LAMP-TR-064 MDA0949-6C-1250
CAR-TR-961 N660010028910//IIS9987944
CS-TR-4217 February 2001

A Point Matching Algorithm for Automatic Groundtruth Generation

Doe-Wan Kim and Tapas Kanungo

Language and Media Processing Laboratory
Center for Automation Research
University of Maryland
College Park, MD 20742
{dwkim,kanungo}@cfar.umd.edu

Abstract

Geometric groundtruth at the character, word, and line levels is crucial for developing and evaluating optical character recognition (OCR) algorithms. Kanungo and Haralick proposed a closed-loop methodology for generating character-level groundtruth for rescanned images. In this paper, we present a robust version of their methodology. We grouped the feature points and used a feature point registration algorithm on the grouped feature point set to estimate the transformation. The Euclidean distance between character centroids was used as the error metric. We performed experiments on the University of Washington data set.

This work was funded in part by the Department of Defense under Contract MDA0949-6C-1250, the Defense Advanced Research Projects Agency under Contract N660010028910, and the National Science Foundation under Grant IIS9987944.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
FEB 2001		**00-02-2001 to 00-02-2001**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
A Point Matching Algorithm for Automatic Groundtruth Generation	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Language and Media Processing Laboratory,Institute for Advanced Computer Studies,University of Maryland,College Park,MD,20742-3275**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**36**	

LAMP-TR-064
CAR-TR-961
CS-TR-4217

MDA0949-6C-1250
N660010028910//IIS9987944
February 2001

A Point Matching Algorithm for Automatic Groundtruth Generation

Doe-Wan Kim and Tapas Kanungo

A Point Matching Algorithm for Automatic Groundtruth Generation

Doe-Wan Kim and Tapas Kanungo

Language and Media Processing Laboratory
Center for Automation Research
University of Maryland
College Park, MD 20742
{dwkim,kanungo}@cfar.umd.edu

Abstract

Geometric groundtruth at the character, word, and line levels is crucial for developing and evaluating optical character recognition (OCR) algorithms. Kanungo and Haralick proposed a closed-loop methodology for generating character-level groundtruth for rescanned images. In this paper, we present a robust version of their methodology. We grouped the feature points and used a feature point registration algorithm on the grouped feature point set to estimate the transformation. The Euclidean distance between character centroids was used as the error metric. We performed experiments on the University of Washington data set.

This work was funded in part by the Department of Defense under Contract MDA0949-6C-1250, the Defense Advanced Research Projects Agency under Contract N660010028910, and the National Science Foundation under Grant IIS9987944.

1 Introduction

Character, word, and line-level geometric groundtruth is crucial for optical character recognition (OCR) algorithm development and evaluation. Such groundtruth is typically created manually and therefore its creation is time-consuming, expensive, and prone to human errors.

Consider a case in which researchers already have geometric groundtruth for a small set of document images but would like to use these document-groundtruth pairs to bootstrap the construction of a larger (more varied) data set. Two scenarios are possible. In the first scenario, the groundtruth for the set of original real document images is created manually, and in the second scenario, the groundtruth for the set of original synthetic document images is generated automatically. In both cases the algorithm developer would like to print, photocopy, fax and rescan the original document images and then automatically generate the geometric groundtruth for the rescanned documents.

In this paper, we present a point matching based algorithm to automatically generate the groundtruth for rescanned images. The algorithm extracts feature points from the original and rescanned images and then registers the two images using a point matching algorithm. The groundtruth for the rescanned images is then generated by transforming the groundtruth of the original images.

In Chapter 2, related research is summarized. The automatic groundtruth generation methodology is outlined in Chapter 3, and the matching algorithms are discussed in Chapter 4. We discuss the impact of image pattern complexity on image registration in Chapter 5. The error metric and experimental protocol for conducting controlled experiments are discussed in Chapter 7. Experimental results are presented in Chapter 8. In Chapter 9, image registration is used for generating groundtruth for microfilmed and faxed images. Finally, in Chapter 10, we provide our conclusions.

Part of the work presented in this paper appeared in DAS2000 [15].

2 Previous Work

Kanungo and Haralick [13, 14] proposed a methodology for automatically generating the groundtruth of a rescanned image by estimating the transformation between two images and then transforming the groundtruth using the estimated transformation. They estimated the transformation from corresponding pairs of feature points. Four corner points of the images were used as feature points to estimate the transformation. The point matching registration algorithm was then improved by using a robust local template matching algorithm. However, their method is not robust when part of the image is missing or there are extra feature points in the image. This drawback can be overcome by using all the available feature points. Hobby [10] improved the registration by considering all feature points. He used a direct search optimization method to minimize the mismatch in the estimated transformation. However, his method finds a local minimum instead of a global minimum. More recently, Viard-Gaudin et al. [25] proposed a methodology for creating groundtruth for handwritten documents. They designed a database of online and offline handwritten data. They manually determined corresponding points in the online and offline domain and then estimated the affine transformation between the two

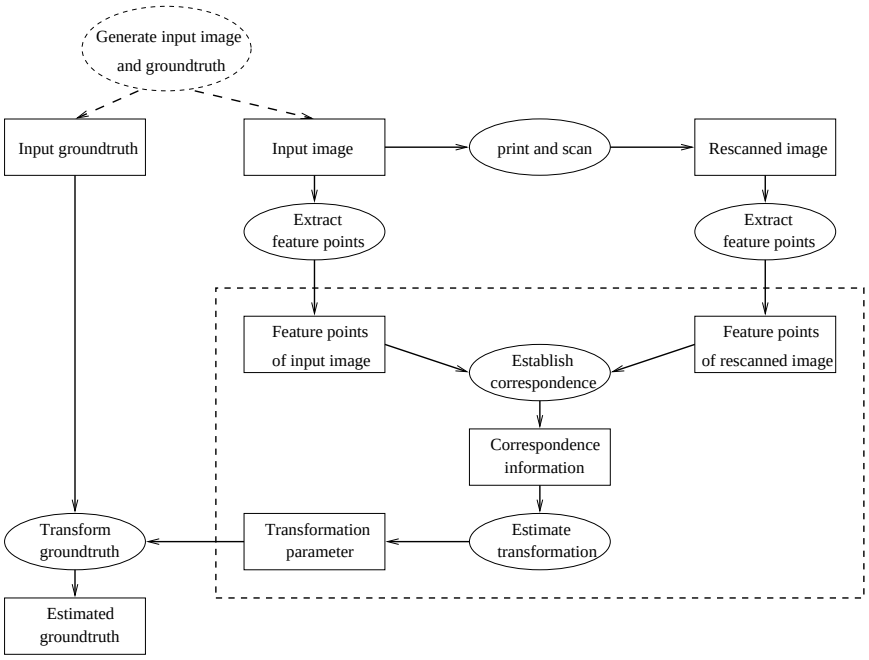


Figure 1: The automatic closed-loop methodology of Kanungo and Haralick.

coordinate systems.

Numerous feature point matching algorithms have been reported in the literature. Baird [1] used feature points to do image matching. Breuel [3] also proposed an algorithm for feature point matching. He estimated the transformation by subdividing the transformation space. Huttenlocher et al. [11, 12] used a branch-and-bound algorithm using Hausdorff distance as the distance measure. They used the distance transform to determine nearest neighbors. Mount et al. [20] proposed a modified branch-and-bound algorithm based on partial Hausdorff distance. They used kd-tree-based nearest neighbor searching to find correspondences. These algorithms are discussed in more detail in Section 4.2.

3 The Automatic Groundtruthing Methodology

Given an image and its groundtruth information, we wish to generate groundtruth for an image which is a transformed (scanned, photocopied, microfilmed, faxed, etc.) version of the original image. The basic idea is to estimate the transformation between the two images and then transform the groundtruth information using the estimated transformation.

Figure 1 illustrates the methodology that Kanungo and Haralick [14] used for generating groundtruth information for real images. Four corner points of the images were used as feature points to estimate the transformation. The four feature points, p_1, p_2, p_3 and p_4 were determined by the following equations:

$$p_1 = \arg\min_{a_i}(x(a_i) + y(a_i)),$$

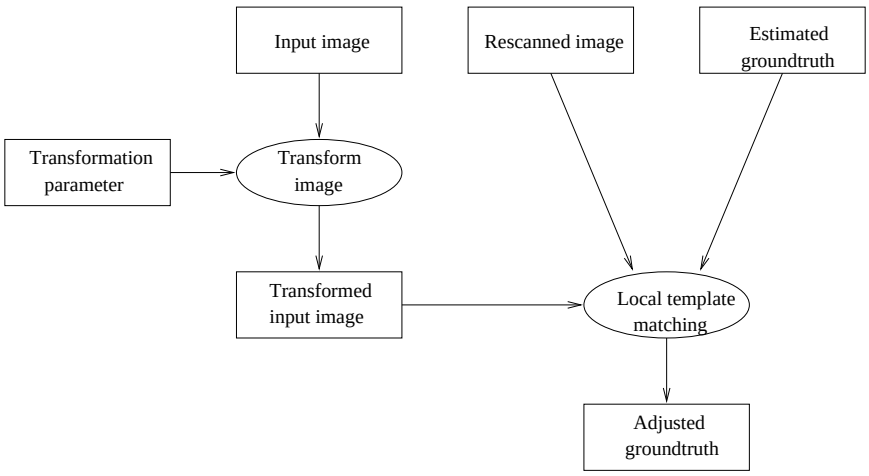


Figure 2: Local template matching.

$$p_2 = \arg\max_{b_i}(x(b_i) - y(b_i)),$$

$$p_3 = \arg\min_{c_i}(x(c_i) + y(c_i)),$$

$$p_4 = \arg\max_{d_i}(x(d_i) - y(d_i)),$$

where a_i, b_i, c_i and d_i are respectively the upper-left, upper-right, lower-right, and lower-left corners of the bounding boxes of each connected component in the image. More improvement is achieved by applying a local template matching algorithm described in Figure 2. The dashed rectangle in Figure 1 is the module that is being replaced by the algorithm described in this paper.

First we extract the connected components of the original and transformed images. The number of connected components in a typical document image is 1000-5000, which makes the running time of the estimation procedure too large. To reduce the complexity of the problem, we group the connected components. The groups are approximately at the word level. As a result of grouping, the number of feature points to be considered is reduced to about 20-25% of its original size. We explain the feature point grouping procedure in Section 4.1.

Using the two feature point sets, one from the original image and the other from the transformed image, we estimate the transformation by using the feature point registration algorithms described in Section 4.2. Figure 3 shows an illustration of this procedure.

4 The Matching Algorithm

We need to find the correspondence and the transformation between two point sets. There are two major steps in the matching procedure: (i) feature point grouping and (ii) feature point registration.

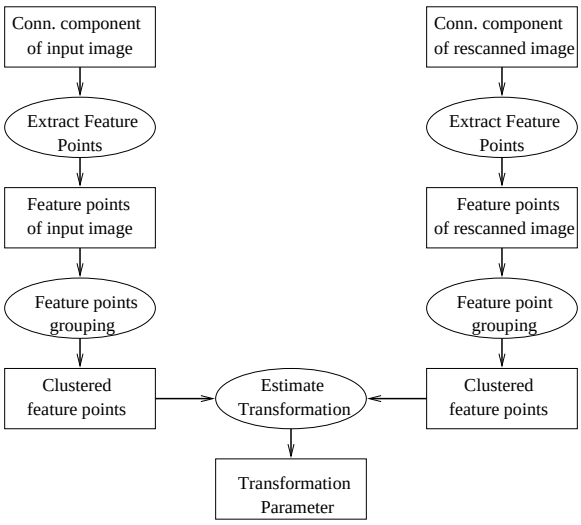


Figure 3: The automatic registration methodology.

