Proc. VIIth Digital Image Computing: Techniques and Applications, Sun C., Talbot H., Ourselin S. and Adriaansen T. (Eds.), 10-12 Dec. 2003, Sydney

Significance-Based Multi-View Hausdorff Distance for Non-Rigid 3-D Object Registration

Yongsheng Gao

School of Microelectronic Engineering,
Griffith University, QLD 4111, Australia
yongsheng.gao@griffith.edu.au

Abstract. In this paper, we propose a new similarity measure to combine multiple views of 3-D objects for non-rigid registration. It is a metric that integrates multiple 2-D view features representing a visual identity of a 3-D object seen from different viewpoints. The robustness to non-rigid distortions is achieved by the proximity correspondence manner. The human face, a typical non-rigid object, was chosen to evaluate the capability of the proposed object matching technique. Very encouraging results were obtained which showed that the proposed Significant-Based Multi-View Hausdorff Distance (SMVHD) provides a new fusion method for non-rigid 3-D object registration.

1 Introduction

Non-rigid 3-D object matching presents a significant challenge to the computer vision and pattern recognition researchers due to geometric differences in representing the test and the model objects. To create a non-rigid object matching system, one strategy is to employ a deformable matching method [6,15,16]. Recently, attention has been focused on the techniques of fusing multiple images of the objects subject to nonlinear geometric deformations. Since the matching method has to be tolerant to distortions between the test object and the identical model, this tolerance might average out the differences that make individual objects unique. The difficulty lies in the design of the matching technique to tackle the problem of increased/large intra-class variation while keeping inter-class discrimination capability. In this paper, we present a nonlinear similarity measure to fuse multiple images from different viewpoints for non-rigid 3-D object registration.

The Hausdorff distance is a shape comparison metric to measure the dissimilarity between two point sets. Unlike most shape comparison methods that establish a point-to-point correspondence between the test point set and the model point set, it can be calculated without explicit pairing of points of the test set with the model set. The lack of point-to-point matching delivers speed and tolerance to local non-rigid distortions. The Hausdorff distance has been successfully applied to binary image comparison in computer vision [1,2,3]. Huttenlocher *et al.* [1] presented efficient algorithms to compute the Hausdorff distance between all possible relative positions of a binary image and a model for object location. The Hausdorff distance under translation and rotation was used to locate objects in a scene. It was shown that the method is quite tolerant to small position errors such as that occur with edge detectors and other feature extraction methods. Rucklidge [2] applied the Hausdorff distance to the task of locating an affine transformation of a model in an image. Efficient rasterized searching and location techniques were proposed to detect a planar object that has undergone weak-perspective projection. Takács [3] used Hausdorff distance to measure the similarity of two binary frontal face images. The above studies indicated that the Hausdorff distance could be used for non-rigid matching due to its robustness to local and non-rigid distortions. However, the Hausdorff distance is a metric for holistic similarity measuring from a single exemplar view to another view. Each point in the test set finds its match among all the points in the model set, i.e. all the points in the model set are candidate matches, or vice versa. It encounters problems/limitations for the applications as in the following cases.

1. In many object recognition applications, it is important to be able to identify objects by fusing multiple exemplar view images either due to the high accuracy requirement of the application or insufficient identity information from a single exemplar view. Each point in one view of the test object has to find its match among the points in the corresponding view of the model instead of among all the points in all views of the model.
2. In some non-rigid classification applications, such as the recognition of scissors and pliers [4], the objects are decomposed into rigid subparts and the feature points on one subpart of the input object has to be matched to the points on the same subpart of the model.
3. In the current definition of the Hausdorff distance, the contribution of each point is equal. However, this "assumption" is not always appropriate. Due to different segmentation techniques, the prominence of each point in representing the object may vary from point to point.

In this paper, we propose a novel Significance-Based Multi-View Hausdorff Distance (SMVHD) for the combined comparison of multiple views of two non-rigid objects. The contribution of each point depends on its significance to the identity representation of the object. The proposed method was evaluated on the matching of the human face, a typical non-rigid object. Encouraging results were obtained from all the three evaluating experiments under normal condition, non-rigid distortions generated by smiling expression and speaking action.

