Conflation is not a well defined term. Here it is treated as many of the issues associated with merging, combining, and comparing geographic datasets. The previous discussion of registration describes some of the aspects what is an early step of the conflation process. However, other issues to consider are: Data structure compatibility, Accuracy compatibility, Scale of measurement compatibility (e.g. nominal, ordinal, interval, and ratio data), and the various logical, mathematical, and statistical means by which two or more datasets can be combined or compared. The number of ways different kinds of datasets can be combined or compared clearly becomes quite large. Several examples of various types are listed here:

Co-Registering Datasets

1. Registering an image to a geo-referenced vector dataset

Potential purpose: Compare satellite image to census data

2. Registering one image to another image

Potential purpose: Create new images based on map algebra, measure cross-corr elation between the two images, etc.

Combining Datasets

1. Intersecting two vector datasets

Potential purposes: Site selection based on combined attributes

2. Joining tabular data to spatial data

Potential purposes: Data compression, ease of spatial query, etc.

Spatial Scale Resampling/Registration

1. Aggregating a fine scale image to another coarser scaled image

Potential purpose: Allow for map algebra and statistical comparison

Note: There are many decisions to be made here: e.g. Type of aggregation

2. Dissolving a high resolution polygon coverage to a lower resolution

Potential purpose: Generalization, Check for scale dependance, etc

Data Model/Structure Conversions

1. Converting an image to polygons

Potential purpose: create a vegetation map from classified satellite data

2. Convert a vector dataset to a raster image

Potential purpose: allow for comparison of nighttime satellite image to census data

These are only a few of many examples of what can be loosely defined as conflation. Our take on the big picture with respect to conflation is that the merging of spatial data allows for manipulation, analysis, and comparison within and between co-spatially extensive data. It also allows for the creation of datasets derived from diverse conflated data.

Follow-up Units

Unit 34 - Using Overlay Operators

Resources

A nice Image Processing Tutorial on the Web:
[outdated link removed]

Bernstein, R., 1983. Image Geometry and Rectification. Chapter 21 in The Manual of Remote Sensing. R. N. Colwell, ed., Bethesda, MD. American Society of Photogrammetry, 1:875-881.

Campbell, J. B. 1987. Introduction to Remote Sensing. The Guilford Press. 551 pp.

Lillesand, T. M. and R. W. Kiefer. 1994. Remote Sensing and Image Interpretation, 3rd Ed. John Wiley and Sons, Inc. 750pp.

Jensen, J. R. 1996. Introductory Digital Image Processing: A Remote Sensing Perspective. Prentice Hall Inc., 316pp.

Clarke, K.C. 1955 Analytic and Computer Cartography Prentice Hall 334 pp.

Created: May 14, 1997. Last updated: October 5, 1998.

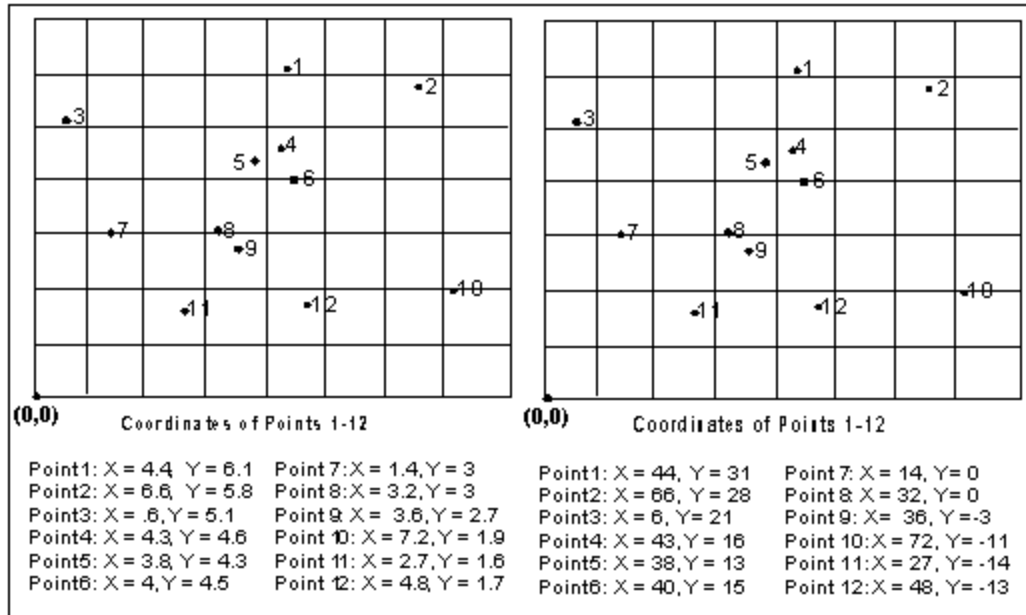
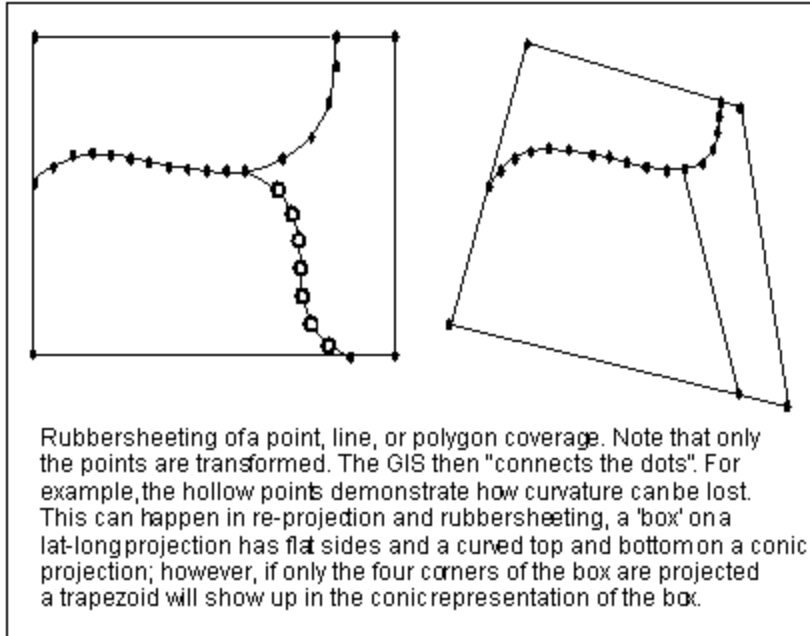