4.1 Feature point grouping

To reduce the size of the problem, we group connected components at the word token level. Let B be the set of bounding boxes, $NN^k(b)$ be the k nearest neighbors of bounding box b, PQ be a priority queue, τ be a threshold, and $root(b)$ be the root of b, which is initialized to be b. The key of the priority queue is the distance between the two bounding boxes. Bounding boxes with the smallest distance appear on top of the queue. In selecting the threshold, we used the threshold selection method of Kittler and Illingworth [16].

The thresholding works as follows. Assume that the observations come from a mixture of two Gaussian distributions having respective means and variances (μ_1, σ_1^2) and (μ_2, σ_2^2) and respective proportions q_1 and q_2. We determine the threshold T that results in $q_1, q_2, \mu_1, \mu_2, \sigma_1, \sigma_2$. They minimize the Kullback directed divergence [18] J from the observed histogram $P(1), \ldots, P(I)$ to the unknown mixture distribution f, where

$$J = \sum_{i=1}^{I} P(i) \log[\frac{P(i)}{f(i)}] = \sum_{i=1}^{I} P(i) \log P(i) - \sum_{i=1}^{I} P(i) \log f(i)$$

and

$$f(i) = \frac{q_1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}(\frac{i-\mu_1}{\sigma_1})^2} + \frac{q_2}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}(\frac{i-\mu_2}{\sigma_2})^2}$$

Because the first term of J does not depend on the unknown parameters, the minimization can be done by minimizing the second term. Assume that the modes are well separated. Then for some threshold t that separates the two modes

$$f(i) \approx \begin{cases} \frac{q_1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}(\frac{i-\mu_1}{\sigma_1})^2} & , i \leq t \\ \frac{q_2}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}(\frac{i-\mu_2}{\sigma_2})^2} & , i > t \end{cases}$$

The function $H(t)$ to be minimized can then be written as

$$H(t) = -\sum_{i=1}^{t} P(i) \log \frac{q_1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}(\frac{i-\mu_1}{\sigma_1})^2} - \sum_{i=t+1}^{I} P(i) \log \frac{q_2}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}(\frac{i-\mu_2}{\sigma_2})^2}$$

```
Input
      B: Set of bounding boxes
      I: Input image
Output
       Set of grouped bounding boxes
begin
      for all b ∈ B
          root(b) ← b
          for all b' ∈ NN^k(b)
              put (b, b') into PQ
      pair (b, b') ← pair with smallest distance of PQ
      while distance of (b, b') < τ
      do
          if root(b) ≠ root(b')
          then for all b'' with root(b')
                    root(b'') = root(b)
      end
end
```

Figure 4: The feature point grouping algorithm.

From the assumption of well-separated modes, the mean and variance estimated from $P(1), \ldots P(t)$ will be close to μ_1 and σ_1, and the same for the second part. By using these estimated values, we can evaluate $H(t)$ for each t. We choose the threshold t which minimizes $H(t)$.

The grouping algorithm is illustrated in Figure 4. In Figure 5, we show an image overlaid with bounding boxes of the grouped connected components. This sample image contains 2127 connected components, and 442 groups. We can see that these groups are approximately at the word level. Grouping takes less than 10 seconds per image when run on a Sun Ultra-Sparc 5 with clock speed 361.2 MHz.

4.2 Feature point based registration algorithms

With feature points generated by the methodology described in Section 4.1, we need to estimate the transformation between the two sets of feature points. In this section, we discuss several registration algorithms that can be used for this purpose. All the algorithms work on feature points, and therefore we can use any of these methods for our matching problem. The algorithms take two sets of feature points as input, and estimate the transformation between them. We also need to give the bounds for the initial search space.

4.2.1 Huttenlocher et al.'s algorithm

Huttenlocher et al. [11, 12] proposed a feature matching algorithm using the Hausdorff distance as a similarity measure. A set of transformations (a cell) is defined such that

ROBOTEX: An Autonomous Mobile Robot for Precise Surveying *

Xavier LEBEGUE and J. K. AGGARWAL
Computer and Vision Research Center,
Dept. of Electrical and Computer Engr.,
The University of Texas at Austin,
Austin, Texas 78712-1084, U.S.A.

Abstract. The RoboTex project aims at automatically constructing an exact CAD representation of buildings using a mobile robot. This paper reports on the current status of the project. The hardware of the robot is described, with special emphasis on issues relating to measurement accuracy, and algorithms used to process the sequences of monocular images acquired by the robot are presented. Results of automatic indoor surveying are shown and compared to direct measurements in the scene. The techniques developed here have important applications in architectural surveying, scene understanding, and precise robot navigation.

1 Introduction

This paper describes RoboTex, a mobile robot especially designed for building accurate 3-D maps of its environment. The goal of the RoboTex project is to enable a robot to automatically explore a building to construct a very accurate CAD representation. This CAD representation should be as close as possible to what an architect would generate.

Traditionally, the tasks of a robot's perception system are to detect obstacles, find the free space, and estimate the position of the robot in the world. Here, the focus is on building a useful 3-D description of the world. Our 3-D representation of the environment differs primarily from representations used by other robots in that:

1. It must concentrate on *semantically significant* features.

2. It must be more accurate than is strictly necessary for navigation alone.

To satisfy the first constraint, we chose to concentrate on straight edges with particular orientations in the 3-D scene. Typically, there are three prominent 3-D orientations in indoor scenes and outdoor urban scenes: the vertical, and two horizontal orientations perpendicular to each other. Our approach considers only polyhedral objects with such edges. This assumption holds for most large architectural features such as walls, doorways, floors, and ceilings. The second constraint, accuracy, has multiple implications for both the hardware and the software of the robot.

*This research was supported in part by the DoD Joint Services Electronics Program through the Air Force Office of Scientific Research (AFSC) Contract F49620-89-C-0044, and in part by the Army Research Office under contract DAAL03-91-G-0050.

Figure 5: Sample document image overlaid with the bounding boxes of the grouped connected components.

Input
    I: Original point set
    R: Transformed point set
Output
    $\hat{t}$: Estimated transformation
begin
    Initialize a cell C_0 to contain all transformations of interest.
    Initialize a list of cells L with C_0.
    while cell_size > threshold
        for each cell $c \in L$
            if c can contain a transformation t s.t. $H_{LK}(I, t(R)) \leq \tau$
            add c to interesting list IL
        Create a new L with smaller cells s.t. they completely cover IL.
end

where $H_{LK}(I, t(R)) = \max(h_L(I, t(R)), h_K(t(R), I))$,
and $h_K(t(R), I) = K^{th}_{r \in t(R)} \min_{i \in I} \| i - r \|$.

Figure 6: Huttenlocher et al.'s algorithm.

the optimum transformation lies inside this cell. A list of interesting cells is created and initialized to be this cell. Let I be the original points and R be the transformed points. For each cell in the list, determine whether it is possible that the cell contains a transformation t for which $H_{LK}(I, t(R)) \leq \tau$, where

$$H_{LK}(I, t(R)) = \max(h_L(I, t(R)), h_K(t(R), I)),$$

and

$$h_K(t(R), I) = K^{th}_{r \in t(R)} \min_{i \in I} \| i - r \| .$$

If the rule is satisfied, the cell is marked as interesting. Once the entire list has been scanned, a new list of smaller cells (of the same size) is constructed such that it completely covers the interesting cells. This step is repeated until the cell size is smaller than a threshold. Figure 6 shows the pseudo-code of this algorithm.

4.2.2 Breuel's algorithm

Breuel [3] proposed a registration algorithm called RAST (Recognition using Adaptive Subdivisions of Transformation space). We define a *box* to be a set of transformations. Initially, the box contains all the transformations we would like to consider. The algorithm finds all possible correspondences between the two feature point sets and evaluates the quality of the match resulting from this set of correspondences.

If the upper bound on the best possible match is either (i) smaller than the required minimum quality or (ii) smaller than the best solution found so far, we abandon this

```
Input
      I: Original point sets
      R: Transformed point sets
Output
       t̂: Estimated transformation
SearchBox(box, depth, candidates)
begin
      intersecting = all candidates that intersect box
      containing = all candidates that contain box
      axis = depth mod 4
      if evaluate(intersecting) ≤ best_Quality then return
      else if(candidates = containing) or (depth > max_Depth)
           then best_Quality=evaluate(intersecting)
                best_Box = box
                return
      else SearchBox(left(box,axis), depth+1, intersecting)
           SearchBox(right(box,axis), depth+1, intersecting)
end
RAST(constraints, max_Depth, min_Quality)
begin
      best_Quality = min_Quality
      best_Box = none
      SearchBox(entire_box, 0, constraints)
      return best_Box
end
```

Figure 7: Breuel's algorithm.

part of the transformation space. Otherwise, we subdivide the current box into smaller regions and repeat the same procedure recursively. This process terminates when all boxes have correspondences, or when a threshold is reached. The RAST algorithm is given in Figure 7.

4.2.3 Mount et al.'s algorithm

Mount et al. [20] proposed a branch-and-bound algorithm for feature point matching. They used the partial Hausdorff distance [11] as the similarity measure. Given point sets A and B and parameter k, the partial Hausdorff distance is defined as

$$H_k(I, R) = k^{th}_{i \in I} \min_{r \in R} dist(i, r).$$

Let T be the range of the affine transformation, and ϵ be the error bound. The basic approach of the branch-and-bound algorithm is as follows. For a given T, we first compute the upper and lower bounds on similarity. Next, a priority queue is constructed such that the element that has the largest size is on top of the queue. In each iteration, we pick up the largest element from the priority queue and see if its similarity lower bound is

```
Input
      I: Original point sets
      R: Transformed point sets
      T: Initial search space
Output
       t̂: Estimated transformation
begin
      construct and initialize PQ with given T
      while PQ size ≠ 0 and best_similarity > ε
      do
          T ← next element in PQ
          compute lower bound of similarity for T
          if lower bound of T > best_similarity - ε
          then kill this cell and proceed to the next one
          compute upper bound of similarity for T
          if upper bound of T < best_similarity
          then update best_similarity and transformation
          split T into T1 and T2
          insert T1 and T2 into PQ
      end
end
```

Figure 8: Mount et al's algorithm.

better than the current best similarity. If not, we kill that element and proceed to the next largest element. Otherwise, we compute the upper bound and check if it is better than the current best similarity. If it is, we (i) update the best similarity to be the upper bound of the current element, (ii) update the best transformation, (iii) split the element into two parts along the longest side,and (iv) insert both new elements into the priority queue. This process is iterated until we achieve the target similarity or there are no more elements to be processed in the queue. In computing the upper and lower bounds of a given range of transformation, we use the kd-tree-based nearest neighbor searching algorithm proposed in [2, 7]. The matching algorithm is illustrated in Figure 8.

