A PIECEWISE AFFINE MODEL FOR IMAGE REGISTRATION IN NONRIGID MOTION ANALYSIS

Guna Seetharaman[1] Ged Gasperas[2] and Kannapppan Palaniappan[3]

[1]The Center for Advanced Computer Studies
University of Louisiana at Lafayette
Lafayette, LA 70504-4330
guna@cacs.usl.edu

[2]Digi Dyne Inc
1170 Dolres St
San Francisco, CA
gasperas@digi-dyne.com

[3]Dept of Computer Eng. & Computer Sci.
University of Missouri-Columbia
Columbia, MO 65211
palani@cecs.missouri.edu

ABSTRACT

A piecewise-affine image registration method is proposed to compute the displacement field in an image sequence of an aerodynamic object experiencing heavy winds in a wind tunnel. Our method is useful for tracking objects whose net 3-D motion is fully characterized by a non-rigid motion and a significant component that is common to all parts of the object. A set of control points have been introduced on the object surface to help divide each image into a set of triangles and enable tracking each triangular area. The computed velocity field is piecewise affine for each triangle and is continuous across the boundary between any two adjacent triangles. Of interest is the accuracy of the registration process, and its estimate based on local moments and shape distortion that can be computed from the original image and the motion compensated version of its registered pair. The method has been applied to several images and the experimental results are presented.

Key words: Image registration, motion analysis, optical-flow, and MPEG techniques.

1. INTRODUCTION

The purpose of image registration is to enable a systematic way of analyzing two or more images related by an underlying context. In general, the pair of images will be drawn from a video image sequence which is temporal in nature. The most widely used instance of such application is the MPEG based systems where an image is described in terms of a previous image and a suitably coded difference image. In contrast, there exist a number of applications based on stereoscopic imagery where the images depict a scene observed from two vantage points. Stereoscopy has been used in many robotic systems since the seventies.

Block based computation of 2-D displacements widely used in MPEG systems are primarily concerned with data compression and do not associate any interpretation to the computed displacements. To a large extent these systems have applied affine model due to its simplicity, computational efficiency and factors governing real-time implementation. Two recent works on MPEG systems use a set of triangles to compute [1, 2] and describe the difference between consecutive frames in a video image sequence.

Oceanographic image sequence analysis applies registration for tracking clouds and mesoscale features such as turbulence and depressions associated with hurricanes[3]. These systems are unable to use the block based techniques because of the non-rigid [4] nature and the vast variation in the granularities at which the features manifest in these images. A more recent article by Chou *et. al.*, [5] presents the framework and a mathematical analysis for piecewise computation of the displacement fields in an oceanographic image sequence, using a set of triangles.

The focus of this paper is a piecewise-affine image registration technique for 3D motion analysis of complex objects in a temporal image sequence. More specifically we are interested in computing the change in surface reflectance properties of an object being subjected to heavy winds. The object has been painted with a pressure sensitive paint whose reflectance property changes with the surface pressure. Also, the object has several parts that tend to vibrate as it moves and thus requires a 3D non-rigid motion model. The ratio of pixel intensity between the original image and the motion compensated image would facilitate the computation of surface pressure. Conventional optical flow techniques would be inapplicable. It is also required deliver the result in the original 3D coordinate system pertaining to the 3D object, by using stereo-imaging. Image registration is central in both extract-

0-7803-6297-7/00/$10.00 © 2000 IEEE

ing the instantaneous 3D state of the scene from the stereo pairs, and registering two consecutive frames in each video sequence.

2. THE REGISTRATION MODEL

Let a column vector $\boldsymbol{x} = (x_1, x_2)$ represent the position of an arbitrary point in the image plane. Also, let $\boldsymbol{X} = \{\boldsymbol{x}_i\}, i = 1, 2, \cdots, I$; and, $\boldsymbol{Y} = \{\boldsymbol{y}_j\}, j = 1, 2, \cdots, J$; be the set of points extracted from the two images to be registered. In general $I \neq J$ for complex 3D scenes involving objects that are rotated about an axis parallel to the image plane. Then the image registration, by definition assigns each $\boldsymbol{x}_i \in \boldsymbol{X}$, a corresponding member $\boldsymbol{y}_j$, subject to certain relations $c(\boldsymbol{x}_i) = \boldsymbol{y}_j \in \boldsymbol{Y}$. In the case of 3D motion analysis based on *point correspondence* [6], the set $\boldsymbol{X}$ is made of small number of points. However, in a typical image registration application, both $\boldsymbol{X}$ and $\boldsymbol{Y}$ are allowed to become as large as the entire population of pixels in the image-grid.

