SAND2007-6581P

A Statistical Approach to Image Registration

Katherine (Hansen) Simonson '84
Sandia National Laboratories
Albuquerque, New Mexico

Mathematics Department Seminar
Middlebury College, Middlebury VT
16-October-2007

Tel: 505-844-6551
Email: kmsimon@sandia.gov

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

This research was funded by Sandia's Laboratory Directed Research and Development (LDRD) Program, whose sponsorship is gratefully acknowledged.

Reference: *Simonson, K.M., Drescher Jr, S.M, and Tanner, F.R. "A Statistics-Based Approach to Binary Image Registration with Uncertainty Analysis", TPAMI vol. 29 no. 1, p. 112-125, 2007.*

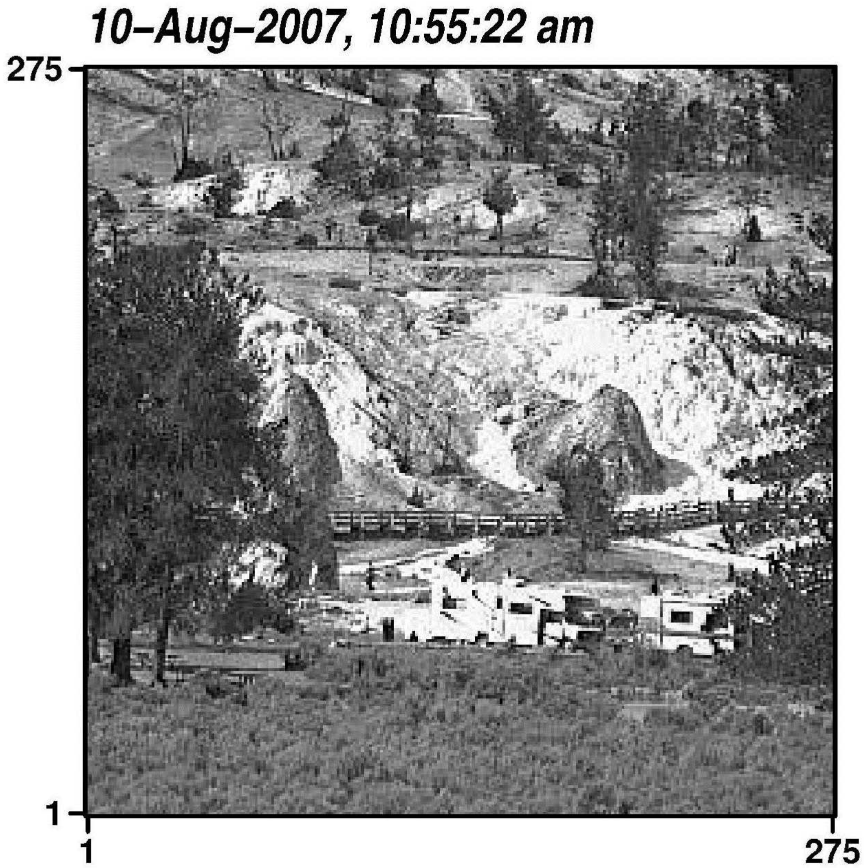
Outline

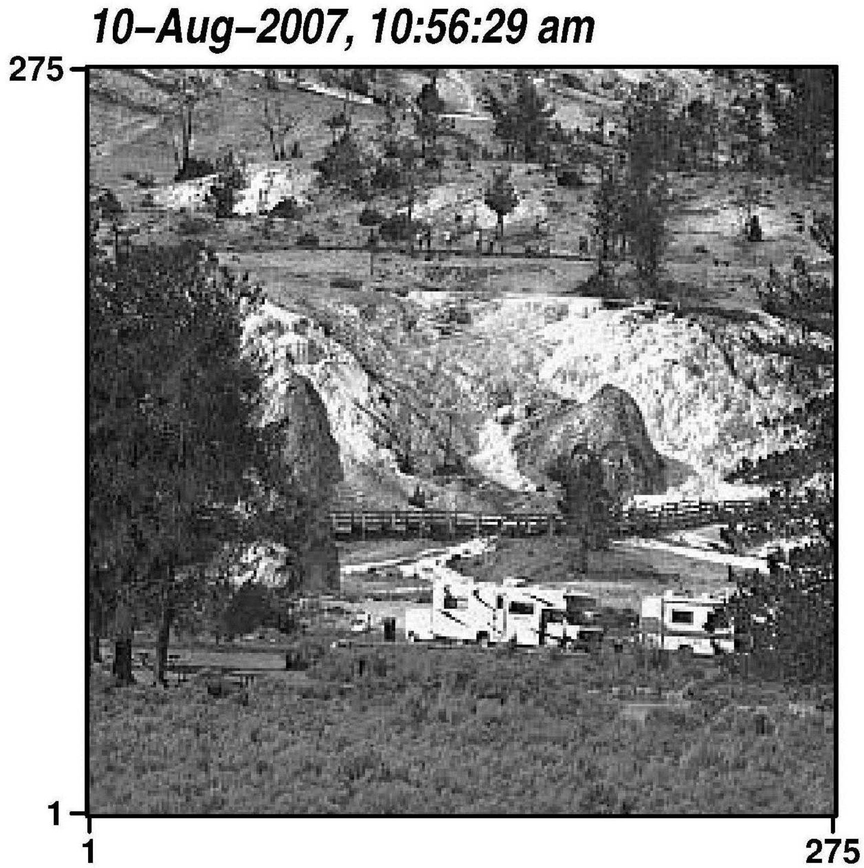
- Motivation
- Mathematical Approach
- Examples
- Discussion

Image Registration

- ***Image Registration*** **refers to the matching (or alignment) of two or more pictures taken, for example, at different times, by different sensors, or from different viewpoints.**
- **Applications of image registration abound:**
  - **In *medical diagnostics*, pairs of images taken at different times can be used to detect new malignancies or track the progress of degenerative diseases.**
  - **In *geologic change detection*, multi-temporal satellite images are compared to identify unstable slopes that are prone to landslides or debris flows.**
  - **In *autonomous navigation and tracking*, platform (e.g., aircraft or robot) motion can be derived from sequences of terrain images.**
- **In all of these cases, it is essential to properly register the images prior to quantifying differences between them.**