4.2.4 Hobby's algorithm

Hobby proposed a new approach to the registration problem [10]. In this algorithm, he defined a mismatch function and found the minimum values using direct search optimization methods, such as Nelder-Mead's [21] and Torczon's [24] algorithms. The mismatch function is defined as follows. Let R be the real image with connected component C^R, and I be the ideal image with groundtruth G^I. Using the initial affine transformation T^θ, transform C^R to $C^{\theta R}$. Then for each $g_i^I \in G^I$, we can choose the $c_j^{\theta R} \in C^{\theta R}$ such that the distance $d(g_i^I, c_j^{\theta R})$ is minimized. Then apply a standard vector norm (he used the L_4 norm) to the resulting list of d values.

Input

I: Original point sets

R: Transformed point sets

T^θ: Initial transformation from R to I

Output

θ: Estimated transformation

$\theta = (t_{xx}, t_{xy}, t_{yx}, t_{yy}, t_x, t_y)$

$d(g_i^I, c_j^{\theta R})$: Distance measure between $g_i^I, c_j^{\theta R}$

begin

Let C^R be the connected components of R

Let G^I be the groundtruth of I

$c_i^R \in C^R, g_i^I \in G^I$

$C^{\theta R} = T^\theta(C^R), c_i^{\theta R} \in C^{\theta R}$

for each $g_i^I \in G^I$

for each $c_j^{\theta R} \in C^{\theta R}$

compute $d(g_i^I, c_j^{\theta R})$

$k_i = arg\min_j d(g_i^I, c_j^{\theta R})$

Find θ that minimizes the function

$f(\theta; I, G^I, R) = \sqrt[4]{\sum_{g_i^I \in G^I}(d(g_i^I, c_{k_i}^{\theta R}))^4}$

end

Figure 9: Hobby's algorithm.

The distance measure d is defined as follows. Assume that we have two boxes A and B. The distance between them is defined to be

$$\begin{aligned} d(A,B) = \quad & \min(d_f(A_{x1}, A_{x2}, B_{x1}, B_{x2}) + d_f(B_{y1}, A_{y2}, B_{y1}, B_{y2}), \\ & d_f(B_{x1}, B_{x2}, A_{x1}, A_{x2}) + d_f(B_{y1}, B_{y2}, A_{y1}, A_{y2})) \\ & + d_p(A_{x2} - A_{x1}, B_{x2} - B_{x1}) + d_p(A_{y2} - A_{y1}, B_{y2} - B_{y1}) \end{aligned}$$

where the $x1$ and $x2$ subscripts refer to a box's minimum and maximum x coordinates and the $y1$ and $y2$ subscripts refer to a box's minimum and maximum y coordinate. d_f and d_p are defined to be

$$d_f(x_1, x_2, x_3, x_4) = \begin{cases} 0 & \text{if } x_3 \le x_1 \text{ and } x_2 \le x_4 \\ \min(\mid x_3 - x_1 \mid, \mid x_4 - x_2 \mid) & \\ + \max(0, x_2 - x_1 - (x_4 - x_3)) & \text{otherwise} \end{cases}$$

$$d_p(a,b) = \max(0, max(a,b) - 8\min(a,b)).$$

He used four corner points of the image as used by Kanungo and Haralick [14] to estimate the initial affine transformation T^θ. Figure 9 shows his algorithm. More details about this algorithm and the distance measure can be found in [10].

5 The Impact of Pattern Complexity on Image Registration

It is clear that the performance of the registration algorithms described in Section 4.2 depends on the number of feature points to be registered. However, the complexity of the image may also affect algorithm performance.

In this section, we examine the impact of the complexity of an image on the objective function and the algorithm performance. For all the experiments described in this section, we fixed the number of points in each image to be 500.

5.1 Impact on objective function

Two extreme cases are considered, one with an asymmetric image, and the other with a highly symmetric image. Figure 10 is an example of an asymmetric image. This image consists of 500 data points on 8 line segments; most of the lines are not parallel to each other. In this image, the gaps between points on the line segments are varying, making the line segments asymmetric. For this data set, we can anticipate that the objective function should converge to the global minimum smoothly (there would not be many local minima). To show the six-dimensional objective function, we fixed five parameters while varying one parameter around the optimal solution.

Figure 11 shows the impact of changes in the first four parameters of the affine transformation. The impact of changes in the two translation parameters is shown in Figure 12. As we anticipated, there are very few local minima, making it faster for the algorithm to converge to the global minimum.

Figure 13 is an example of a symmetric image. This image consists of 500 data points with 50 parallel line segments. In this case, we fix the gap between points to be constant. For this image, we can anticipate that there are many local minima in the objective function, because if we translate the image by the distance between the points/lines, this results in another good match (even though not as good as the global minimum). Therefore, the objective function will have some periodic structure in the translation direction. Figure 15 shows this behavior of the objective function. For both x and y translations, there are periodic local minima in the objective function. In fact, the periods correspond to the distances between points in the two directions. For the other affine parameters, similar behavior is observed, causing the objective function to have many local minima as the parameters change. This behavior is shown in Figure 14.

5.2 Impact on algorithm performance

The shape of the objective function affects the performance of the algorithm. There are several algorithms that can find the global optimum when there are numerous local minima. However, if the objective function has many local minima, these algorithms have difficulty in finding the global one. When we ran the branch-and-bound algorithm described in Section 4.2.3, it took 39 seconds on the asymmetric image of Figure 10, and 138 minutes on the symmetric image of Figure 13.

Table 1 shows the running times of the branch-and-bound algorithm when applied to

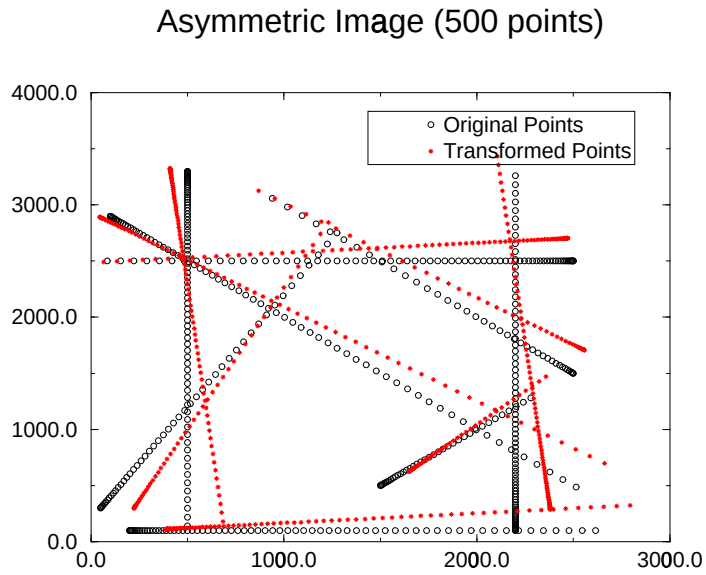


Figure 10: Layout of asymmetric image.

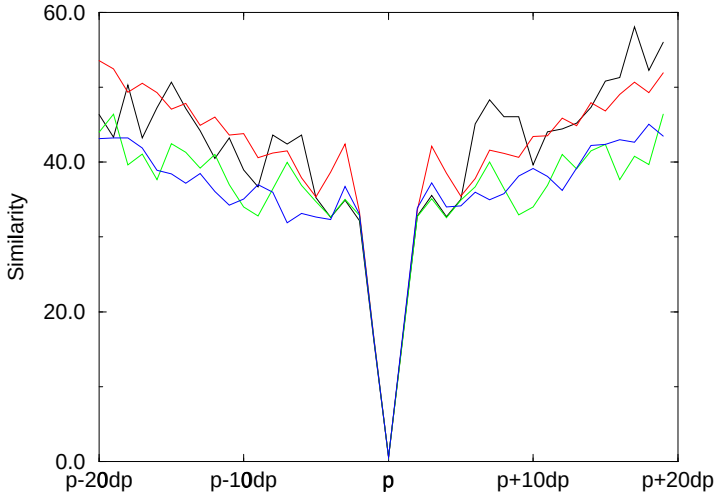


Figure 11: Objective function of asymmetric image.

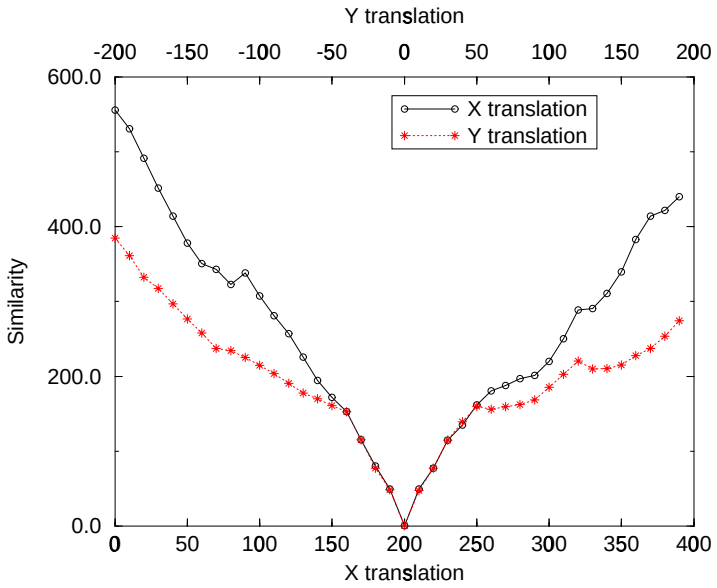


Figure 12: Objective function of asymmetric image (translation).

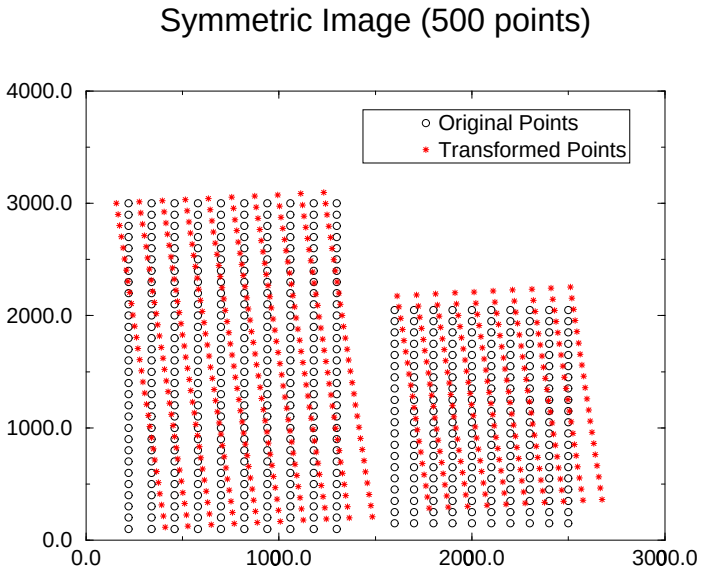


Figure 13: Layout of symmetric image.

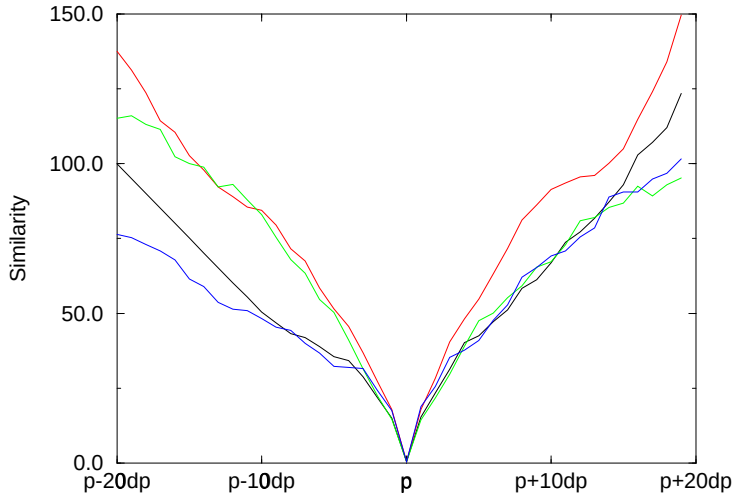


Figure 14: Objective function of symmetric image.

