Registering multiview range data to create 3D computer objects

Article *in* IEEE Transactions on Pattern Analysis and Machine Intelligence · September 1995

DOI: 10.1109/34.400574 · Source: IEEE Xplore

Martin David Levine
McGill University

help to identify occlusion. The ASSET-2 system will be integrated with a structure-from-motion system which recovers world structure in a *static* environment. ASSET-2 will be used to segment out and track moving vehicles, and the static part of the scene will then be tracked by the structure-from-motion system.

REFERENCES

[1] S.M. Smith and J.M. Brady, "A scene segmenter; visual tracking of moving vehicles," *Engineering Applications of Artificial Intelligence*, vol. 7, no. 2, pp. 191-204, Apr. 1994.

[2] M. Irani, B. Rousso, and S. Peleg, "Detecting and tracking multiple moving objects using temporal integration," *Proc. Second European Conf. Computer Vision*, pp. 282-287, Springer-Verlag, 1992.

[3] F.G. Meyer and P. Bouthemy, "Region-based tracking using affine motion models in long image sequences," *Computer Vision, Graphics, and Image Processing: Image Understanding*, vol. 60, no. 2, pp. 119-140, Sept. 1994.

[4] P. Bouthemy and E. François, "Motion segmentation and qualitative dynamic scene analysis from an image sequence," *Int'l J. Computer Vision*, vol. 10, no. 2, pp. 157-182, 1993.

[5] S.M. Smith, "Feature based image sequence understanding," DPhil thesis, Robotics Research Group, Dept. of Eng. Science, Oxford Univ., 1992.

[6] J.M. Brady, "Seeds of perception," *Proc. Alvey Vision Conf.*, pp. 259-265, 1987.

[7] S.M. Smith, "A new class of corner finder," *Proc. Third British Machine Vision Conf.*, 1992.

[8] C.G. Harris and M. Stephens, "A combined corner and edge detector," *Fourth Alvey Vision Conf.*, pp. 147-151, 1988.

[9] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[10] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Proc. First Int'l Conf. Computer Vision*, pp. 259-268, 1987.

Registering Multiview Range Data to Create 3D Computer Objects

Gérard Blais and Martin D. Levine

***Abstract*—This research deals with the problem of range image registration for the purpose of building surface models of three-dimensional objects. The registration task involves finding the translation and rotation parameters which properly align overlapping views of the object so as to reconstruct from these *partial* surfaces, an *integrated* surface representation of the object.**

The approach taken is to express the registration task as an optimization problem. We define a function which measures the quality of the alignment between the partial surfaces contained in two range images as produced by a set of motion parameters. This function computes a sum of Euclidean distances between a set of control points on one of the surfaces to corresponding points on the other. The strength of this approach resides in the method used to determine point correspondences across range images. It is based on reversing the rangefinder calibration process, resulting in a set of equations which can be used to directly compute the location of a point in a range image corresponding to an arbitrary point in three-dimensional space.

A stochastic optimization technique, very fast simulated reannealing (VFSR), is used to minimize the cost function.

Dual-view registration experiments yielded excellent results in very reasonable computational time. A multiview registration experiment was also performed, but a large processing time was required. A complete surface model of a typical 3D object was then constructed from the integration of its multiple partial views. The effectiveness with which registration of range images can be accomplished makes this method attractive for many practical applications where surface models of 3D objects must be constructed.

***Index Terms*—Range, multiview, 3D, image registration, simulated annealing, surface models, suface integration, rangefinder calibration.**

I. INTRODUCTION

Given N views of an object in a scene, each one describing the 3D structure of the object as seen from a particular viewpoint. We wish to find N rigid motion transformations $T_1, T_2, \ldots, T_N$ that specify the true positions of the rangefinder with respect to a unique frame of reference (arbitrarily chosen and usually the frame of one of the views). Suppose that each range view $i(i = 1, \ldots, N)$ consists of a set of 3D points S_i expressed in the coordinate frame of the rangefinder. The transformation T_i transforms the points S_i of range image i into a new set of points $S_i' = T(S_i)$ in which the 3D coordinates of the points are expressed in a unique coordinate frame. By transforming the sets of points of all N range views, we can generate a new set of 3D points which is the union of all transformed sets $S_1', S_2', \ldots, S_N'$, namely

$$S = \bigcup_{i=1}^{N} T_i(S_i) = \bigcup_{i=1}^{N} S_i'$$

This new set of points represents the surface boundary model of the object defined by all the views.