In the following, a brief introduction of the Hausdorff distance is presented in Section 2. In Section 3, the generic Significance-Based Multi-View Hausdorff Distance is proposed and described in detail. Encouraging experimental results are reported in Section 4. Finally, the paper is concluded in Section 5.

2 The Hausdorff Distance

The Hausdorff distance is a shape comparison metric based on binary images. It is a distance defined between two point sets. Given two finite point sets *M* = { m_1 , m_2 , ..., m_k } (representing a model) and *T* = { t_1 , t_2 , ..., t_n } (representing a test image), the Hausdorff distance is defined as

$$H(M,T) = \max(h(M,T), h(T,M)) \tag{1}$$

where

$$h(M,T) = \max_{m_i \in M} \min_{t_j \in T} \left\| m_i - t_j \right\| \tag{2}$$

and $\| m_i - t_j \|$ is Euclidean norm on the points of *M* and *T*. The function *h(M,T)* is called the directed Hausdorff distance from *M* to *T*. It identifies the point $m_i \in M$ that is the farthest from any point of *T* and measures the distance from m_i to its nearest neighbor in *T*. The Hausdorff distance *H(M,T)* is the maximum of *h(M,T)* and *h(T,M)*. Thus, it measures the degree of mismatch between two sets by measuring the distance of the point of *M* that is farthest from any point of *T* and vice versa. In this study, a new generic Significance-Based Multi-View Hausdorff Distance (SMVHD) is proposed for the multiple views fused similarity measuring of non-rigid objects.

3 Significance-Based Multi-View Hausdorff Distance

Let $M = \{M^1, M^2, \cdots, M^n\}$ be a model object consisting of *n* views from *n* different viewpoints and $T = \{T^1, T^2, \cdots, T^n\}$ be a test object consisting of *n* views from the same viewpoints as in the model. Define $M^1 = \{m_1^1, m_2^1, \cdots, m_{p_1}^1\}$, $M^2 = \{m_1^2, m_2^2, \cdots, m_{p_2}^2\}$,…, $M^n = \{m_1^n, m_2^n, \cdots, m_{p_n}^n\}$ to be *n* point sets representing the features in the *n* model views of object *M* and $T^1 = \{t_1^1, t_2^1, \cdots, t_{q_1}^1\}$, $T^2 = \{t_1^2, t_2^2, \cdots, t_{q_2}^2\}$,…, $T^n = \{t_1^n, t_2^n, \cdots, t_{q_n}^n\}$ to be *n* point sets representing the features in the *n* views of the test object *T* from the same viewpoints as in the model *M*. A point $t_i^k \in T^k$ in the kth view of the test object *T* has to be only allowed to find its match in the corresponding view M^k of the model *M*.

The Significance-Based Multi-View Hausdorff Distance (SMVHD) is defined as the distance between the model object (image set) *M* and the input object (image set) *T*:

$$H_{SMVHD}(M,T) = \max(h_{SMVHD}(M,T), h_{SMVHD}(T,M)) \tag{3}$$

The directed SMVHDs from *M* to *T* and from *T* to *M* are defined in (4) and (5).

$$h_{SMVHD}(M,T) = \frac{1}{\sum_{k=1}^{n}\sum_{m_i^k \in M^k} Sig_{m_i^k t_j^k}} \sum_{k=1}^{n}\sum_{m_i^k \in M^k} (Sig_{m_i^k t_j^k} \bullet \min_{t_j^k \in T^k} \left\| m_i^k - t_j^k \right\|) \qquad (4)$$

where $Sig_{m_i^k t_j^k}$ is the average significance of point m_i^k and its corresponding point t_j^k. The superscript *k* stands for the kth view of the model set *M* or the test set *T*, and *n* is the number of views in representing an object. Unlike the conventional Hausdorff distance that the contributions from all the points are equal, every distance of a matched pair in SMVHD is weighted by the average significance of m_i^k and t_j^k because its contribution to $h_{SMVHD}(M,T)$ is assumed to be proportional to the significances of the two matched points. This definition makes the SMVHD a more generic measure than the conventional Hausdorff distance to cover the problem three mentioned in the Introduction, i.e. the prominance of each point in representing an object could vary from point to point due to different segmentation techniques employed.