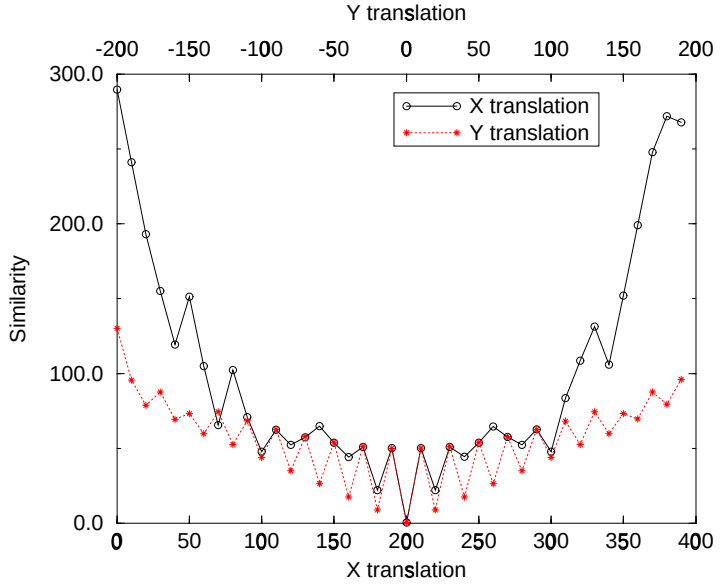


Figure 15: Objective function of symmetric image (translation).

Image type	Number of lines	Gap type	Running time
Asymmetric	8	Variable	39 sec.
Asymmetric	8	Constant	51 sec.
Symmetric	8	Variable/diff. direction	54 sec.
Symmetric	8	Variable/same direction	98 sec.
Symmetric	50	Variable	68 min.
Symmetric	50	Constant	138 min.

Table 1: Timing information on images with various complexities.

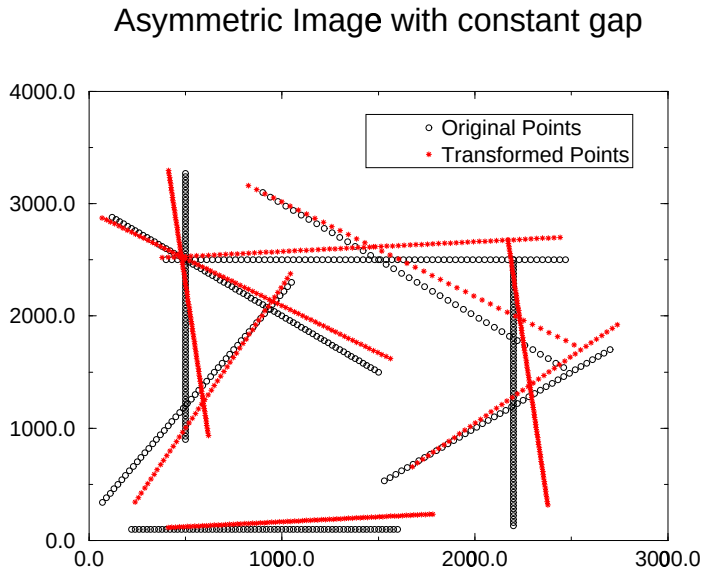


Figure 16: Asymmetric image with constant gaps.

images with various types of complexity. Figures 16–18 show the layout of these images. From this timing information, we can see that as the image becomes more symmetric, the running time for registration increases. In many cases, document images are highly symmetric, having similar layout to that in Figure 13. This fact tells us that registration of document images usually takes more time than for more asymmetric images, such as satellite images and video images.

6 Attributed Point Matching

To improve algorithm performance, we introduce the notion of attributes of feature points into the similarity measure. Attributes can be color, area, width, height, aspect ratio, or number of black pixels. The similarity measure is now a function of the distance between the points as well as the similarity between their attributes. We use the number of black pixels as an attribute of the feature points. As discussed in Chapter 4, a feature point represents a group of connected components. Therefore, we can count the number of black pixels in each group of connected components.

Now we need to define the similarity measure for the attribute. Let Δn_b be the differ-

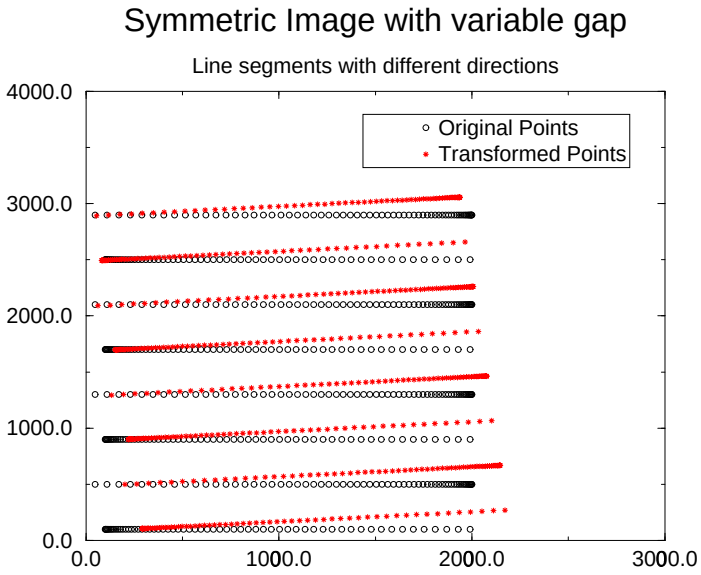


Figure 17: Symmetric image with variable gaps (different directions).

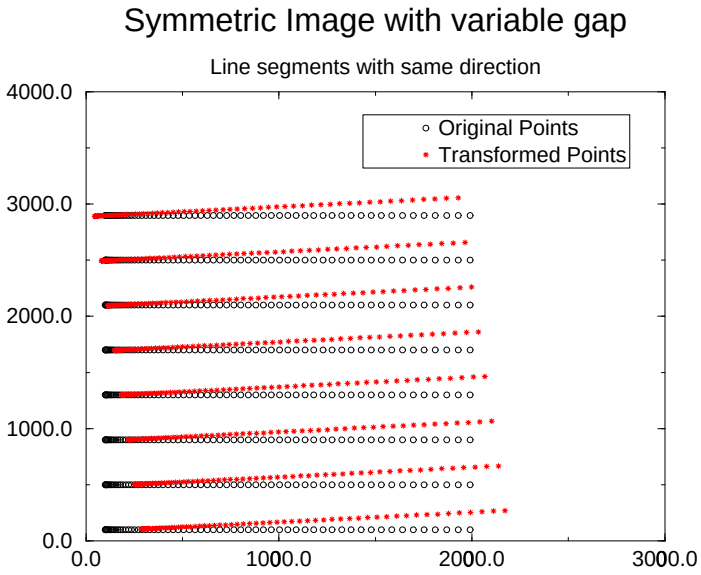


Figure 18: Symmetric image with variable gaps (same direction).

ence between the numbers of black pixels in two feature points, and d be the Euclidean distance between them. Then the new similarity sim_a is defined to be

$$sim_a = p\frac{1}{\lambda_1}\exp(-\frac{\Delta n_b}{\lambda_1}) + (1-p)\frac{1}{\lambda_2}\exp(-\frac{d}{\lambda_2}),$$

$$\text{where } \lambda_1 = E[\Delta n_b], \lambda_2 = E[d] \text{ ,and } 0 \leq p \leq 1.$$

By changing p we can control the weight of the attribute. For example, if we use only the distance, we can set p to be 0, so that the first term of sim is 0. When the distance is 0, the similarity is also 0, and when the distance goes to infinity, the similarity approaches 1.

Instead of partial Hausdorff distance, we use the new attributed similarity as the similarity measure for Mount et al.'s algorithm described in Section 4.2.3. In Figures 19–22 we show the behavior of the algorithm for the two similarity measures. The image contains 30 randomly generated points. We then remove 10% of the points, introduce the same number of outlier points, and transform the image with a 5° rotation and an x translation of 50. The running time for partial Hausdorff distance is 41 seconds, whereas it takes 26 seconds for attributed similarity. For comparison, we multiply the attributed similarity by 100 so that the similarity has the range [1,100] instead of [0,1]. Figure 19 is the graph of best similarity at each iteration. We observe that the attributed similarity decreases faster than the partial Hausdorff distance.

In Figure 20 we compare the maximum size of the cell at each iteration for two similarity measures. The attributed measure also decreases faster in this case. The number of active cells is important in terms of system resources. The maximum number of active cells represents the memory usage of the algorithm. As we observe in Figure 21, the maximum number of active cells for attributed similarity is less than half that for partial Hausdorff distance. Figure 22 shows the best similarity as a function of the search tree level. We observe that they are similar to each other, and therefore we can suppose that in both cases they take similar paths in the search tree to reach the optimal solution.

7 Error Metric and Experimental Protocol

7.1 Error metric

For the analysis of the experimental results, we need to define an error criterion. Let G be the set of groundtruth elements $g_i, i = 1, \cdots, N,$, where N is the number of characters in the image. Typically, g_i is a tuple: $g_i = (x_i, y_i, w_i, h_i, f_i) \in R \times R \times R^+ \times R^+ \times \mathcal{F}$, where, x_i, y_i are the x- and y-coordinates of the upper-left corner of the character-level bounding box, w_i, h_i are the width and height of that bounding box, and f_i is the font. Let θ and $\hat{\theta}$ denote the true and estimated transformations respectively. We can get the groundtruth for the rescanned image by transforming G using the estimated transformation. Then we can define G^θ and $G^{\hat{\theta}}$ to be the set of transformed groundtruth elements as follows:

$$G^\theta = T^\theta(G) \text{ with elements } g_i^\theta = (x_i^\theta, y_i^\theta, w_i^\theta, h_i^\theta, f_i^\theta)$$
$$G^{\hat{\theta}} = T^{\hat{\theta}}(G) \text{ with elements } g_i^{\hat{\theta}} = (x_i^{\hat{\theta}}, y_i^{\hat{\theta}}, w_i^{\hat{\theta}}, h_i^{\hat{\theta}}, f_i^{\hat{\theta}}).$$

We can compute g_i^θ and $g_i^{\hat{\theta}}$ as follows:

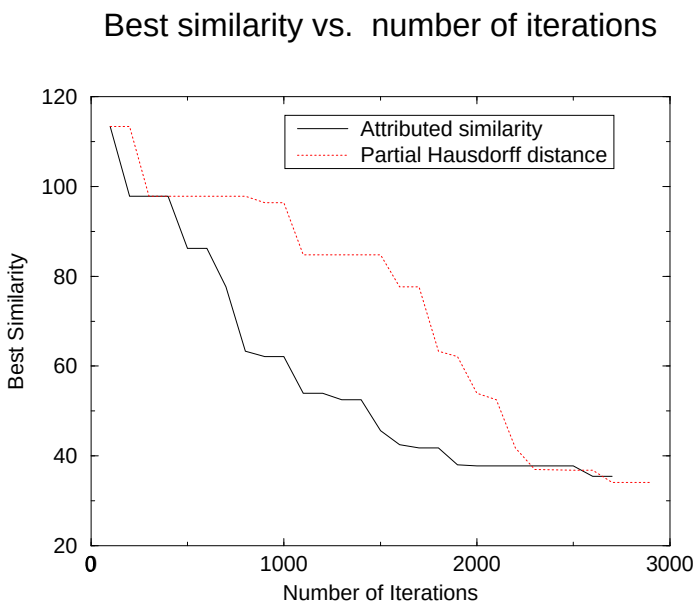


Figure 19: Best similarity vs. number of iterations

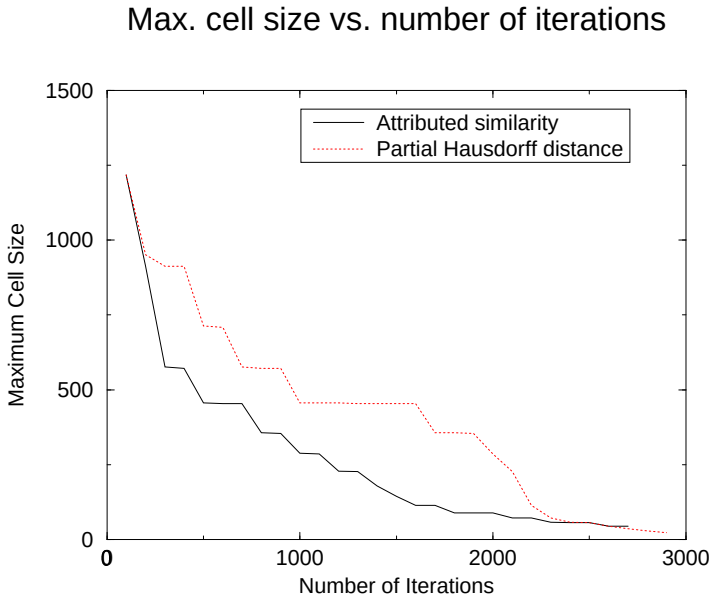


Figure 20: Maximum cell size vs. number of iterations

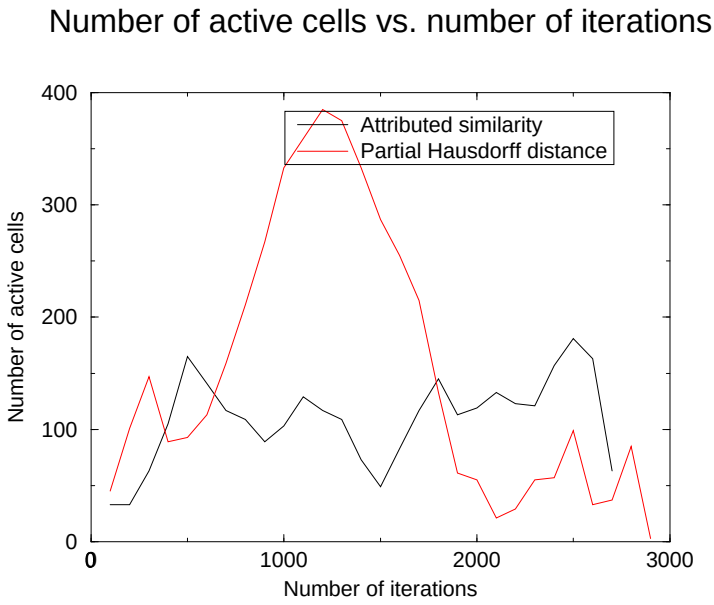


Figure 21: Number of active cells vs. number of iterations

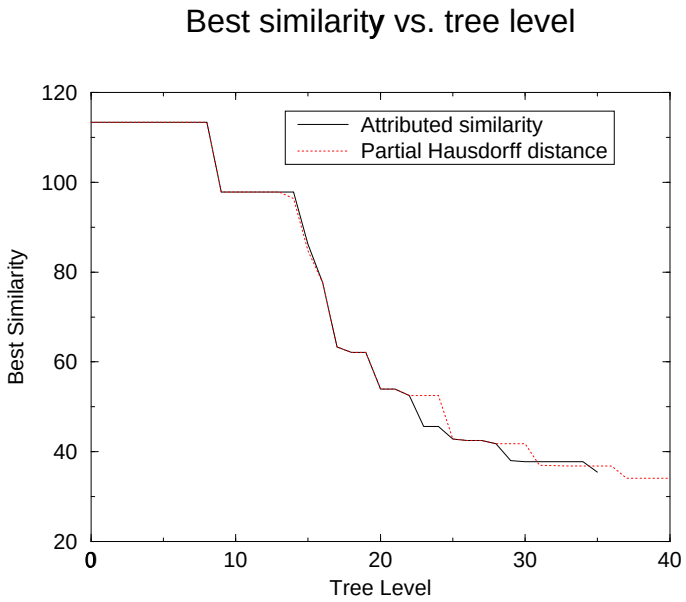


Figure 22: Best similarity vs. tree level

$$(x_i^\theta, y_i^\theta)^t = T^\theta(x_i, y_i)^t, (x_i^{\hat{\theta}}, y^{\hat{\theta}})^t = T^{\hat{\theta}}(x_i, y_i)^t.$$

To define w_i^θ, h_i^θ and $w_i^{\hat{\theta}}, h_i^{\hat{\theta}}$, let u_i, v_i be the x- and y-coordinates of the lower-right corner of the bounding box:

$$u_i = x_i + w_i,\ v_i = y_i + h_i$$
$$(u_i^\theta, v_i^\theta)^t = T^\theta(u_i, v_i)^t, (u_i^{\hat{\theta}}, v^{\hat{\theta}})^t = T^{\hat{\theta}}(u_i, v_i)^t$$
$$w_i^\theta = u_i^\theta - x_i^\theta,\ h_i^\theta = v_i^\theta - y_i^\theta$$
$$w_i^{\hat{\theta}} = u_i^{\hat{\theta}} - x_i^{\hat{\theta}},\ h_i^{\hat{\theta}} = v_i^{\hat{\theta}} - y_i^{\hat{\theta}}.$$

Also, we assume that $f_i^\theta = f_i^{\hat{\theta}} = f_i$. The Euclidean distance between the centroids of the corresponding bounding boxes δ_i is defined as

$$\delta_i = \|Centroid(g_i^\theta), Centroid(g_i^{\hat{\theta}})\|.$$

Then the mean and maximum error measures for an image can be defined as follows:

$$\rho_{mean}(G^\theta, G^{\hat{\theta}}) = \tfrac{1}{N}\textstyle\sum_{i=1}^{n} \delta_i$$
$$\rho_{max}(G^\theta, G^{\hat{\theta}}) = \max_i\{\delta_1, \cdots \delta_N\}$$

7.2 Experimental methodology and protocol

Our experiment was performed on the University of Washington data set [22]. This data set contains journal images with character-level geometric groundtruth. We performed two experiments, one on non-rotated images and the other on rotated images.

The experiment on non-rotated images was performed on 450 images. These images were generated by transforming 10 randomly selected images from the University of Washington data set by 45 different transformations. The rotation angle R was set at zero and the scale S and translation X_t, Y_t parameters were selected from the following sets:

$$S = \{65\%, 80\%, 100\%, 120\%, 135\%\},$$
$$X_t = \{-50, 0, 50\},\ Y_t = \{-100, 0, 100\}.$$

The initial search space was $60\% \sim 140\%$ for scale, $-100 \sim 100$ for X translation, and $-200 \sim 200$ for Y translation.

For the experiment on rotated images, we generated another 450 images from the same 10 images. For each image, we have 45 different transformations described as follows: We choose the scale parameter value from the set

$$S = \{65\%, 80\%, 100\%, 120\%, 135\%\},$$

rotation from the set

$$R = \{0^\circ, 1^\circ, 3^\circ\},$$

and the X, Y translations from the set

$$(X_t, Y_t) = \{(0, 0), (50, 0), (100, 0)\}.$$

The initial search space was $60\% \sim 140\%$ for scale, $-10^\circ \sim 10^\circ$ for rotation, $-100 \sim 100$ for X translation and $-200 \sim 200$ for Y translation.

8 Results and Discussion

In this section we describe the results of our controlled experiments. We used the branch-and-bound method described in Section 4.2.3 for feature point registration.

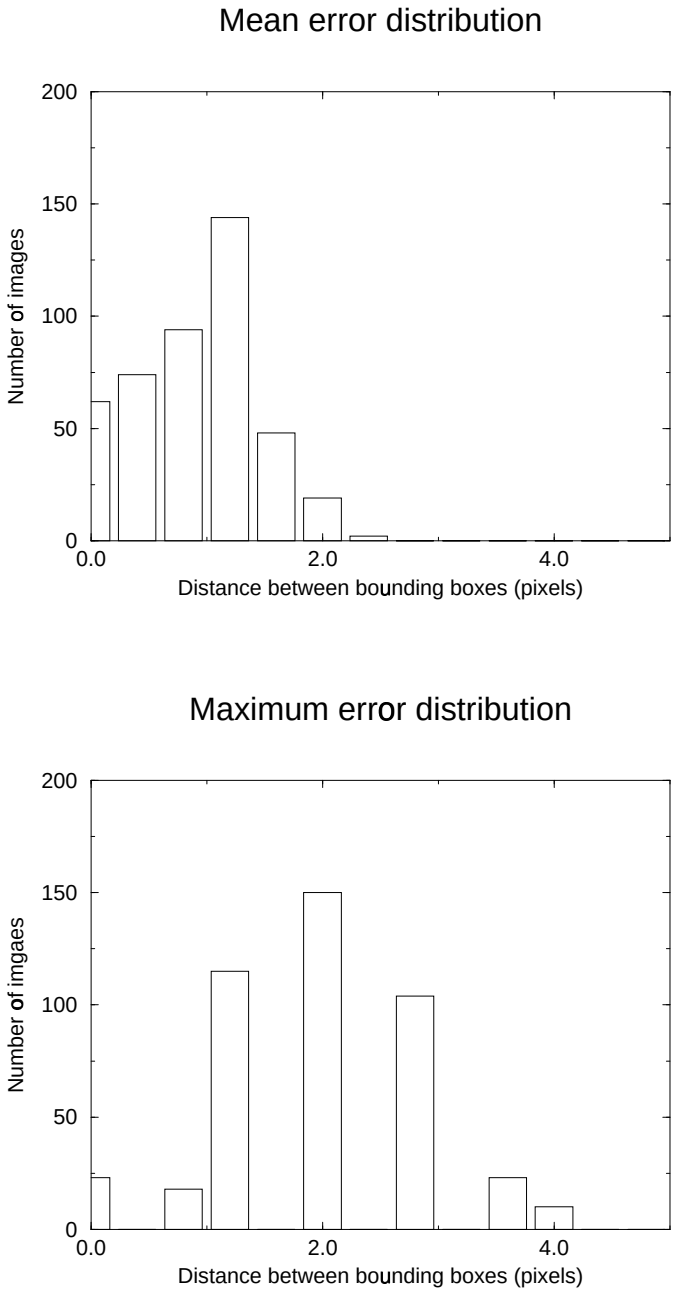


Figure 23: Distributions of mean and maximum errors.