A novel approach for solving this problem of range image registration is presented in this paper. It is a relatively simple method and robust to noise in the range data and positioning errors in the mechanical apparatus used for data acquisition. For example, the

Manuscript received Oct. 28, 1993; revised Mar. 27, 1995; recommended for acceptance by R. Nevatia.

Gérard Blais is with Bell-Northern Research, Verdun, Quebec, Canada; email: gblais@bnr.ca.

Martin Levine is with Center for Intelligent Machines & Dept. of Electrical Engineering, McGill University, Montreal, Quebec, Canada;
e-mail: levine@cim.mcgill.ca.

IEEECS Log Number P95098.

method is suitable for use with an eye-in-hand system. In this situation, the rangefinder is attached to the end of a robot manipulator which is generally known to be a rather sloppy absolute positioning device.

To achieve accurate registration, a cost function is defined which indicates the quality of registration between two range views by a sum of distances between corresponding points in each view. The range views are registered by determining the 3D rigid transformation which minimizes the cost function. The novelty of this approach is a method for reversing the calibration process of the rangefinder which permits point correspondences between range views to be computed directly. This results in an extremely fast method for computing the distance between range views, as required by the evaluation of the cost function. Stochastic search is used to find the transformation which minimizes the cost function in a reliable manner, even in the presence of the multiple local minima that always characterize the cost function.

The methods used for range image registration can be divided into two main categories. The first avoids the registration problem altogether by relying on precisely calibrated mechanical equipment to determine the motion transformation between views. The second category involves methods that derive the registration transformation between range images from the information contained in the range images and other information provided by the acquisition system [1], [3], [5], [6], [9], [12], [13]. In most cases, a coarse *estimate* of the transformation between each pair of range views is part of the available information.

In Section II, the essential aspects of our registration method are presented and registration is formulated as an optimization problem. Section III discusses the testing of the method by performing various registration experiments. Finally, Section IV offers some concluding remarks.

II. Registration Using Inverse Camera Calibration

A. Registration Method Overview

In this research, we have used range images of maximum size 256×256. In order to derive the 3D coordinates of each sampled surface point, the rangefinder is calibrated before data acquisition so that, given the index i and j in the rectangular array for a given point, and given the depth measured for the point, its coordinates (x, y, z) with respect to the camera's reference frame can be computed directly. If one thinks of each image point as being sampled by a different laser ray, then the indices i and j would specify which ray sampled each point.

The principal idea behind our method is to reverse this process so that the indices are computed from the coordinates by inverse calibration. This inverse calibration permits us to match points across range views. Given a transformation T from range image one to range image two, a 3D point (x, y, z) in range image 1 is transformed to (x', y', z') in image two's reference frame. Using the inverse calibration we are able to determine directly the indices (i, j) of the ray in image two closest (Euclidean distance) to that transformed 3D point. Since every point in a range image is obtained by sampling the surface with a different ray, the point in image two associated with the (i, j) ray is thus taken as the corresponding point.

To perform the registration, control points (pixels) are selected from the first range view by uniform subsampling. These are mapped by a rigid 3D transformation T into the second view's reference frame. Each transformed control point is then associated with a point in the other view. This point-to-point correspondence is directly established through the inverse calibration process. A distance measure, based on a sum of Euclidean distances between the transformed control points of the first view and their respective corresponding points in the second view, is computed. The objective is to find that transformation T which minimizes this distance measure. A transformation estimate T_e, obtained from the acquisition apparatus, is used to constrain the number of possible transformations. Thus, a finite search space is delimited around the estimate T_e. Only those transformations inside this search space are considered as potential solutions for the registration. The inverse calibration process is discussed in detail in [2].