4 Experimental Results

The capability of the proposed SMVHD was evaluated on the human face, a typical non-rigid object. The face database [7] from the University of Stirling was used to evaluate the effectiveness of SMVHD on three views fused recognition under non-rigid distortions due to smiling and speaking.

The feature used in the experiments is dominant points of a face with merit. The merit is used as the significance of a point in SMVHD matching. Edges are the most fundamental features of objects in the 3D world. The edges in an image reflect large local intensity changes that are caused by the geometrical structure of the object, the characteristics of surface reflectance of the object, and viewing direction. In this study, the edges of the images were extracted using an edge detector based on the algorithm of Nevatia [8] followed by a thinning process to generate one pixel wide edge curves. Then, the dynamic two strip algorithm (Dyn2S) [9] was utilized to detect dominant points on the edge curves. In Dyn2S algorithm, a strip is fitted to the left and right of each point on the curve, and the points inside each strip are approximated as a straight line. The orientation and width of the strip are adjusted automatically. Longer and narrower strips are favored. A measure of “merit” based on the strips lengths and widths, and the angle between the strips is calculated. This merit provides an objective evaluation of the prominent strength of each point, which can be used as the significance of the point in the SMVHD calculation. The results of applying these processes on the frontal and profile faces are illustrated in Figure 1.

Since the matching test has only to be applied on facial area, i.e. the non-rigid object, pre-processing was conducted to crop the facial area and remove the influences of hair and clothes. An automatic profile location algorithm was employed to detect the nose tip and chin points. The distance between these two points was used as a reference to normalize image size, align face position and crop facial area. For the frontal faces, the eye locations were detected manually for normalization, alignment and cropping. Some automatic face/eye detection algorithms can be found in [10-14]. Figure 1 gives some examples of the cropped faces.

The database [7] from the University of Stirling contains 311 images of 35 people (18 females and 17 males). 31 people (16 females and 15 males) have complete image set, which contains 3 poses (frontal view, ¾ view and profile view) and 3 expressions (neutral expression, smiling and speaking), and thus can be used as our testing data sets.

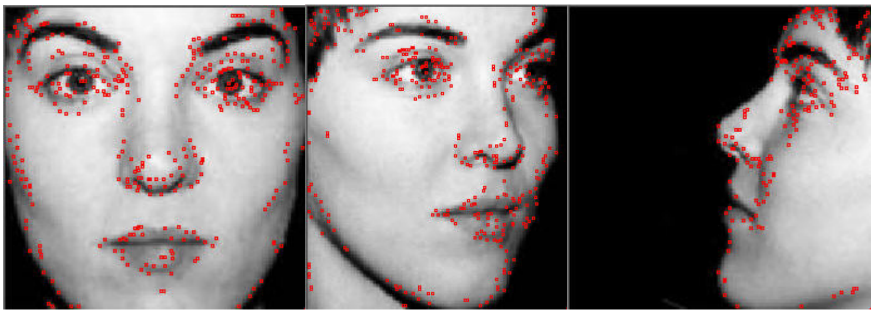

Figure 1 Examples of facial feature points (red dots) superimposed on the face images.

In the experiments, the frontal, ¾ and profile views of neutral expression from each person were used as the three views of one model of the person. The algorithm was tested using the three views taken under non-rigid deformations due to smiling expression and speaking action, respectively. Note that the experiment was a single model based object matching. Each model or input object is represented by three images from different viewpoints. There is significant non-rigid distortion between the input object and the corresponding model.

The recognition rates of the proposed SMVHD on non-rigid distortion due to smiling and speaking are summarized in Table 1 together with the accuracy of MHD. It is found that the MHD method on single view matching only achieved an average accuracy of 72.05% for non-rigid distortion due to smiling and 75.27% for non-rigid distortion due to speaking. The proposed SMVHD significantly improved the recognition rate up to 93.55%. This is a very encouraging recognition accuracy considering the fact that faces are very similar from person to person with small inter-class variations and the non-rigid distortions produce large intra-class variations. 29.84% and 24.29% of increases in accuracy for distortions due to smiling and speaking are significant and attractive. The experimental results reveal that the SMVHD effectively provides a new similarity measure for image fusion, which could be used for non-rigid object matching.