Application Example: Area Surveillance

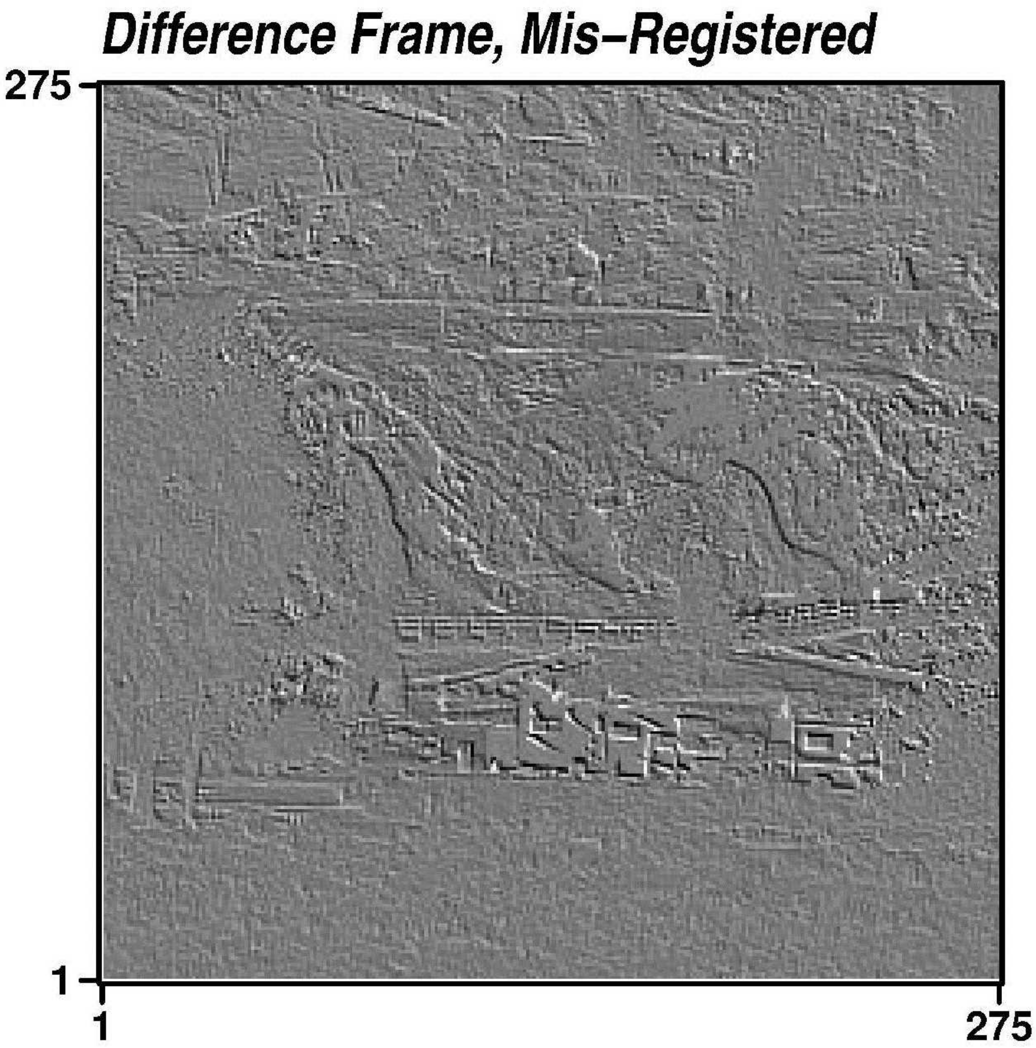




Two frames from a low-resolution digital camera, taken 67 seconds apart.
Can you see what has changed in the scene?
It is difficult to do so in a timely manner!

Frames from Mammoth Hot Springs Webcam
http://www.jacksonholewy.net/web_cams/jh_mammoth_web_cam.php

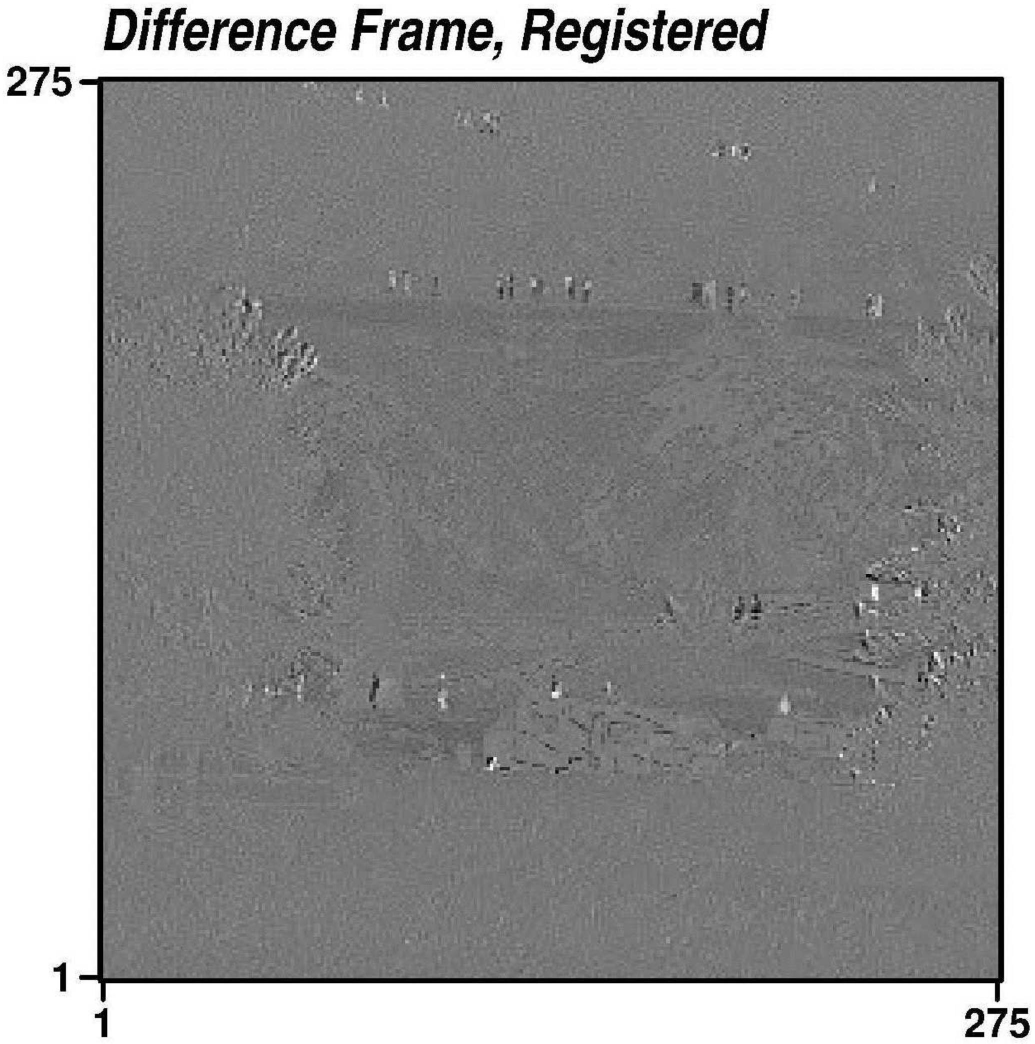
Area Surveillance, 2



In many surveillance applications, it is standard practice to generate difference frames, by subtracting the first image from the second.

If the two images are not properly registered prior to differencing, most of the signal seen in the difference frame is due to spatial gradients in the scene, rather than actual change.

Area Surveillance, 3



When the images are properly registered prior to differencing, the *changes in the scene stand out clearly* in the difference frame.

Here, the motion of each person in the scene is easily detected.

A person wearing light clothing will show up as a bright detection in their current position, and a dark detection in their previous position.