By minimizing the sum of Euclidean distances between all control points in one view and their respective corresponding points in the other, the distance between these views is minimized. Since the sum of distances is a minimum when surface regions that are common to both views coincide, we can conclude that the views are registered. The sum of Euclidean distances is the basis for a cost function used by an optimization algorithm. This cost function will be described in detail in Section II.B.

B. Registration as Optimization

Let S_c be a set of control points taken from the total set of points in the first view. S_c, obtained by uniform subsampling, is a subset of all the sampled points in that view. Let T be the transformation which takes a point in the first view and expresses it in the reference coordinate frame of the second. If $\vec{\mathbf{p}}$ is a point in the first view, then $T(\vec{\mathbf{p}})$ is the same point expressed in the second view's coordinate frame. We specify a rigid 3D transformation by six motion parameters, consisting of three translations t_x, t_y and t_z (which define $\vec{\mathbf{d}}$) and three rotation angles r_x, r_y and r_z (which define $\mathbf{R}$). $T(\vec{\mathbf{p}})$ is simply given by

$$T(\vec{\mathbf{p}}) = \mathbf{R}\vec{\mathbf{p}} + \vec{\mathbf{d}} \tag{1}$$

Let $C(\,)$ be the correspondence function defined by the camera's inverse calibration equations. If $\vec{\mathbf{q}}$ is a point whose coordinates are expressed in the second view's coordinate frame, then $C(\vec{\mathbf{q}})$ is the point in the second view whose associated ray is closest to point $\vec{\mathbf{q}}$. The inputs to the function are the coordinates (x, y, z) of a 3D point. From these coordinates, the indices i and j of the closest ray in the second range image are found using the inverse calibration equations. The result is the 3D point in the range image at location (i, j). It is possible that the indices i and j found by the inverse calibration equations do not represent any valid point in the range image. This would be the case if the indices i and j represented a datum point that has been discarded during the preprocessing of the range images. Also, because a range image has a maximum of 256 by 256 sampled points, it is possible that the values of i and j computed from the inverse calibration equations are outside the allowable range of the indices, which must be between 1 and 256. In such cases, $C(\,)$ would return *undefined* as a result to indicate that no correspondence has been found. Given a transformation T, we define a cost function for T as follows (see [2] for details.):

$$cost(T) = \sum_{\vec{\mathbf{p}} \in S_c} d\left(T(\vec{\mathbf{p}}), C\left(T(\vec{\mathbf{p}})\right)\right) \tag{2}$$

where $d(\,)$ is the 3D Euclidean distance (L_2 norm) between two points. The cost function is an indication of the registration quality of the transformation T. The greater the accumulated distance between points in the views due to a transformation, the higher the cost of this transformation will be. Therefore, the transformation yielding the best registration of the range images will be the one with the lowest cost. An optimization search can then be applied to find the transformation T which minimizes this cost function.

When the function $C(T(p))$ finds no corresponding point for a given control point in S_c, the L_2 norm is undefined. In this case we arbitrarily set the distance function $d(\,)$ to 0.

Another issue is improper point correspondences between two range views; this can occur at the edges of the object. If the distance between the points in such a correspondence is very large, it will detrimentally affect the value of the cost function. Therefore, we limit the effect of the Euclidean distance by introducing a threshold τ as follows [7]:

$$d(\vec{p}_1,\vec{p}_2)=\begin{cases}\|\vec{p}_1-\vec{p}_2\| & ;\ \text{if } \vec{p}_2 \text{ is defined and } \|\vec{p}_1-\vec{p}_2\|\le\tau \\ \tau & ;\ \text{if } \vec{p}_2 \text{ is defined and } \|\vec{p}_1-\vec{p}_2\|>\tau \\ 0 & ;\ \text{if } \vec{p}_2 \text{ is undefined}\end{cases} \tag{3}$$