8.1 Experiments on non-rotated images

To analyze the results, we generate the histogram of estimation errors. As discussed in Section 7.1, we calculate $\rho_{mean}(G^{\theta}, G^{\hat{\theta}})$ and $\rho_{max}(G^{\theta}, G^{\hat{\theta}})$ for each image pair. For the set of images O, the number of images that had errors in the range Δ is counted. The following is the notation for this analysis. Let O be the set of images, T be the set of transformations, Δ be the width of the range, I be the set of transformed images, and $\mathcal{G}$ be the set of groundtruth elements G_i. The histograms of the mean and maximum error, $H_{mean}(k; O, T, \Delta)$ and $H_{max}(k; O, T, \Delta)$, are defined as follows:

$$H_{mean}(k; O, T, \Delta) = \|\{i \in I \mid \frac{(k-1)\Delta}{2} < \rho_{mean}(G_i^{\theta}, G_i^{\hat{\theta}}) \leq \frac{(k+1)\Delta}{2}\}\|$$

$$H_{max}(k; O, T, \Delta) = \|\{i \in I \mid \frac{(k-1)\Delta}{2} < \rho_{max}(G_i^{\theta}, G_i^{\hat{\theta}}) \leq \frac{(k+1)\Delta}{2}\}\|$$

We have 450 transformed images for which groundtruth is estimated. The histograms of the mean and maximum error distributions of this image set are shown in Figure 23. We set Δ to be 0.4 pixel.

From the results, we see that the estimated groundtruth is close to the true groundtruth with less than 3 pixels of mean error and 5 pixels of maximum error. The mean of the mean error is 1.09 pixels, and the mean of the maximum error is 2.16 pixels. The estimation takes $10 \sim 15$ minutes per image when run on a Sun Ultra-Sparc 5 with clock speed 361.2 MHz.

8.2 Experiment on rotated images

The same methodology as that for the non-rotated images was used for the experiment on rotated images. Figures 24, 25, and 26 are the distributions of mean errors for the rotated images.

For non-rotated images, we have a similar result to that in Section 8.1, with most of the mean errors less than 3 pixels. However, for the images rotated by 1°, the mean errors become larger, about 40 pixels, and for 3° rotated images, the average of the mean errors is about 100 pixels.

9 Application: Registration for microfilmed and faxed images

9.1 Image registration for microfilmed images

In this section an experiment on microfilmed images is discussed. Assume that we are given a set of images with known groundtruth, and corresponding microfilmed images. We wish to generate the groundtruth for the microfilmed images from the available groundtruth.

In general, microfilmed images have the following features:

1. Large black areas around the image (similar to photocopied images)
2. A lot of small black pixels (so-called salt-and-pepper noise)

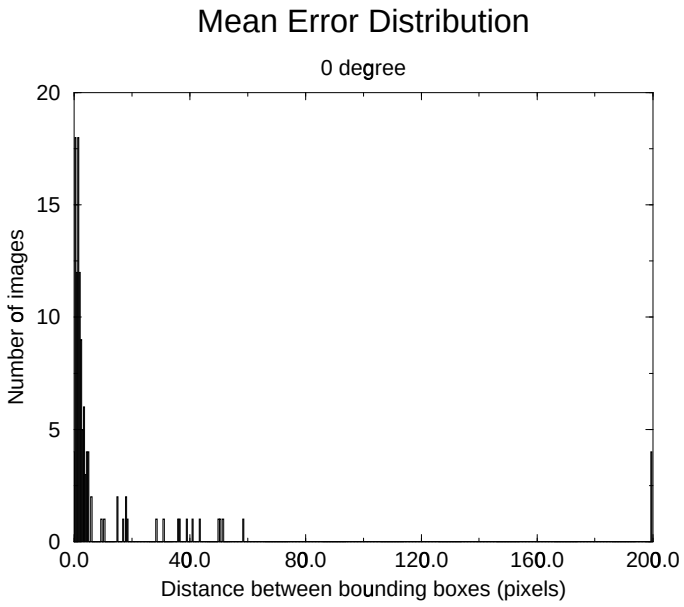


Figure 24: Distribution of mean errors for 0° rotated images.

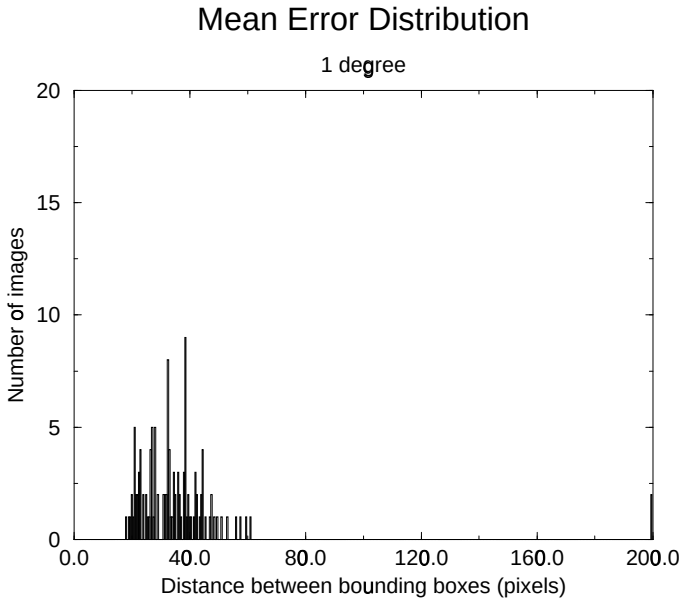


Figure 25: Distribution of mean errors for 1° rotated images.

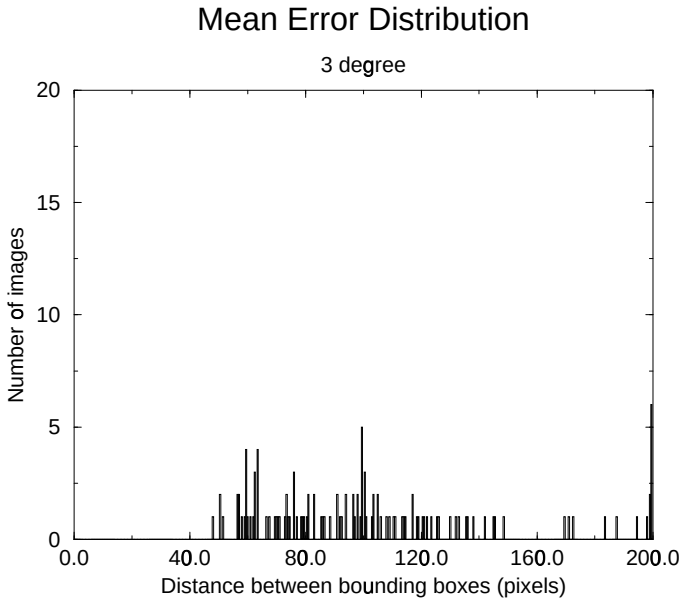


Figure 26: Distribution of mean errors for 3° rotated images.

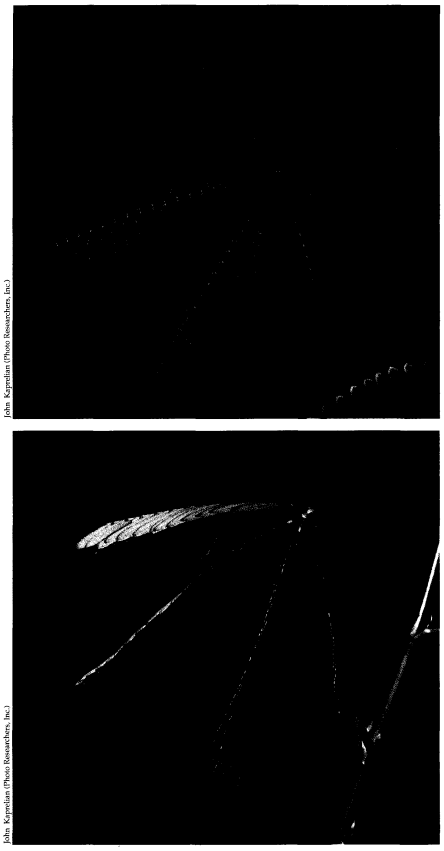
3. Many broken characters

4. Many merged characters.

Because of these features, it is helpful to filter out the connected components whose area is too small (case 2) or too large (case 1). Also, using feature point grouping as described in Section 4.1 helps, especially in cases 3 and 4. Consider the case in which many characters are broken apart in a microfilmed image (see Figures 27 and 28). In many cases, these broken parts are still very close to each other. In most cases the grouping algorithm regroups them.

Another case is when the characters are joined. In this case we have relatively large connected components. However, the grouped result will be similar to that of the original image, because in most cases, the joint characters are not larger than words. Therefore we still have reasonable feature points for the original and microfilmed images. This matters, because the registration algorithm is based on the feature points, and if we do not provide a good correspondence, it is obvious that the registration algorithm cannot give us a good result.

Figures 29 and 30 are corresponding original and microfilmed images. Figure 31 is the microfilmed image overlaid with the estimated groundtruth information. We used the methodology discussed in Section 3; the groundtruth is at the word and zone level in DAFS [8] format. The registration algorithm of Breuel [3], described in Section 4.2.2, was used. The experiment was conducted on the University of Washington III data set [22] with 978 images, and the corresponding microfilmed images. The registration took about 17 minutes per image when run on a Sun Ultra-Sparc 10 with clock speed 481.7 MHz.



John Kaprelian (Photo Researchers, Inc.)

John Kaprelian (Photo Researchers, Inc.)

Figure 14. ***Mimosa pudica*** **plays possum when touched. Normally, this plant stands upright** ***(top)*****. But when the plant is touched, an action potential causes the touched leaves and stems to droop and appear dead. The first action potential triggers action potentials in other parts of the plant, and these stems and leaves also droop; soon the entire plant appears to be dead** ***(bottom)*****.**

providing a secure trap. Then nearby secretory cells exude enzymes, forming a little stomach that digests the insect.

One of the best-known examples of plant behavior comes from *Mimosa pudica*, often called the sensitive plant. When the leaves of the plant are touched, they bend over and appear dead. The drooping arises from a mechanically driven action potential. Moreover, an action potential propagates from the stimulated region throughout the plant. This causes drooping in the rest of the plant, a defense mechanism apparently designed to make the whole plant look unappealing.

Not all plant action potentials, however, cause obvious responses. In *Luffa*—the plant whose gourd or fruit is used for "loofah" sponges—action potentials cause a transient inhibition of growth. And in a variety of flowers, pollen landing on the stigma generates an action potential, which may be involved in subsequent pollination or the maturation process. In tomato seedlings, a mechanical wound induces electrical activity that causes the accumulation of proteins that limit further damage to the plant.

Electrical phenomena control many responses in plants. In a characean alga, we understand many of the details of the mechanism that leads from a duck's nip on the plant to the cessation of protoplasmic streaming. But we are just beginning to address the similarities between the electrical excitability in characean algae and higher plants, let alone animals. In any case, it is apparent that plants can perform long-distance communication through electrical signals, such as the passing of information from a mechanical stimulus from one *Mimosa* stem to another. Many biologists continue to describe electrical excitability as part of the animal world. In the future, we should think of plants as excitable too.