Technical Literature

- ***Thousands of papers on image registration*** **have been published in the engineering literature.**
  - **A SciSearch query on “image <and> registration” for the years 1970 – 2007 turns up 2,428 articles published by the various journals of the Institute of Electrical and Electronics Engineers (IEEE).**

- ***Very few authors have considered the uncertainty*** **associated with registration solutions.**
  - **Refining the query to “image <and> registration <and> uncertainty” returns only 29 articles.**

- **However, *uncertainty measures are essential* for many autonomous applications.**
  - **Solutions that are dubious (involving poorly focused imagery or terrain that is obscured by clouds or smoke) must be distinguished from those that are highly reliable (based on clear images of highly structured scenes).**

Reporting Uncertainty

- **Measurement and sampling uncertainty are familiar concepts. Suppose we take a random sample of 21 voters' preferences:**

Lincoln:	**12**
Nixon:	**9**

- **Based on these poll results, the "best" estimate of candidate Lincoln's support is 12/21 = 57.1429%. But in meaningful terms, is this a better estimate than 57.143%? Or 57.14%? Or 57%?** ***No one would be well served by a report stating simply: "Lincoln has the support of 57% of the population!"***

- **In statistical terms, Lincoln's result is "consistent with" support ranging from 33% to 81%. The responsible media would report "The margin of error is ± 24%", instantly leading the reader to the correct conclusion: Based on these data, the race is "Too close to call".**

Image Registration Uncertainty

- **Our goal is to develop a technique for autonomously measuring the uncertainty associated with image registration solutions.**

***Some image pairs are easy,* with good focus and resolution. The lighting conditions and imaging geometry are similar. We expect to achieve very precise registration.**

http://landsat.usgs.gov/gallery/detail/441/

***Other pairs are more difficult,* with contrast reversals and lighting differences, along with seasonal changes in vegetation. Registration to a fraction of a pixel is probably not feasible.**

http://www.paakoridge.com/livecam/largeshot.asp

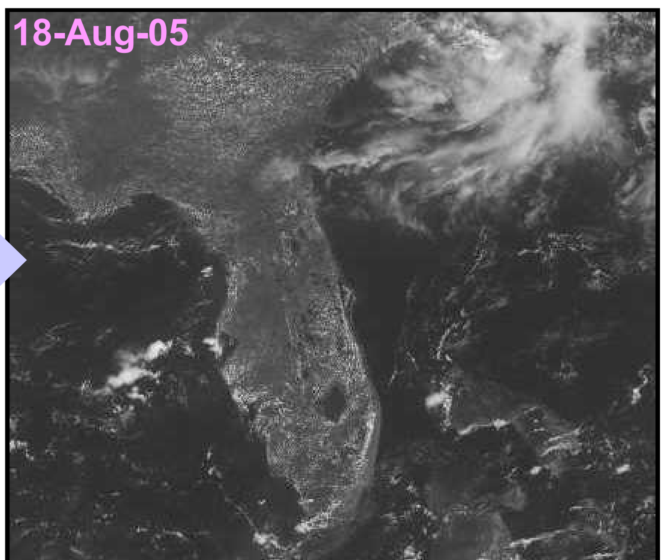
Registration Uncertainty, cont’d

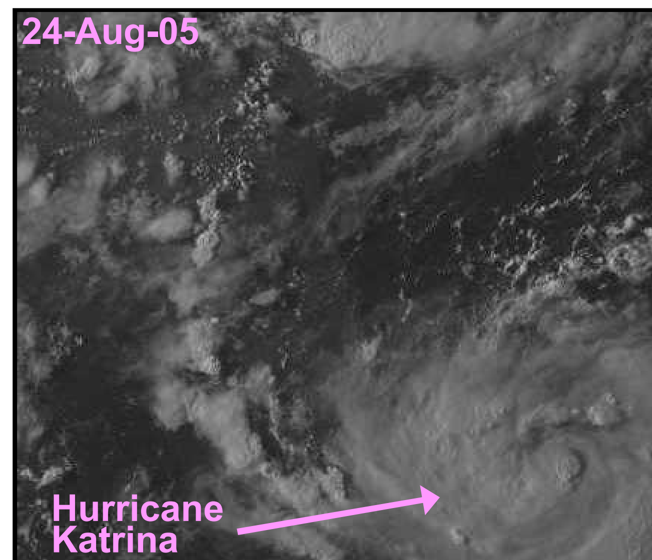
***When much of the scene is obscured** in one or both images, registration relies on a limited number of matching features, and may be imprecise.*

http://www.digiwx-sandia.com/default.aspx

***When registration is not feasible,** we want the method to fail! The user should be informed that a confident solution cannot be achieved.*

http://weather.msfc.nasa.gov/GOES/



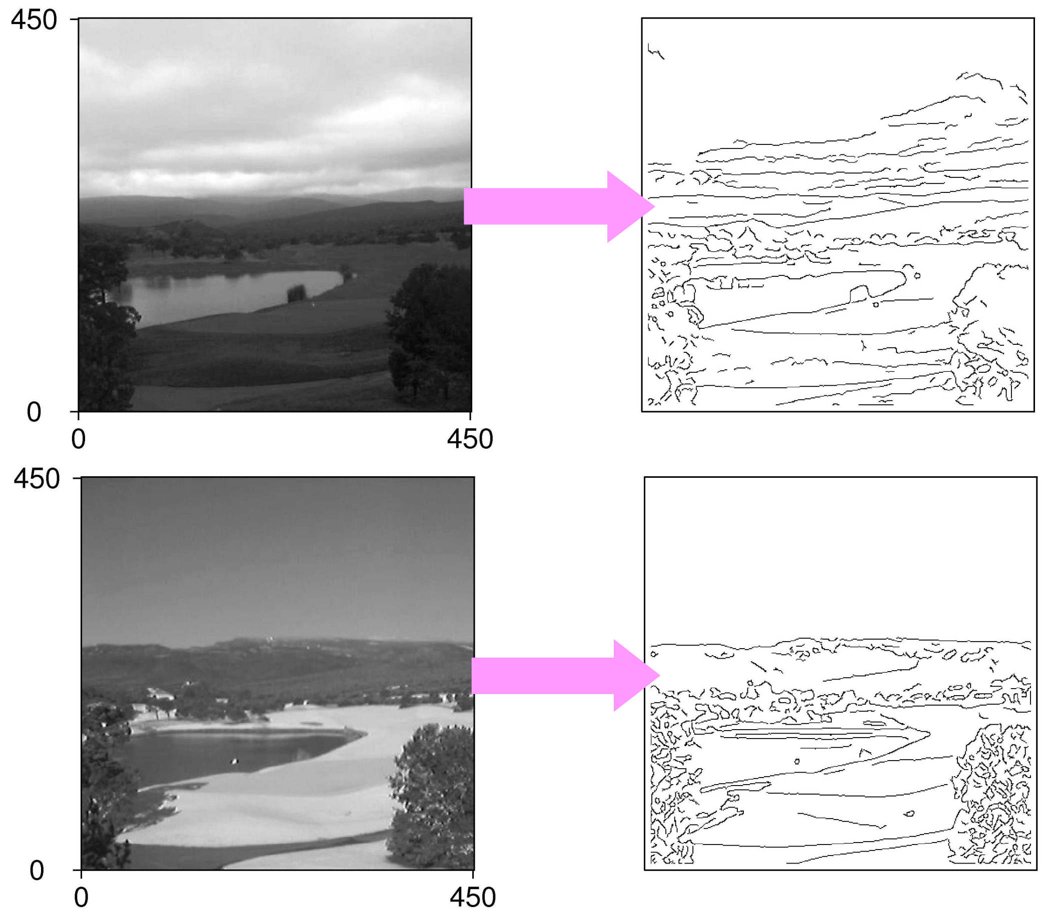


Outline

- Motivation
- Mathematical Approach
- Examples
- Discussion

Binary Transformation

- **To allow for registration across multiple wavelengths, and improve robustness in the presence of contrast reversals and high-frequency artifacts, registration is preceded by an edge detection step.**
  - **Greyscale images are converted to binary edge images.**