Because the distance measure returns 0 when no correspondence point is found, a transformation minimizing the number of correspondences would yield a minimum cost value. However, this is undesirable since very poor transformations will likely yield very few correspondences by definition. To alleviate this problem, the sum of the distances can be normalized by the number of correspondences. Let $S_c(T)$ be the set of all control points for which a correspondence exists under the transformation T. We redefine the cost function as follows:

$$cost(T)=\frac{\sum_{\vec{\mathbf{p}}\in S_c} d\left(T(\vec{\mathbf{p}}),C\left(T(\vec{\mathbf{p}})\right)\right)}{\|S_c(T)\|} \tag{4}$$

There still exists a problem with the cost function defined in this way. Because no penalty is assigned to transformations yielding few correspondences, the cost function will nevertheless be a minimum when no correspondences are established between views. This issue can be handled by enforcing an overlap factor Ω between views. The overlap generated by a transformation T is simply the total number of correspondences $\|S_c(T)\|$ divided by the total number of control points $\|S_c\|$. Because of the threshold τ, and because control points without correspondences result in a distance value for $d(\,)$ of 0, it is clear that no transformation can yield a cost value greater than the cardinality of S_c times the threshold value τ. This idea is used to set the maximum value of the cost function. With this in mind, we redefine the cost function as follows:

$$cost(T)=\begin{cases}\dfrac{\sum_{\vec{\mathbf{p}}\in S_c} d\left(T(\vec{\mathbf{p}}),C\left(T(\vec{\mathbf{p}})\right)\right)}{\|S_c(T)\|} & ;\ \text{if } \dfrac{\|S_c(T)\|}{\|S_c\|}\ge\Omega \\ \tau\|S_c\| & ;\ \text{if } \dfrac{\|S_c(T)\|}{\|S_c\|}<\Omega\end{cases} \tag{5}$$

Note that by establishing point-to-point correspondence across range images using the inverse camera calibration equations, we in fact compute the sum of the distances between the two range views in the direction of the rays. This is different from the intuitive way of evaluating distance, where the distance between a point on a surface to the other surface is taken as either the perpendicular distance or the distance to the closest point. One may argue that these give better indications of the distance between two views. However, as optimization progresses, and the registration between the two views improves, the distance along the scan lines will approach the perpendicular distance.

III. Experiments and Results

This section presents some of the experiments conducted in conjunction with this research. A brief description of the experimental setup and the data acquisition process is given in Section III.A.

Considerable experimentation was carried out to determine the characteristics of the cost function. For most range images tried, these usually indicated the presence of a single global optimum surrounded by multiple local optima. As a consequence, to ensure successful minimization for many different types of object surfaces, it became evident that we had to rely on a robust search method. A conventional gradient descent approach would be inadequate! Thus very fast simulated annealing (VFSR), a stochastic optimization method, was used to minimize the cost function [8]. See [2] for a detailed discussion of the search parameters used for VFSR and the values of these parameters yielding optimum performance for registration.

Section III.B. presents various dual-view registration experiments. Each consists of the registration of two range views obtained by sampling an object from two different viewing positions. Finally, a multiview registration experiment is presented in Section III.C. The notion of local/global optimization arising when registering multiple views of an object is examined and a solution is presented.

A. Experimental Setup and Data Acquisition

An eye-in-hand system was used for the acquisition of the range images. It consisted of a rangefinder camera attached to the end effector of a PUMA 560 robot arm. The robot is inverted and mounted on the ceiling to permit easier positioning of the camera for viewing objects at various angles.

An alternate method was a turntable. While maintaining the rangefinder camera in a fixed position, a precision turntable was used to accurately rotate objects so that sampling them from different viewpoints could be achieved. The position of the turntable can be specified as an absolute angular value in degrees.

We have determined by experimentation that the sampling error of the laser rangefinder is Gaussian distributed and that a linear relationship exists between the average sampling error (average of the absolute values) and the object distance [2]. In most experiments conducted for this research, data acquisition was performed with the rangefinder at a distance of around 40 centimeters from the object, sometimes more depending on its size and shape. At this distance the average error in the measured distance of a sampled point is approximately 0.625 millimeters.