There are at least four fundamentally different approaches [6, 7, 8, 9] to extract the 3D motion parameters and partially recover 3D structure of 3D objects from two or three video images. These methods have independently concluded that the displacement field produced by a rigid planar patch could be characterized by eight parameters involving quadratic terms in x, y coordinates.

Thus, we model the components of $\boldsymbol{y}$ in terms of two second order polynomials, of the form:

$$
\begin{aligned}
\Phi: \quad & \boldsymbol{x}_i \longrightarrow \boldsymbol{y}_j \\
& y_{j1} = a_0 + a_1 x_{i1} + a_2 x_{i2} + a_3 x_{i1}^2 + a_4 x_{i1} x_{i2} + a_5 x_{i2}^2 \\
& y_{j2} = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i1}^2 + b_4 x_{i1} x_{i2} + b_5 x_{i2}^2
\end{aligned}
$$

where, the vectors $\boldsymbol{a} = (a_0, \cdots, a_5)$ and $\boldsymbol{b} = (b_0, \cdots, b_5)$ characterize the mapping uniquely. At least five points in the image are necessary to compute $\boldsymbol{a}$ and $\boldsymbol{b}$. However, more observation would be required to overcome the errors introduced due to the discretization of the pixel coordinates. An ideal registration should produce a minimum value of the error:

$$\epsilon^2 = \int_{x,y} [f_1(\boldsymbol{x}) - f_2(\Phi(\boldsymbol{x}; \boldsymbol{a}, \boldsymbol{b}))]^2 dx dy$$

where $f_1(\cdot)$ and $f_2(\cdot)$ are the first and second video images.

Even a simple multifaceted object such a cube produces a set of nonlinear displacement fields each corresponding to one visible face, as it moves in the field of view of an observer. The work described in [8] assumes that the image has been segmented already, and then proceeds to fit an optimal motion field for each segment. Whereas the comprehensive approach due to Adiv [7] attempts both segmentation and motion-field parameter estimation using a one pass two stage approach. In contrast, the point-correspondence technique [6] implicitly assumes a single rigid motion-field characterized by eight pure parameters, and does not attempt to partition the points. Another work due to Strickland [10] describes both the process of registration, and a k-nearest neighbors set to facilitate the computation of local nonlinearities.

2.1. Piecewise-Affine Model of Registration

Affine transform exhibits several geometric properties that are desirable to analyze the space using piecewise linear functions. It maps triangles into triangles, ellipses into ellipses, and most importantly an elementary convex area into a convex area. So we choose to model the image as set of piecewise-convex areas that is triangles each of which is undergoing a different (affine) displacement, which can be uniquely computed from the observed displacement of its vertices. A *barry centric* interpolation technique introduced in [5] produces a displacement field that is continuous across the interface between any two adjacent triangles. This model offers rich way of decomposing a nonlinear displacement field computation into a simple piecewise affine parameter estimation.

The approach described here extends the ideas introduced in [5]. The approach described here is to: partition the 2-D image plane into several triangular shaped partitions; characterize the local changes in the image shape/intensity by a suitable affine transform for each partition; and, compute the motion compensated image in each partition by a piecewise bilinear function in (x, y). The approach requires a moderate number of points called control points spread across the image. A simple geometric algorithm is used to decompose the image plane into a large number of non-overlapping triangles. Each of these triangles will form a patch for fitting piecewise linear functions to fit the overall displacement field. Two consecutive frames of a test sequence are shown in Figure 1 and Figure 2. Images were first segmented to extract the control points from each frame, and then divided into a number of non overlapping triangles, as shown in Figure 1.