Edge detection algorithms identify strong spatial gradients in scene intensity.

We use the Canny edge detector [Canny, 1984] and the Local edge detector [Drescher, 2005].

While greyscale features may change, many of the edge positions remain fixed. Image registration takes place in the edge-detected domain.

Translation Confidence Regions

- Two images may differ from one another is various ways: translations, rotations, affine transformations, warping, etc. *Our method is limited to solving for translations:* (Δrow, Δcol).

  - It can be used to validate higher-order transformations.
  - The user must specify a maximum shift (MAXSHIFT) in each dimension.

- In Statistics, uncertainty is quantified using confidence regions, which contain the values of the parameter(s) of interest that are "consistent with" the available data.

  - We compute a 95% confidence region in the 2D space of (Δrow, Δcol).
  - The region contains the "best" translation, along with all other candidate translations that are consistent with the data.

- Consistency is measured in terms of the probability that the region will contain the true parameter values.

  - A 95% confidence region should be constructed in such a manner that it will have a 95% chance of containing the true shift.

Three-Step Algorithm

- **It is common practice to run registration algorithms on small image patches known as “chips” or “blocks”.**
  - **This enables registration of images containing regions that match poorly due to cloud cover, changes in vegetation, or object motion.**
  - **It also reduces the computational burden for large images.**

- **Our registration algorithm has three basic steps:**

  1 – *Generation of a preliminary list of chips* to be tested;

  2 – *Acceptance testing* for single chips; and

  3 – *Calculation of a joint translation confidence region* over the set of accepted chips.

- **Steps 2 and 3 proceed cyclically until certain exit criteria are met.**
  - **Solutions are rejected if they do not contain enough chips, are imprecise, or have a low overall edge matching percentage.**

Preliminary Chip List

- **First, *choose a chip size* based on the image dimensions and the scene structure. We generally use chips of about 25 × 25 pixels.**

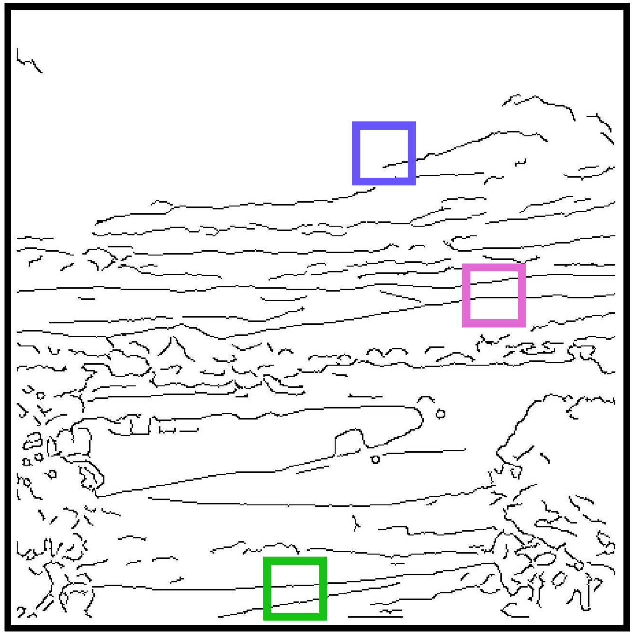
- **Now, generate the preliminary list by *searching along a coarse grid* of candidate chip centers. At each point (*R, C*) on the grid, identify the locations of the edge pixels on the first binary image.**

  - **If there are insufficient edges (less than 40), move on to the next grid point.**
  - **If there are enough edge pixels, look for matches: At each allowable translation (*h, k*), count the number of edges on the first binary image that are also edges on the second binary image, shifted by *h* rows and *k* columns.**
  - **If at least 35% of the edge pixels are matched for at least one translation (*h, k*), add chip (*R, C*) to the preliminary list and test its neighbors (not on the grid) at *(h,k)* as well. Add neighboring chips with edge match percentage at least 35%.**
  - **Move on to the next grid point.**

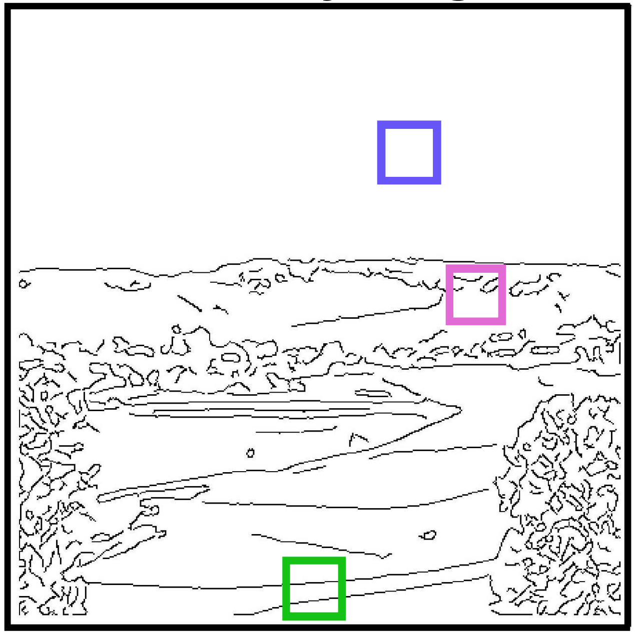
- ***Sort the preliminary chip list* by maximum match percentage, and remove chips that are too close to a higher-ranked candidate.**

Candidate Chip Locations

First Binary Image



Second Binary Image



Chip 1: Rejected, too few edge pixels on image #1
Chip 2: Rejected, low match percentage between images
Chip 3: Good candidate, added to preliminary chip list.