With this in mind, a range surface can be seen as a perfect 3D representation of the surface of an object plus some added noise. The latter is Gaussian distributed with a standard deviation proportional to the object distance. When registering two range views, we therefore expect the minimum average Euclidean distance between corresponding points in each view to be twice the mean absolute sampling error (the errors in each view get added). The cost function computes an approximation to this average distance. Therefore, when two views are properly registered, we would expect the minimum cost function value to be around 1.25 millimeters, which is twice the mean absolute error for an object scanned at a distance of 40 centimeters.

Before registration, it was necessary to preprocess by computer each of the range images. This was done using simple segmentation to remove the background level and any spurious large impulses.

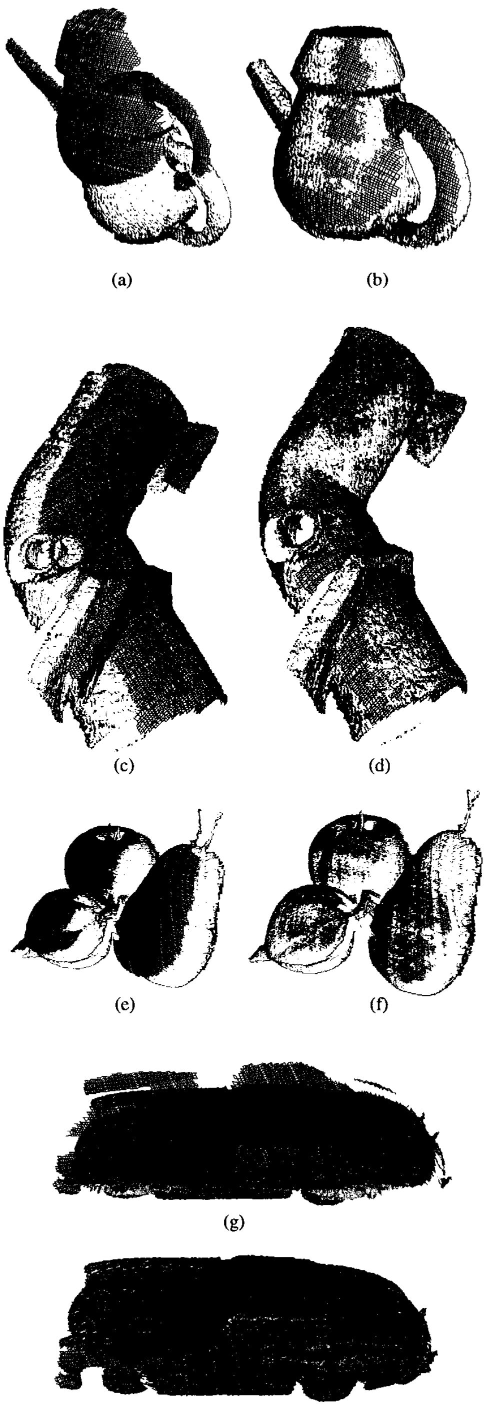

Fig. 1. Composite image of four different dual-view registration experiments. Figs. 1a and 1b show two views of a teapot before (related by the transformation estimate) and after registration (related by the optimum transformation found by the algorithm), respectively. Similarly, Figs. 1c and d, e and f, and g and h show the registration of two views of a metal pipe, fruits, and a model car, respectively.

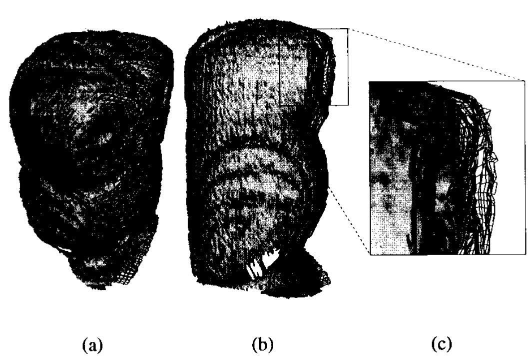

Fig. 2. This figure illustrates the problem arising when registering multiple views without performing global optimization. Local optimization is performed to register range views two at a time. In Fig. 2a we observe that, although the fit between consecutive views is optimized, the first (shown as a grid) and last view are poorly registered. Fig. 2b shows the same group of registered range views seen from the side, illustrating the gap between the views more clearly. Fig. 2c is an enlarged portion of Fig. 2b, emphasizing the large distance between the first and last view.