Acknowledgments

Thanks to Drs. Atsushi Furuno, Owen Hamill, Roger Spanswick, Mark Staves, Robert Turgeon and Scott Wayne for their comments on this manuscript.

Bibliography

Barry, W. H. 1968. Coupling of excitation and cessation of cyclosis in *Nitella*: Role of divalent cations. *Journal of Cell Physiology* 72:153–160.

Beilby, M. J. 1984. Calcium and plant action potentials. *Plant, Cell and Environment* 7:415–421.

Blinks, L. R. 1936. The effect of current flow on bioelectric potential. III. *Nitella*. *Journal of Gen-*

Figure 27: Original image.

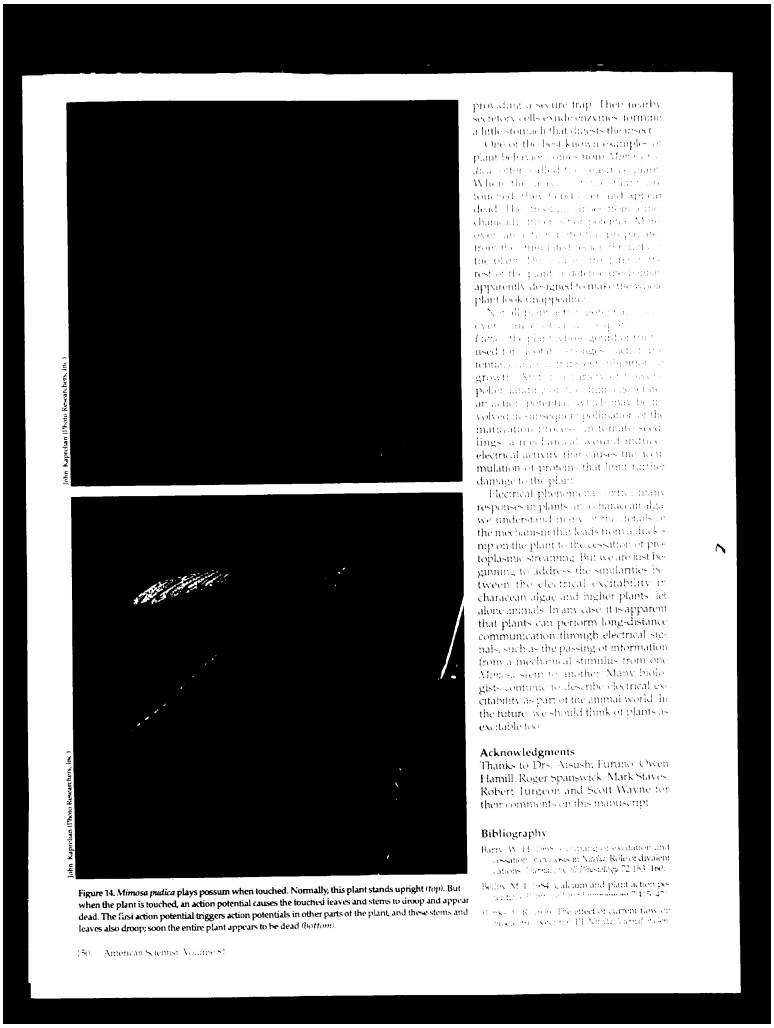


Figure 28: Microfilmed image with broken connected components.

other adaptations. Hence, there is no more reason to believe that the brain is a *tabula rasa* than to believe that the stomach is a general digester designed to track the foods an organism may encounter.

Differences in research strategies

In its pure form, DA focuses on differences in LRS between individuals encountering different environments, and uses the methods of behavioural ecology to study these differences. EP, in its purest form, uses the methods of evolutionary biology and experimental psychology to study the naturally selected design of psychological mechanisms. Consider how these two types of researcher might approach testing the Trivers–Willard[13] hypothesis about the allocation of parental investment to male and female progeny.

Trivers and Willard argued that if (1) variance of male LRS exceeded that of female LRS, (2) the relative health and dominance of mothers is passed on to their progeny, and (3) healthy or dominant males obtain more matings than males lacking these attributes, then (4) females will be selected to allocate investment in progeny as a function of their health or dominance. Clutton-Brock *et al.*[14], in a comprehensive study of red deer (*Cervus elaphus*), found considerable support for the hypothesis. Sons born to mothers above median rank were more reproductively successful than their daughters, while daughters born to subordinate mothers were more reproductively successful than their sons. Moreover, the ratio of sons to daughters produced by dominant mothers was higher than for subordinate mothers. Because the sex ratio and reproductive success were key dependent variables in this study, it is similar to some studies of sex allocation done by DAs and described by Sieff[15].

An evolutionary psychologist attempting to test the Trivers–Willard hypothesis would first construct a *selection model* relating sexual dimorphism in variance in reproductive success in males and females and health or status of mother to the benefits of differential investment in sons and daughters[16]. Varying the parameters of the model would provide a description of how sex allocation *might have been selected* for in a particular species. The model would be used in conjunction with information about the natural history of the species to explore the parameter space of the independent variables to determine whether a 'window' of opportunity could have existed for the evolution of the putative adaptation. If the results of the modelling suggested that the evolution of the adaptation is plausible, a theory of the nature of the adaptation, specified in terms of decision rules assumed to be instantiated in neural hardware, would be formulated. The dependent variables would be outputs from the decision process affecting nursing time, amount of protection from predators, etc., given to sons and daughters, rather than fitness measures or behaviours assumed to enhance fitness. Attitudes, values, intentions and motives would be measured in human studies. A decision rule might be something like: 'If subordinate and physically weak, be more responsive to the needs of daughters than of sons; but if strong and dominant be more attentive to the needs of sons than of daughters'. It would be necessary to formulate a theory of the relation between ancestral and current environments.

Such a theory requires a model of how the crucial independent variables, which are measures of adaptation-relevant external and internal environmental variables, are represented to the ancestral adaptation. Dominance, for example, might have been represented in terms of posture, frequency of unreciprocated threat displays, or resources held by different ancestral individuals. Once the decision rules that describe the adaptation

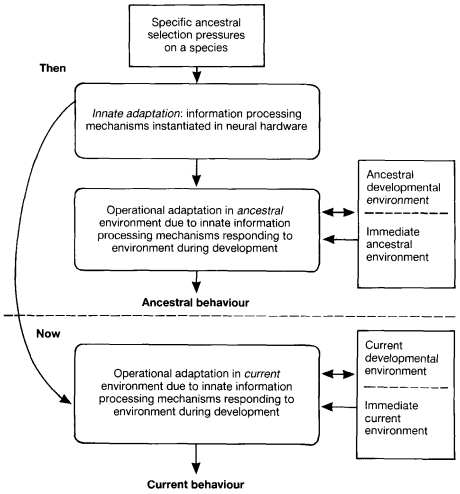
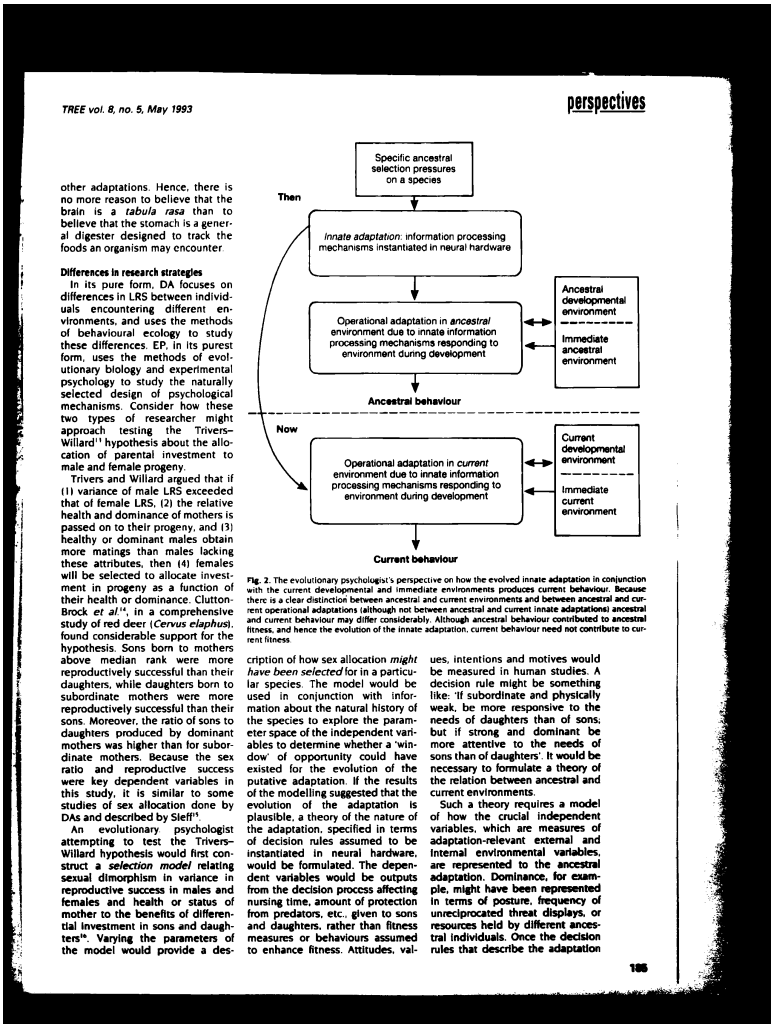


Fig. 2. The evolutionary psychologist's perspective on how the evolved innate adaptation in conjunction with the current developmental and immediate environments produces current behaviour. Because there is a clear distinction between ancestral and current environments and between ancestral and current operational adaptations (although not between ancestral and current innate adaptations) ancestral and current behaviour may differ considerably. Although ancestral behaviour contributed to ancestral fitness, and hence the evolution of the innate adaptation, current behaviour need not contribute to current fitness.

Figure 29: Original image to be registered.



TREE vol. 8, no. 5, May 1993

perspectives

other adaptations. Hence, there is no more reason to believe that the brain is a *tabula rasa* than to believe that the stomach is a general digester designed to track the foods an organism may encounter.

Differences in research strategies

In its pure form, DA focuses on differences in LRS between individuals encountering different environments, and uses the methods of behavioural ecology to study these differences. EP, in its purest form, uses the methods of evolutionary biology and experimental psychology to study the naturally selected design of psychological mechanisms. Consider how these two types of researcher might approach testing the Trivers–Willard[11] hypothesis about the allocation of parental investment to male and female progeny.

Trivers and Willard argued that if (1) variance of male LRS exceeded that of female LRS, (2) the relative health and dominance of mothers is passed on to their progeny, and (3) healthy or dominant males obtain more matings than males lacking these attributes, then (4) females will be selected to allocate investment in progeny as a function of their health or dominance. Clutton-Brock *et al.*[14], in a comprehensive study of red deer (*Cervus elaphus*), found considerable support for the hypothesis. Sons born to mothers above median rank were more reproductively successful than their daughters, while daughters born to subordinate mothers were more reproductively successful than their sons. Moreover, the ratio of sons to daughters produced by dominant mothers was higher than for subordinate mothers. Because the sex ratio and reproductive success were key dependent variables in this study, it is similar to some studies of sex allocation done by DAs and described by Sieff[15].