The registration process described here involves four parts. The first step is to extract the control points in each of the images. The second step is to establish the correspondence between two sets of control points each representing an image. Third step is to compute a dense velocity map of the apparent two dimensional displacement in the observed images. And finally enable some mechanism to relate the pixels to specific three dimensional locations on the object surface. We have implemented the last step by using a second camera and employing standard stereo analysis. Our set up in essence records instantaneous stereo images, and thus utilizes both temporal and spatial image sequence analyses.

2.2. Our multistage registration procedure

PHASE 1: The extraction of control points. As stated earlier, the objects has been painted uniformly with a pressure sensitive paint; and, a number of pressure-opaque dark points called *control-points* are also painted on the surface. The position of these points on the 3D model of the object is usually known at the time they are introduced. A multi-resolution edge detector using $\nabla^2 G$ mask was applied to identify the control points. The resulting edge-map was post processed to identify only the circular features, and thus locate only the control points. Then a iterative thresholding operator was applied to estimate the extent of the control points, as well as accurately compute the centroid of each control point.

PHASE 2: The registration of control points. This step was accomplished in two stages. First, a global trend was computed using the scatter the estimated center $\boldsymbol{\mu}$ and the scatter matrix $\boldsymbol{\Sigma}$ of both $\boldsymbol{X}$ and $\boldsymbol{Y}$. The underlying assumption is that both sets can be considered a set of random samples; and, at least 60% of points in these two sets are rigidly related through a single affine transform. After computing the globally consistent affine transform, each $\boldsymbol{y}_j$ was transformed into a corresponding $\widehat{\boldsymbol{x}_j}$. Then, a nearest-neighbor technique was applied to identify the optimal pairs $(\boldsymbol{x}_i, \hat{\boldsymbol{x}}_j)$, where, $\boldsymbol{x}_i \in \boldsymbol{X}$ and $\hat{\boldsymbol{x}}_j \in \hat{\boldsymbol{X}}$.

PHASE 3: Partitioning of the images into triangles. First we identify only those points in $\boldsymbol{X}$, for which a corresponding $\boldsymbol{y}$ has been established. We construct a spatial graph made of these points as its vertices. A simple variation of Kruscal algorithm for producing minimal spanning trees has been used to identify non overlapping triangle. Then, a convex hull computing algorithm was applied to these points to identify the net area covered by these points. All pixels within the convex hull share the same property in that their local displacement can be uniquely computed based on the three distinct measurements, each coming from one vertex of the triangle in which they are present. A controlled dilation of the convex hull was used to associate outside pixels to the nearest triangles, hence its local motion parameters.

PHASE 4: Computation of motion compensated image. The displacement of each pixel $\boldsymbol{x}$ in a triangle is computed based on its vertices. The resulting value is used to compute the motion compensated image. Thus the registered version of the second image is now made available in the frame of the first image.

3. EXPERIMENTAL EVALUATION

The object has been monitored by two video cameras. Also, the 3D coordinate of each control point was computed through stereo imaging, as well as, physical geometric measurements. The object was placed in many desired ways in 3D space to produce various types of displacement fields. Our investigation focussed on estimating the disparity between the actual (nonlinear) displacement field and piecewise affine interpolated displacement fields. The mean pixel displacement error between the actual (nonlinear) displacement field and the piecewise affine interpolated field, was computed using, a moment based approach. In addition, the visual similarity between the original image, and the motion compensated version of the second image, was also factored into evaluating the overall performance.

The proposed approach is also useful for providing interpolated or extrapolated 2D and 3D motion vectors for input to deformable motion sequence analysis algorithms[4]. This would facilitate speeding up the search process in applications where the imagery is non-uniformly sampled in time such as in meteorological remote sensing and cloud tracking[3]. Under non-uniform temporal sampling the maximum amount of local deformation will vary over time and is usually difficult to handle automatically. A closed triangulation based on the object only without triangle coverage of the background can be constructed, if needed, by using robust motion segmentation algorithms[11]. We are currently extending this work to oceanographic images. We have come up with abstract tokens based on multi-thresholded-image centroids. Initial experiments with oceanographic images suggest that shape distortion must be taken into account, in evaluating the performance.

REFERENCES

[1] Wael Badawy and Magdy Bayoumi. Low Power VLSI Architecture for 2D-Mesh Video Object Motion Tracking. In *Proceedings of the IEEE Comp. Soc Workshop on VLSI 2000*, pages 67–72, April 2000.