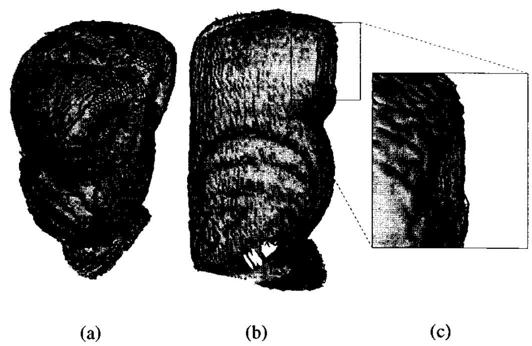

Fig. 3. This figure shows the results obtained when registering multiple views using global optimization. Six range views were registered simultaneously. The result is shown seen from above (Fig. 3a) and from the side (Fig. 3b). Fig. 3c is an enlargement of a section of Fig. 3b illustrating how closely the first (shown as a grid) and last view are registered.

B. Dual-View Registration Experiments

This section presents various experiments realized with the registration algorithm described previously. All experiments involve registering two range views obtained from different 3D objects.These are shown in Fig. 1. The range views of the metal pipe, the fruits and the model car (Figs. 1c, d, e, f, g, and h, respectively) were acquired with the eye-in-hand robot system. This indicates the ability of the registration algorithm to handle the usually large errors occurring in the initial transformation estimate obtained with such a positioning system.

C. Multiview Registration

The most straightforward way of performing the registration of multiple views of a 3D object is to register the views in pairs. However this is not globally optimal and even small errors at each registration tend to accumulate to produce a large total error. To avoid this problem, the views can be registered simultaneously and the error between the first and last views can be taken into proper consideration. Of course, this makes more sense from a theoretical point of view as well.

A multiview registration experiment was performed for the owl figurine using a total of six range views. Data acquisition was performed with the precision turntable. The object was rotated by 60 degrees between views. Fig. 2 illustrates the problem arising when the range views are registered two at a time.

Fig. 3 shows the result obtained when global optimization is performed to register the six range views. The results obtained with global registration are clearly superior to pairwise local registration. We can now see that the first and last views are properly registered.

IV. Conclusions

This paper presents an approach for the registration of range images. The method relies on formulating the registration task as an optimization problem by defining a cost function which measures the quality of registration between two range views. To do this for a specific rigid 3D transformation, the cost function evaluates the sum of Euclidean distances between control points in one view after transformation and finding corresponding points in the other view. Point correspondence between range views is rapidly established by computationally inverting the set of rangefinder calibration equations before the experiments are initiated. The claim of novelty for this approach is based on this latter aspect. The results achieved clearly indicate the accuracy of the method.

Acknowledgments

We thank Dr. Lester Ingber for kindly providing the VFSR (ASA) computer code as well as participating in valuable discussions. We would also like to thank Marc Bolduc, Samir Shah, Pierre Tremblay, Gilbert Soucy and Kenong Wu for their advice, expertise and technical assistance. G.B. would like to thank NSERC for its financial support in the form of an NSERC Postgraduate Scholarship.

M.D.L. would like to thank CIAR and PRECARN for their financial support. This research was partially supported by the Natural Sciences and Engineering Research Council of Canada by a Strategic Research Grant.

References

[1] P.J. Besl and N.D. McKay, "A method for registration of 3D shapes," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 14, no. 2, pp. 239-256, 1992.

[2] G. Blais, "Creating 3D computer objects by integrating multiview range data, Master's thesis, McGill University, 1993

[3] G. Champleboux, S. Lavall, R. Szeliski, and L. Brunie, "From accurate range imaging sensor calibration to accurate model-based 3D object localization," *CVPR'92,* June 1992.

[4] H.H. Chen, "Pose determination from line-to-plane correspondences: Existence condition and closed-form solutions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 6, pp. 530-541, June 1991.