Single Chip Acceptance Testing

- **Chips from the preliminary list are checked, one at a time, until:**

  1. **A sufficient number have been "accepted" for inclusion in the final registration solution, and:**
  2. **Their joint confidence region is precise enough for the application.**

- **Acceptance of a chip pair (one from each image) is determined by *how precisely they can be registered to one another*, where precision is gauged by the size of a confidence region in translation space.**

- ***We consider only full-pixel shifts*, so the confidence regions will contain one or more integral points in the two-dimensional space of Δ*row* and Δ*col*.**

Matching Edges

- **To test the j^{th} chip pair on the preliminary list, *compute the registration solution for this pair only*. Begin by identifying the edge pixels on this chip for the first image. The count is $nedge_j$.**

- **For translation (h,k), define $V_j(h,k)$ as a binary column vector of length $nedge_j$. The *ith* element of $V_j(h,k)$, denoted $v_{ij}(h,k)$, is set equal to one if the i^{th} edge on chip j for the first image is matched in chip j for the second image, *shifted by h rows and k columns*. Otherwise, $v_{ij}(h,k)$ is set to zero.**

- **Define scalar $S_j(h,k)$ to be the number of edge pixels matched at translation (h,k). This is simply the sum of the elements of $V_j(h,k)$:**

$$S_j(h,k) = \sum_{i=1}^{nedge_j} v_{ij}(h,k).$$

It follows that, for all translations, $0 \le S_j(h,k) \le nedge_j$.

Best Translation

- ***The "best" translation is the one matching the largest number of edge pixels.*** **Let S^*_j be the maximum match count, and let (h^*_j, k^*_j) be the corresponding translation.**

- **The observed proportion of edges matched at the best translation point is given by $p^*_j = S^*_j/nedge_j$. If p^*_j is less than 50%, reject chip j and move on to the next chip on the preliminary list.**

- **Otherwise, *determine which additional translations are "not significantly worse"* than (h^*_j, k^*_j). This is accomplished using a two-sample statistical hypothesis test.**

Statistical Test

- We want to test whether the edge matching performance at candidate translation (h,k) is significantly worse than at (h^*_j, k^*_j).

- The model underlying our test must anticipate ***dependencies between different translations*** at the same edge pixel. In statistical terms, this is known as a ***paired data framework***.

- In addition, the model should account for ***non-constant match probabilities across the different edge pixels*** at the same translation. While some edge pixels (e.g., coastlines) should be matched with high priority, others (perhaps noise-induced edges on the first image) will not likely be found in the second image.

Null and Alternative Hypotheses

- **Denote by *$P_{ij}(h,k)$* the true (unknown) probability that the *i^{th}* edge pixel on the *j^{th}* chip of the first image will be matched in the *j^{th}* chip of the second image, translated by *h* rows and *j* columns.**

- ***The null and alternative hypothesis to be tested are:***

$$H_0: \sum_{i=1}^{nedge_j} P_{ij}(h,k) = \sum_{i=1}^{nedge_j} P_{ij}(h_j^*, k_j^*),$$

$$H_1: \sum_{i=1}^{nedge_j} P_{ij}(h,k) < \sum_{i=1}^{nedge_j} P_{ij}(h_j^*, k_j^*)$$

- **That is, *under H_0, the average match probabilities are equal* at translations *(h,k)* and *(h^*_j, k^*_j)*. Under *H_1*, the average match probability at *(h^*_j, k^*_j)* exceeds that at *(h,k)*.**

McNemar's Statistic

- ***Acceptance or rejection of H_0 is determined based the McNemar statistic,*** **which is appropriate for testing hypotheses about paired binary data. Define the random indicator variables A_{ij} and B_{ij}:**

$$A_{ij} = 1,\ \textit{if edge } i \textit{ is matched at translation } (h_j^*, k_j^*) \textit{ but not at } (h, k),$$
$$= 0,\ \textit{otherwise.}$$

$$B_{ij} = 1,\ \textit{if edge } i \textit{ is matched at translation } (h_j, k_j) \textit{ but not at } (h_j^*, k_j^*),$$
$$= 0,\ \textit{otherwise.}$$

- **The sum of the A_{ij} over all *nedge*$_j$ edge pixels is the number of edges matched at translation (h^*_j, k^*_j) but not (h,k). The sum of the B_{ij} is the number matched at (h,k) but not (h^*_j, k^*_j).**

- ***If the sum of the A_{ij} is much larger than the sum of the B_{ij},*** **this provides evidence against the null hypothesis. *If the two sums are close,* we conclude that translation (h,k) is not significantly worse that the "best" translation (h^*_j, k^*_j) : "Too close to call".**

McNemar's Statistic, cont'd

- **The test statistic is given by (McNemar, 1947):**

$$R_j = \frac{\sum_{i=1}^{nedge_j} A_{ij} - \sum_{i=1}^{nedge_j} B_{ij} - 1}{\sqrt{\sum A_{ij} - \sum B_{ij}}}.$$

- **The distribution of R_j can be computed using the exact distribution or a normal approximation.** ***The test is one-sided, with the null hypothesis rejected for large R_j.***

- **For chip *j*, a 95% confidence region for the true translation contains the "best" translation, (h^*_j, k^*_j),** ***along with any other candidate translations for which the null hypothesis was not rejected.***

Chip Acceptance

- **Once the translation confidence region for chip *j* has been computed, the chip is *accepted or rejected from the final registration solution* based on the number of full-pixel translations contained in this region.**

- ***We accept chip j if and only if the confidence region contains eight or fewer translations.***

- **Once chip *j* has been accepted, any chips remaining on the candidate list that overlap with chip *j* (share pixels with it) are removed.**

Joint Confidence Region

- When two or more pairs of non-overlapping chips have been selected, a *joint confidence region* can be constructed from the match statistics computed from each pair.

- Recall that we assume from the outset that the true transformation between the two images is a rigid translation; it follows that each pair of chips has the same true (but unknown) shift.

- Construction of the joint confidence region across several chips is accomplished by concatenating the binary edge vectors, $V_j(h,k)$, over each accepted chip j.

- The processing steps are now the same as for a single chip:
  - Compute the best translation, (h*,k*) over the combined edge pixels,
  - Use McNemar's test to identify other translations that are not significantly worse.

Decision Rule

- A compound decision rule is used to determine when to *keep testing additional chips*, *when to stop*, and *whether the final registration solution will be accepted or rejected*.
  - The rule saves run time by specifying conditions under which processing may cease before testing the entire candidate chip list.

- The rule is as follows:
  1. If at least 6 chips have been selected and the confidence region contains just one translation, *the solution is complete*.
  2. If at least 10 chips have been selected, *the solution is complete*.
  3. If all chips on the preliminary list have been tested, *the solution is complete*.
  4. If the solution is complete, at least 3 chips have been selected, the joint confidence region contains no more than 8 translations, and the joint match percentage (over all selected chips) exceeds 35%, *the solution is accepted*.
  5. If the solution is complete but the conditions of step 5 are not met, *the solution is rejected*.

Decision Rule, cont'd

- **The decision rule ensures that the registration solution will be rejected unless it meets several quality criteria:**
  - **Requiring a minimum of three chips ensures that the solution is not unduly influenced by a single feature.**
  - **The solution is rejected if the joint confidence region is insufficiently precise: that is, if it contains too many translations.**
  - **If the combined match percentage at the best registration point is too low, this suggests that the validity of the rigid translation model underlying the registration algorithm may be called into question.**
- ***It is better to supply the user with an indication that confident registration is not possible than it is to provide false assurance by reporting only a "best" solution!*** **This is particularly true for autonomous applications, where there is no person in the loop to check over the solutions.**

Bottom Line

- Extensive simulation studies confirm that 95% joint confidence regions computed using the approach outlined here *contain the correct translation approximately 95% of the time*.