[2] Ahmad Shams. *A Low Power VLSI System for Mesh based Elastic Motion Analysis.* PhD thesis, Center for Advanced Computer Studies. University of Lafayette at Louisiana, 2000.

[3] A. F. Hasler, K. Palaniappan, and *et. al.* High resolution windfields within the inner core and eye of mature tropical cyclone from GOES-1min images. *Bulletin of the American Meteorological Society*, 79(11):2483–2496, 1998.

[4] K. Palaniappan, M. Faisal, C. Kambhamettu, and A. F. Hasler. Implementation of an automatic semi-fluid motion analysis algorithm on a massively parallel computer. In *10 th International Symposium on Parallel Processing Honolulu Hawaii*, April 15-19 1996.

[5] E. C. Chou, S. S. Iyengar, G. Seetharaman, J. Holyer, and M. Lybanon. Velocity vectors for features of sequential oceanographic images. *IEEE Trans. on Geoscience and Remote Sensing*, 36(3):985–998, May 1998.

[6] Roger Y. Tsai and Thomas S. Huang. Estimating Three Dimensional Motion Parameters of a Rigid Planar Patch. *IEEE Trans. on Acoustics Speech and Signal Processing*, ASSP-29:1147–1152, 1981.

[7] Gilad Adiv. Determining 3-D Motion and Surface structure from optical flow generated by several objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-7(No. 6):pp. 384–401, June 1985.

[8] Tzay Y. Young and Guna Seetharaman. A Region Based Approach to tracking 3D Motion in an image sequence . *Advances in Computer Vision and Image Processing*, Vol. 3, 1986.

[9] Guna Seetharaman. Three dimensional perception of image sequences. In *Handbook of Computer Vision by Tzay Y. Young. Academic Press, FL.*, 1994.

[10] R. N. Strickland and Z. Mao. Computing correspondences in a sequence of non rigid images. *Pattern Recognition*, 25(9):901–912, 1992.

[11] C. Y. Huang, K. Palaniappan, and X. Zhuang. Optic flow segmentation and motion estimation using a robust genetic partitioning algorithm. *IEEE Trans. on PAMI*, 17(12):1177–1190, 1995.

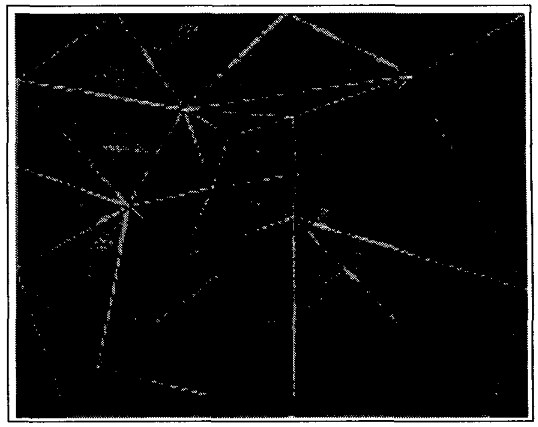

Figure 1. An image of a model B2 with 10 control points. The image plane is divided into a number of triangular regions.

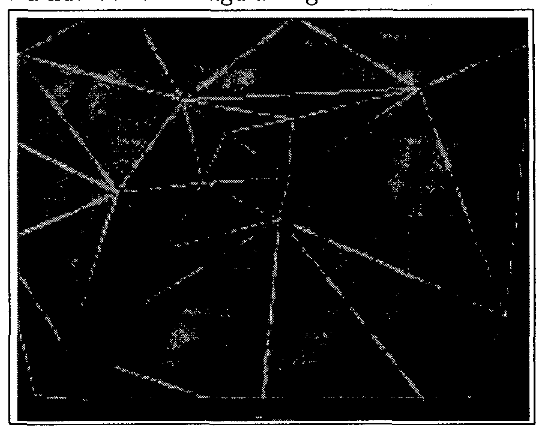

Figure 2. The object has moved and the overall intensity has also changed from the previous frame.

Figure 3. The motion compensated image resolved from the second frame (Fig. 2) into the first frame (Fig. 1).