[5] J-.C. Cheng and H-.S. Don, "A graph matching approach to 3D point correspondences," *Int'l J. of Pattern Recognition and Artificial Intelligence,* vol. 5 no. 3, pp. 399-412, 1991.

[6] Y. Cheng and G. Medioni, "Object modeling by registration of multiple range images," *Proc. 1991 IEEE Int'l Conf. on Robotics and Automation,* 1991.

[7] P. Huber, *Robust Statistics.* New York, Wiley, 1981.

[8] L. Ingber, "Very fast simulated reannealing (VFSR)," *Mathematical and Computer Modeling,* vol. 12, no. 8, pp. 967-973, 1989.

[9] M. Potmesil, "Generating models of solid objects by matching 3D surface segements," *Eighth Int'l Joint Conf. on Artificial Intelligence,* vol. 2, 1983.

[10] B. Sabata and J.K. Aggarwal, "Estimation of motion from a pair of range images: A review," *Computer Vision, Graphics and Image Processing,* vol. 54, no. 3, pp. 309-324, Nov. 1991.

[11] M.A. Shah and R. Jain, "Detecting time-varying corners," *Computer Vision, Graphics and Image Processing,* vol. 28, pp. 345-355, 1984.

[12] G. Soucy, "View correspondence using curvature and motion consistency," Master's thesis, McGill University, 1992.

[13] R. Szeliski, "Estimating motion from sparse range data without correspondence, *IEEE Second Int'l Conf. on Computer Vision,* pp. 207-216, Dec. 1988.

[14] P. Whaite and F.P. Ferrie, "From uncertainty to visual exploration," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no 10, pp. 1,038-1,050, Oct. 1991.

[15] P. Whaite and F.P. Ferrie, "Uncertain views," *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition,* pp. 3-9, Champaign, Ill., June 15-18, 1992.

CASM: A VLSI Chip for Approximate String Matching

Raghu Sastry, N. Ranganathan, and Klinton Remedios

***Abstract*—The edit distance between two strings $a_1, \ldots, a_m$ and $b_1, \ldots, b_n$ is the minimum cost s of a sequence of editing operations (insertions, deletions and substitutions) that convert one string into the other. This paper describes the design and implementation of a linear systolic array chip for computing the edit distance between two strings over a given alphabet. An encoding scheme is proposed which reduces the number of bits required to represent a state in the computation. The architecture is a parallel realization of the standard dynamic programming algorithm proposed by Wagner and Fischer, and can perform approximate string matching for variable edit costs. More importantly, the architecture does not place any constraint on the lengths of the strings that can be compared. It makes use of simple basic cells and requires regular nearest-neighbor communication, which makes it suitable for VLSI implementation. A prototype of this array has been built at the University of South Florida.**

***Index Terms*—Edit distance computation, string-to-string correction problem, very large scale integration (VLSI) implementation, systolic algorithm, special purpose architecture, hardware algorithm.**

I. Introduction

In approximate string matching, also known as the string-to-string correction problem, a similarity measure called the edit distance needs to be computed between two strings. This distance is computed using three editing operations, substitution, deletion, and insertion. Each of these operations has a cost associated with it. The objective of approximate string matching is to determine the minimum cost required to transform one string into another using these three editing operations.

String comparison is an important task in many disciplines. It has applications in information retrieval, pattern recognition [4], [9], [12], error correction, molecular genetics [7], [17], and text search and edit systems [2], [18]. Recent advances in Very Large Scale Integration (VLSI) technology have made the development of special purpose architectures and hardware algorithms for complex, computationally intensive algorithms possible. The attributes of parallelism, concurrency, pipelining, modularity and regularity have become

Manuscript received May 3, 1994; revised Feb. 10, 1995.

R. Sastry is with HAL Computer Systems, Inc., 1315 Dell Ave., Campbell, CA 95008; e-mail: sastry@hal.com.

N. Ranganathan and K. Remedios are with the Center for Microelectronics Research, Dept. of Computer Science and Engineering, University of South Florida, Tampa, FL 33620.

IEEECS Log Number P95074.