An evolutionary psychologist attempting to test the Trivers–Willard hypothesis would first construct a *selection model* relating sexual dimorphism in variance in reproductive success in males and females and health or status of mother to the benefits of differential investment in sons and daughters[16]. Varying the parameters of the model would provide a description of how sex allocation *might have been selected* for in a particular species. The model would be used in conjunction with information about the natural history of the species to explore the parameter space of the independent variables to determine whether a 'window' of opportunity could have existed for the evolution of the putative adaptation. If the results of the modelling suggested that the evolution of the adaptation is plausible, a theory of the nature of the adaptation, specified in terms of decision rules assumed to be instantiated in neural hardware, would be formulated. The dependent variables would be outputs from the decision process affecting nursing time, amount of protection from predators, etc., given to sons and daughters, rather than fitness measures or behaviours assumed to enhance fitness. Attitudes, values, intentions and motives would be measured in human studies. A decision rule might be something like: 'If subordinate and physically weak, be more responsive to the needs of daughters than of sons; but if strong and dominant be more attentive to the needs of sons than of daughters'. It would be necessary to formulate a theory of the relation between ancestral and current environments.

Such a theory requires a model of how the crucial independent variables, which are measures of adaptation-relevant external and internal environmental variables, are represented to the ancestral adaptation. Dominance, for example, might have been represented in terms of posture, frequency of unreciprocated threat displays, or resources held by different ancestral individuals. Once the decision rules that describe the adaptation

Fig. 2. The evolutionary psychologist's perspective on how the evolved innate adaptation in conjunction with the current developmental and immediate environments produces current behaviour. Because there is a clear distinction between ancestral and current environments and between ancestral and current operational adaptations (although not between ancestral and current innate adaptations) ancestral and current behaviour may differ considerably. Although ancestral behaviour contributed to ancestral fitness, and hence the evolution of the innate adaptation, current behaviour need not contribute to current fitness.

185

Figure 30: Microfilmed image to be registered.

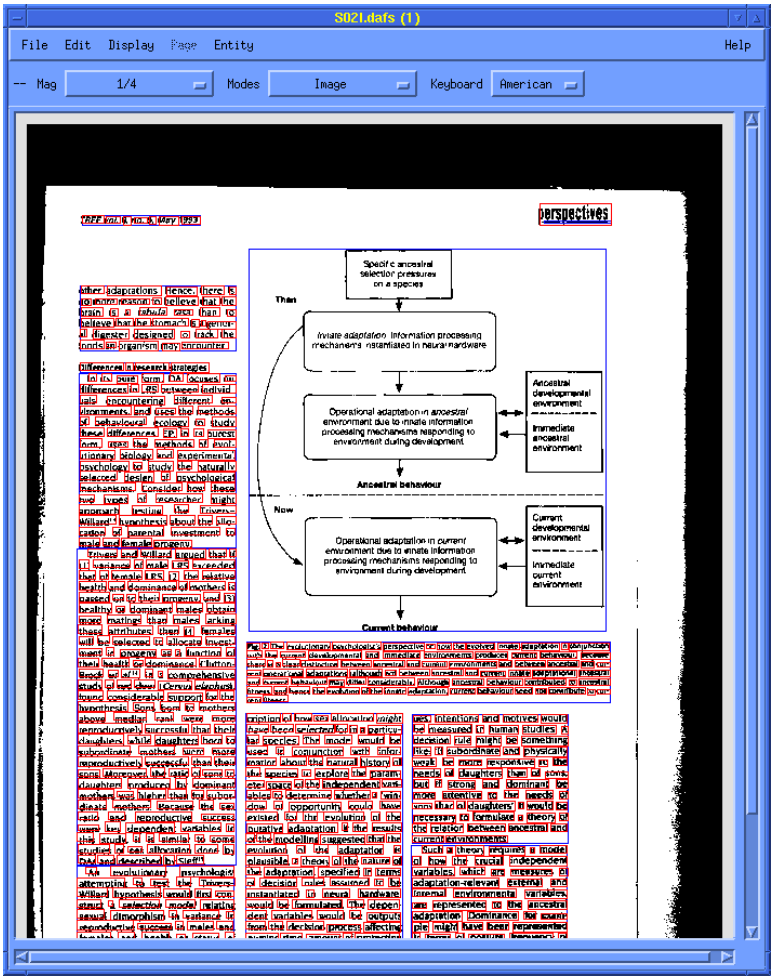


Figure 31: Microfilmed image overlaid with estimated groundtruth.

9.2 Experiment on a faxed image

In Section 9.1, we discussed a methodology for generating groundtruth for microfilmed images. The same methodology can be applied to other images such as photocopied or faxed images. We faxed and rescanned an image, and ran the feature point registration algorithm to produce the groundtruth for this image. Figure 32 shows the faxed image overlaid with the estimated groundtruth.

10 Conclusions

We have proposed an improvement over the automatic groundtruthing algorithm proposed by Kanungo and Haralick. We used feature point grouping to reduce the complexity of the problem. Then we used feature point registration algorithms on the grouped feature point sets to estimate the transformation between two images. To analyze the result of a controlled experiment, we defined the error metric to be the Euclidean distance between the centroids of corresponding characters. Further reduction in groundtruth location error can be achieved by using the local template matching algorithm described by Kanungo and Haralick [13, 14].

The contributions of this paper are:

- We made the image registration process more robust by using all the feature points available from both the original and transformed images. Several point matching algorithms were discussed and used for document image registration.
- We studied the impact of pattern complexity on the registration process. By observing the behavior of the objective function, we found that registration takes more time on symmetric images than on asymmetric ones.
- We also studied attributed point matching. Each feature point can have an attribute, such as color, area, width, height, aspect ratio, or number of black pixels. This attribute can be introduced into the similarity measure to make registration faster and more accurate. We used the number of black pixels as an attribute, and found the best similarity and maximum cell size at each iteration, as well as the number of active cells at each iteration.
- We used our algorithm to create groundtruth for scanned microfilm images and faxed images.

References

[1] H. S. Baird. *Model-Based Image Matching Using Location.* MIT Press, Cambridge, Mass., 1985.

[2] J. L. Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18:509–517, 1975.

[3] T. M. Breuel. Fast recognition using adaptive subdivisions of transformation space. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 445–451, Champaign, Illinois, June 1992.

ROBOTEX: An Autonomous Mobile Robot for Precise Surveying *

Xavier LEBEGUE and J. K. AGGARWAL
Computer and Vision Research Center,
Dept. of Electrical and Computer Engr.,
The University of Texas at Austin,
Austin, Texas 78712-1084, U.S.A.

Abstract. The RoboTex project aims at automatically constructing an exact CAD representation of buildings using a mobile robot. This paper reports on the current status of the project. The hardware of the robot is described, with special emphasis on issues relating to measurement accuracy, and algorithms used to process the sequences of monocular images acquired by the robot are presented. Results of automatic indoor surveying are shown and compared to direct measurements in the scene. The techniques developed here have important applications in architectural surveying, scene understanding, and precise robot navigation.

1 Introduction

This paper describes RoboTex, a mobile robot especially designed for building accurate 3-D maps of its environment. The goal of the RoboTex project is to enable a robot to automatically explore a building to construct a very accurate CAD representation. This CAD representation should be as close as possible to what an architect would generate.

Traditionally, the tasks of a robot's perception system are to detect obstacles, find the free space, and estimate the position of the robot in the world. Here, the focus is on building a useful 3-D description of the world. Our 3-D representation of the environment differs primarily from representations used by other robots in that:

1. It must concentrate on *semantically significant* features.

2. It must be more accurate than is strictly necessary for navigation alone.

To satisfy the first constraint, we chose to concentrate on straight edges with particular orientations in the 3-D scene. Typically, there are three prominent 3-D orientations in indoor scenes and outdoor urban scenes: the vertical, and two horizontal orientations perpendicular to each other. Our approach considers only polyhedral objects with such edges. This assumption holds for most large architectural features such as walls, doorways, floors, and ceilings. The second constraint, accuracy, has multiple implications for both the hardware and the software of the robot.

*This research was supported in part by the DoD Joint Services Electronics Program through the Air Force Office of Scientific Research (AFSC) Contract F49620-89-C-0044, and in part by the Army Research Office under contract DAAL03-91-G-0050.

Figure 32: Estimated groundtruth of faxed image.

[4] L. G. Brown. A survey of image registration techniques. *ACM Computing Surveys*, 24:325–376, 1992.

[5] R. G. Casey and D. R. Ferguson. Intelligent forms processing. *IBM Systems Journal*, 29:435–450, 1990.

[6] D. S. Doermann and A. Rosenfeld. The processing of form documents. In *Proceedings of International Conference on Document Analysis and Recognition*, pages 497–501, Tsukuba, Japan, August 1993.

[7] J. H. Friedman, J. L. Bentley, and R. A. Finkel. An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software*, 3:209–226, 1977.

[8] T. Fruchterman. DAFS: A standard for document and image understanding. In *Proceedings of SDIUT*, pages 94–100, Bowie, MD, October 1995.

[9] R. Haralick and L. Shapiro. *Computer and Robot Vision.* Addison-Wesley, Reading, Mass., 1992.

[10] J. D. Hobby. Matching document images with ground truth. *International Journal on Document Analysis and Recognition*, 1, 1998.

[11] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:850–863, 1993.

[12] D. P. Huttenlocher and W. J. Rucklidge. A multi-resolution technique for comparing images using the Hausdorff distance. Technical Report TR 92-1321, Department of Computer Science, Cornell University, 1992.

[13] T. Kanungo and R. Haralick. Automatic generation of character goundtruth for scanned documents : A closed loop approach. In *Proceedings of IAPR International Conference on Pattern Recognition*, pages 669–675, Vienna, Austria, August 1996.

[14] T. Kanungo and R. Haralick. An automatic closed-loop methodology for generating character groundtruth for scanned documents. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:181–183, 1999.

[15] D. Kim and T. Kanungo. A point matching algorithm for automatic generation of groundtruth for document images. In *Proceedings of IAPR International Workshop on Document Analysis Systems*, pages 475–485, Rio de Janeiro, Brazil, December 2000.

[16] J. Kittler and J. Illingworth. On threshold selection using clustering criteria. *IEEE Transactions on Systems, Man, and Cybernetics*, 15:652–655, 1985.

[17] D. E. Knuth. *TEX: The Program.* Addison-Wesley, Reading, Mass., 1988.

[18] S. Kullback. *Information Theory and Statistics.* Wiley, New York, 1959.

[19] L. Lamport. *LATEX: A Document Preparation System.* Addison-Wesley, Reading, Mass., 2nd edition, 1994.

[20] D. Mount, N. Netanyahu, and J. LeMoigne. Efficient algorithms for robust point pattern matching and applications to image registration. *Pattern Recognition*, 32:17–38, 1999.

[21] J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.

[22] I. Phillips. *Users' Reference Manual.* CD-ROM, UW-III Document Image Database-III.

[23] R. Sedgewick. *Algorithms in C.* Addison-Wesley, Reading, Mass., 1990.

[24] V. Torczon. PDS: Direct search methods for unconstrained optimization on either sequential or parallel machines. Technical Report CRPC-TR92206, Rice University Center for Research on Parallel Computation, 1992.

[25] C. Viard-Gaudin, P. M. Lallican, S. Knerr, and P. Binter. The IRESTE On/Off(IRONOFF) dual handwriting database. In *Proceedings of International Conference on Document Analysis and Recognition*, pages 455–458, Bangalore, India, September 1999.