- *Reporting only the “best” translation* (with the highest match percentage) *can give accuracy rates as low as 20%*, depending on the problem at hand.

- For challenging registration problems (poor focus, unstructured scenes, significant obstruction), *simply reporting a sub-pixel translation solution, without uncertainty bounds, is unacceptable*.

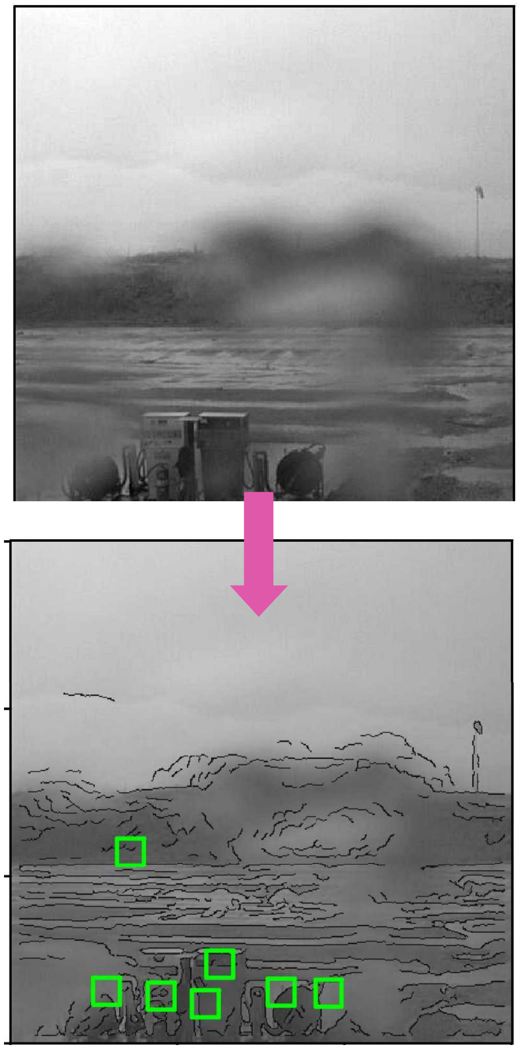
  This is analogous to reporting that a politician has the support of 57.1429% of the population based on a sample of 21 voters!

Outline

- Motivation
- Mathematical Approach
- Examples
- Discussion

Example 1: Partial Obscuration

Greyscale Image 1

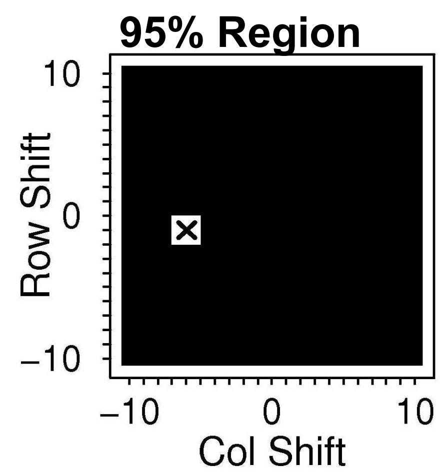


Greyscale Image 2

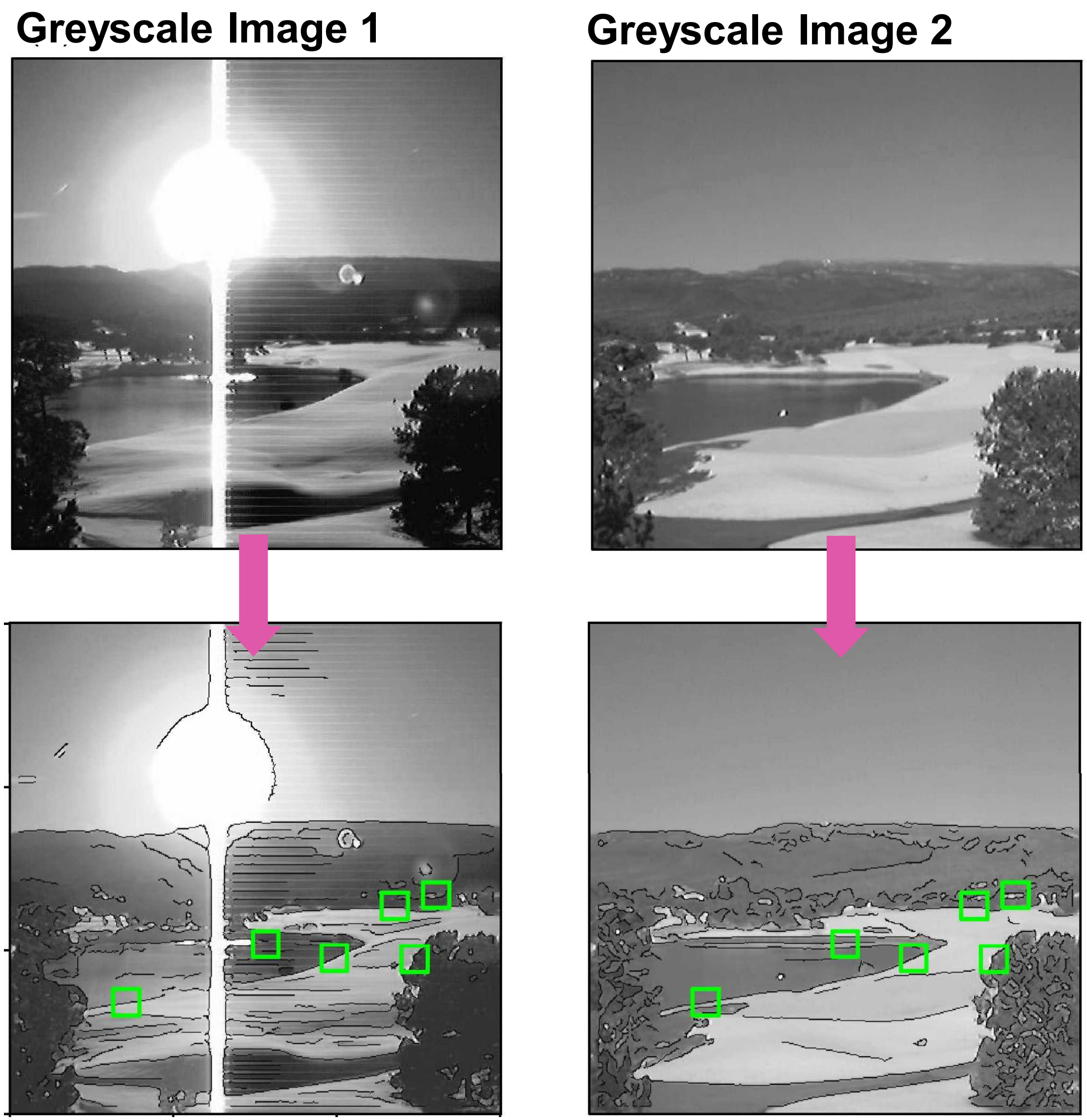
Much of the scene structure is obscured by raindrops in the foreground of the camera.