Table 1 The matching results on non-rigid distortion due to smiling and speaking.

Accuracy	MHD (on single view)			SMVHD (on 3 views)
	Frontal view	¾ view	Side view	
Smiling	70.97%	80.65%	64.52%	93.55%
Speaking	87.10%	80.65%	58.06%	93.55%

5 Conclusion

In this paper, we have proposed a new similarity measure to fuse multiple views of objects for non-rigid object matching. It is a metric that combines multiple 2-D view features representing a visual identity of a 3-D object seen from different viewpoints. The robustness to non-rigid distortions is achieved by the proximity correspondence manner. Experiments had been conducted on typical non-rigid objects, human faces, to evaluate the capability of the proposed SMVHD technique. The experimental results demonstrate that our technique consistently improved the system performance by fusing multiple views from different viewpoints. The improvements were significant when the object has undergone non-rigid deformations in which the single view MHD recognition had unsatisfactory accuracy. Human faces are very similar in structure from person to person and the deformations caused by expression and speaking significantly increased the intra-face variations. The experimental results obtained in this research are thus very encouraging due to the difficulty of face matching under non-rigid deformations. Actually, handling the variability in appearance due to varying expression/deformation is one of the key remaining problems in face recognition. As the proposed SMVHD is a generic similarity measure, we believe that it may also prove useful in other applications such as subpart matching as discussed in the introduction section. These are issues for future work.

References

1. D. P. Huttenlocher, G. A. Klanderman and W. J. Rucklidge, “Comparing Images Using the Hausdorff Distance”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, pp. 850-863, 1993.
2. W. J. Rucklidge, “Efficiently Locating Objects Using the Hausdorff Distance”, International Journal of Computer Vision, Vol. 24, pp. 251-270, 1997.
3. B. Takács, “Comparing Face Images using the Modified Hausdorff Distance”, Pattern Recognition, Vol. 31, pp. 1873-1881, 1998.
4. P. C. Yuen, “Recognition of non-rigid objects”, Proceedings of the 5th International Conference on Image Processing and its Applications. Edinburgh, UK, pp. 85-89, 1998.

5. M. P. Dubuisson and A. K. Jain, "A Modified Hausdorff Distance for Object Matching", Proceedings 12th International Conference on Pattern Recognition, Jerusalem, Israel, pp. 566-568, 1994.
6. Z. M. Kovacs-Vajna, "A fingerprint verification system based on triangular matching and dynamic time warping", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, pp.1266 –1276, 2000.
7. University of Stirling face database. http://pics.psych.stir.ac.uk/
8. R. Nevatia and K. R. Babu, "Linear Feature Extraction and Description", Computer Graphics and Image Processing, Vol. 13, pp. 257-269, 1980.
9. M. K. H. Leung and Y. H. Yang, "Dynamic two-strip algorithm in curve fitting", Pattern Recognition, Vol. 23, pp. 69-79, 1990.
10. E. Saber and A. M. Tekalp, "Frontal-view Face Detection and Facial Feature Extraction Using Color, Shape and Symmetry Based Cost Functions", Pattern Recognition Letters, Vol. 19, pp. 669-680, 1998.
11. K. M. Lam and H. Yan, "Locating and Extracting the Eye in Human Face Images", Pattern Recognition, Vol. 29, pp. 771-779, 1996.
12. H. A. Rowley, S. Baluja and T. Kanade, "Neural Network-Based Face Detection", IEEE Transactions on PAMI, Vol. 20, pp. 23-38, 1998.
13. C. H. Lee, J. S. Kim and K. H. Park, "Automatic Human Face Location in a Complex Background using Motion and Color Information", Pattern Recognition, Vol. 29, pp. 1877-1889, 1996.
14. K. K. Sung and T. Poggio, "Example-Based Learning for View-Based Human Face Detection", IEEE Transactions on PAMI, Vol. 20, pp. 39-51, 1998.
15. A. Goshtasby, "Registration of images with geometric distortions", IEEE Transactions on Geoscience and Remote Sensing, Vol. 26, pp. 60 –64, 1988.
16. A. Hill, C.J. Taylor, A.D. Brett, "A framework for automatic landmark identification using a new method of nonrigid correspondence", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, pp.241-251, 2000.