Figure 1

**Figure 2**

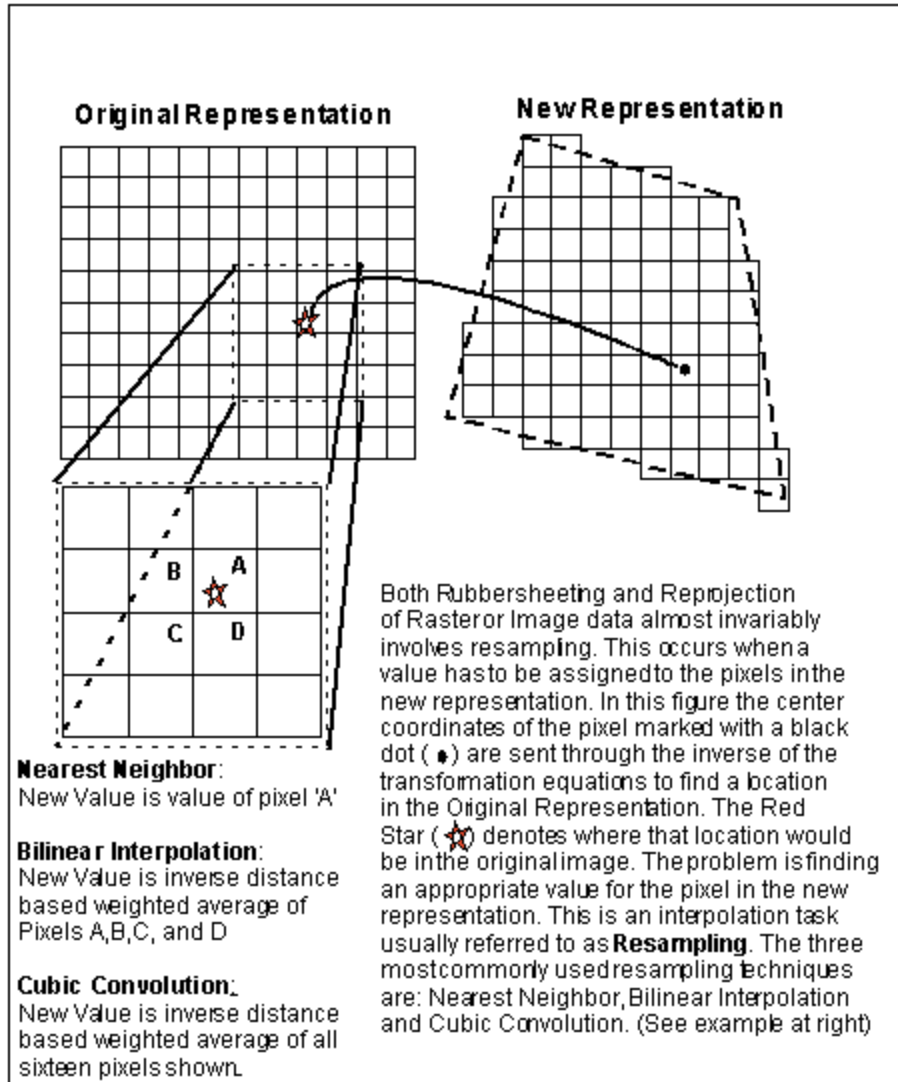


Figure 3

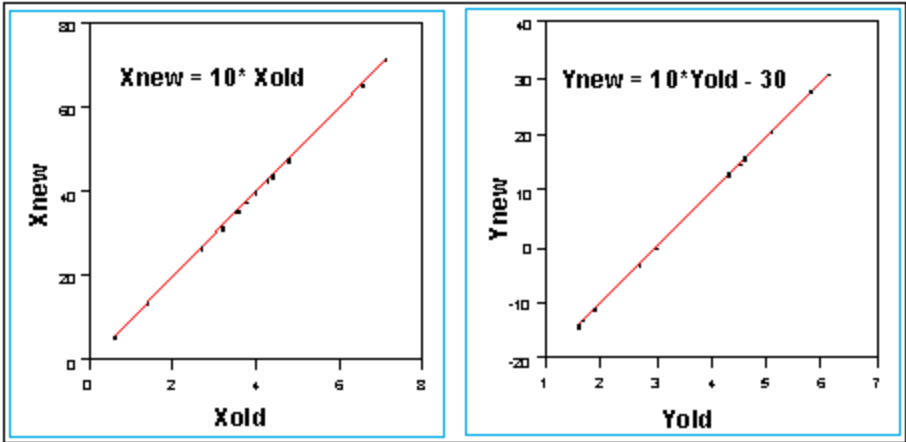


Figure 4