Passing chips (shown in green) are concentrated near the gas pumps, which contain strong vertical and horizontal features. The 95% region contains a single translation.

7 chips, match percent 55.8%



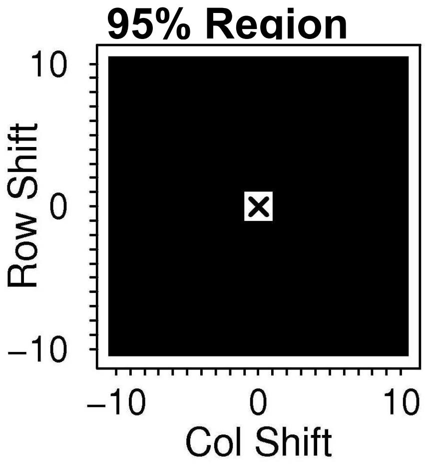
Example 2: Camera Saturation



Successful registration is achieved despite the change in illumination angle and camera saturation in the first image.

Matching chips are found along the treeline and the edge of the water hazard.

6 chips, match percent 52.8%



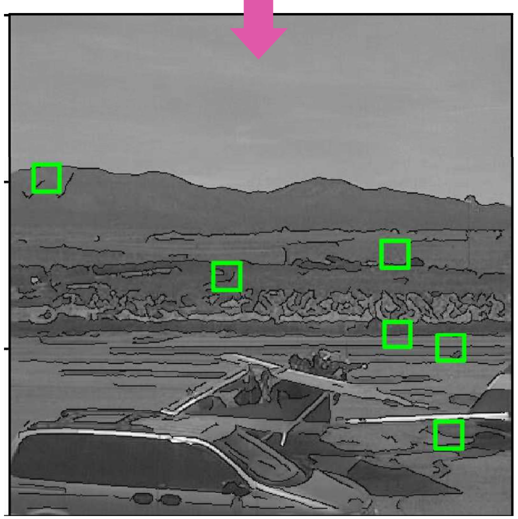
Example 3: Horizontal Uncertainty

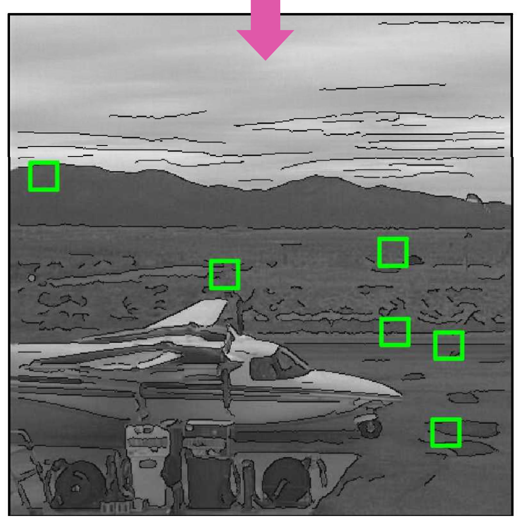
Greyscale Image 1

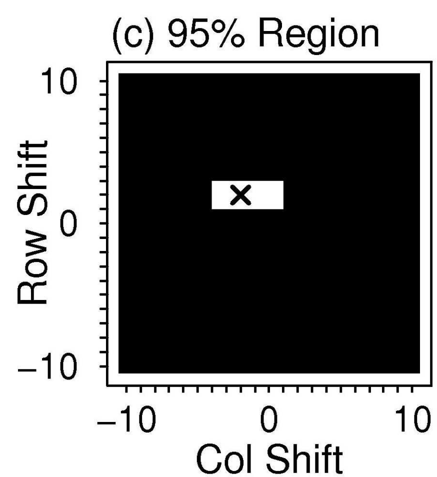
Greyscale Image 2

All of the matched chips have features that are predominantly horizontal. As a result, there is uncertainty regarding the best column shift. The 95% region contains four translations.

4 chips, match percent 47.2%

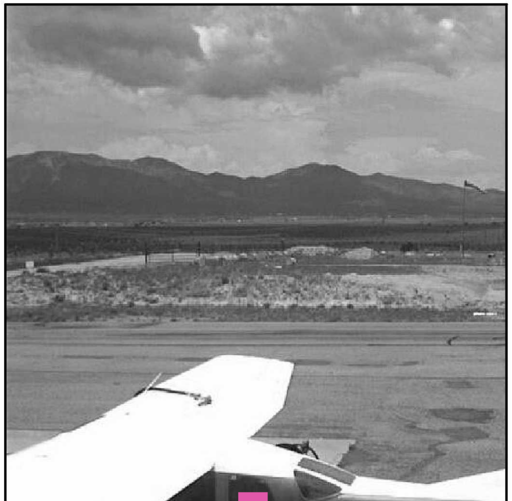






Example 4: Registration Failure

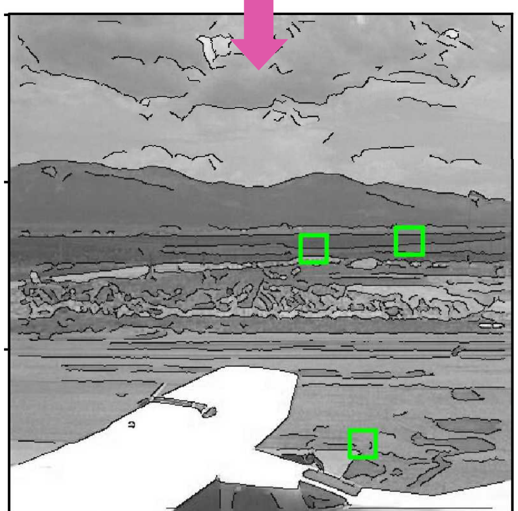
Greyscale Image 1

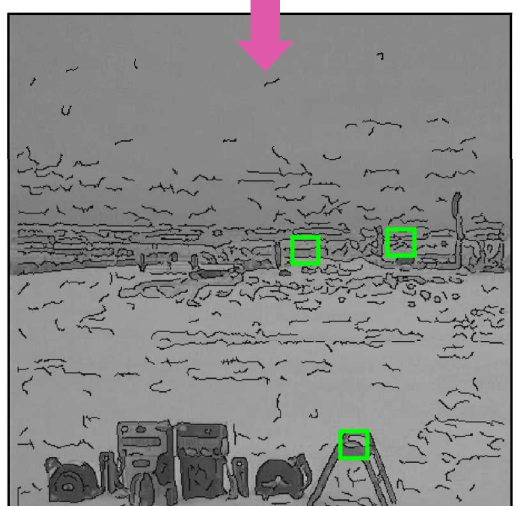


Greyscale Image 2

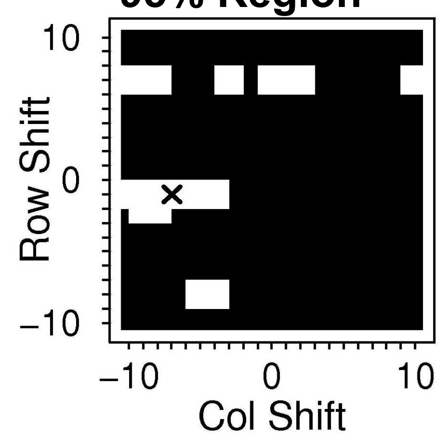
Only 3 matching chips are found, one of which is spurious. The 95% region contains 19 shifts, well above the threshold of 8. The solution is rejected.

3 chips, match percent 40.6%





95% Region



Validating High-Order Transformations, 1

Image 1

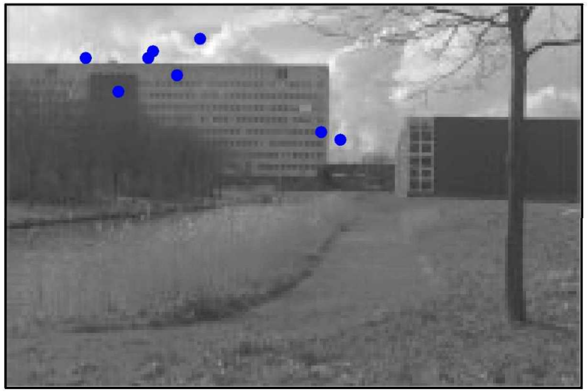


Image 2, Original

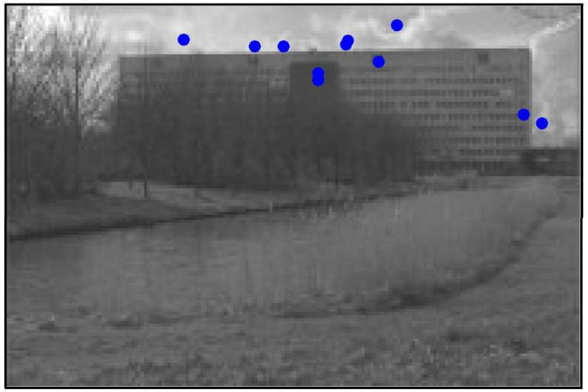
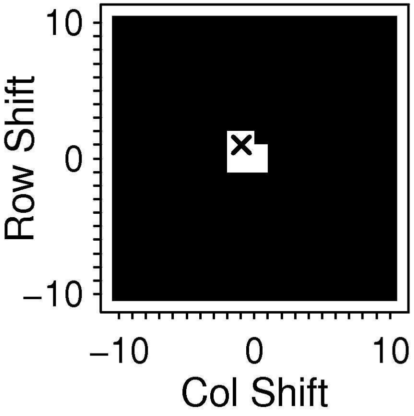


Image 2, Transformed

95% Region



An alternative algorithm (Lowe, 2004) is used to register two images that differ by an affine transformation. Match points are shown in blue.

The transformed image (lower left) is registered to image 1 using our statistical approach. The resulting confidence region (based on 4 chips) is precise and contains the translation (0,0), which reinforces the correctness of the Lowe solution.

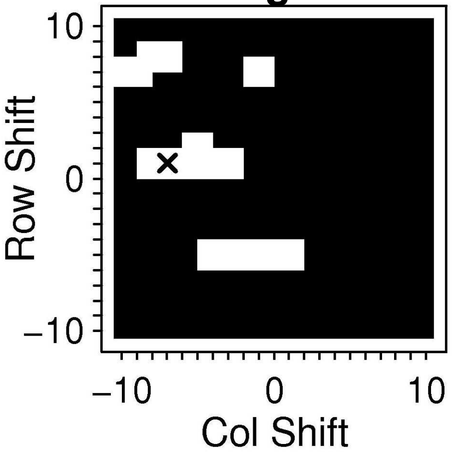
Validating High-Order Transformations, 2

Image 1

Image 2, Original

Image 2, Transformed

95% Region



The Lowe solution (based on pink match points) for this pair of images is incorrect.

Statistical registration of image 1 to the transformed version of image 2 fails: the 95% confidence region, based on 3 chips, contains 18 translations and has a maximum edge match percentage of 25.1%, well below the threshold of 35%.

The statistical approach automatically determines that the Lowe solution is invalid.

Outline

- Motivation
- Mathematical Approach
- Examples
- Discussion

Conclusion

- *The statistical algorithm introduced here is used to register binary edge-detected images.* It has demonstrated solid performance in a number of challenging scenarios: poor image focus, contrast reversals, saturation artifacts, etc.

- In addition to estimating the “best” translation between the images, the *algorithm provides a well defined confidence region* in the space of the registration parameters.

- The user is informed when a confident solution cannot be achieved. *This is particularly important for autonomous operations*.

- While the algorithm itself can only solve for translations, it can be used to *autonomously validate solutions* to higher-order transformation problems.

Image Science, 1

- *Imaging Science is an exciting technical field*, with application in a wide range of problem domains:
  - Environmental Monitoring
  - Biomedical Imaging
  - Humanitarian Operations (Darfur, Bande Aceh, Katrina)
  - Law Enforcement
  - Intelligence, Surveillance, and Reconnaissance (ISR)
  - Consumer Products (entertainment, Google Earth, etc)
- *Jobs are plentiful*, and typically involve working in a team environment, with colleagues from multiple disciplines.
- Depending on the application, the team might include: optical and radar engineers, geophysicists, ecologists, radiologists, astronauts, military officers, computer programmers, graphical designers, etc.

Image Science, 2

- **Image registration is just one of many active areas of research in the general field of imaging science.**

  Others include:

  - ***Pattern recognition*** **(facial ID, license plates, digital mammography)**
  - ***Material ID, plume characterization*** **(hyperspectral sensors)**
  - ***Change detection*** **(Synthetic Aperture Radar, optical, infrared)**
  - ***Image formation and focusing*** **(SAR, exploration seismology)**
  - ***Image restoration*** **(aged and/or low quality photographs and videos)**
  - ***Image compression*** **(limited bandwidth transmission)**
  - ***Foliage penetration*** **(VHF/UHF SAR)**
  - ***Photogrammetry*** **(3-D model building, camera point-of-view)**
  - ***Design tradeoffs*** **(spatial / spectral / temporal resolution; size, weight, power consumption)**

Image Science, 3

- *An undergraduate degree in Mathematics or Computer Science* is an excellent foundation on which to build expertise in imaging science. Be sure to include course work in Statistics.

- Follow up with an MS or PhD in:
  - *Electrical Engineering* (emphasis on Signal and Image Processing); or
  - *Statistics* (emphasis on Engineering Applications).
  - Tuition support through fellowships and internships is readily available at most Engineering schools.

- *Liberal Arts undergraduates* have some natural advantages!
  - Superior oral and written communications skills.
  - Ability to interact productively with team members having vastly different perceptions and areas of expertise.
  - Broad academic background enabling fuller integration of technical, cultural, legal, moral, practical, and aesthetic issues.

Questions?