

Fusion of Infrared and Range Data: Multi-modal Face Images

Xin Chen, Patrick J. Flynn, and Kevin W. Bowyer

Dept. of Computer Science and Engineering, University of Notre Dame,
Notre Dame, IN 46556 USA
{xchen2, flynn, kwb}@nd.edu

Abstract. Infrared and range imagery are intriguing sensing modalities for face recognition systems. They may offer better performance than other modalities due to their robustness to environmental effects and deliberate attempts to obscure identity. Furthermore, a combination of these modalities may offer additional discrimination power. Toward this end, we present a semi-automatic system that captures range and infrared data of a human subject's face, registers and integrates multiple 3D views into one model, and applies the infrared measurements as a registered texture map.

1 Introduction

Although current face recognition systems employing intensity imagery have achieved very good results for faces that are taken in a controlled environment, they perform poorly in less uncontrolled situations. This motivates the use of non-intensity image modalities to supplement (or replace) intensity images [1]. Two major environmental problems in face recognition are illumination and pose variations [2]. Representations of the image and the stored model that are relatively insensitive to changes in illumination and viewpoint are therefore desirable. Examples of such representations include edge maps, image intensity derivatives and directional filter responses. It has been claimed [3] that no single one of these representations is sufficient by itself to withstand lighting, pose, and expression changes. Within-class variability introduced by changes in illumination is larger than the between-class variability in the data, which is why the influence of varying ambient illumination severely affects classification performance [4]. Thermal imagery of faces is nearly invariant to changes in ambient illumination [5], and may therefore yield lower within-class variability than intensity, while maintaining sufficient between-class variability to ensure uniqueness [1].

Well-known face recognition techniques, (for example, PCA), not only successfully applies to infrared images [6], they also perform better on infrared imagery than on visible imagery in most conditions [7] [8]. Calibrated 3D (range) images of the face are also minimally affected by photometric or scale variations. Therefore, they are receiving increasing attention in face recognition applications. Gordon [9] developed a curvature-based system employing Cyberware cylindrical scans. Beumier and Acheroy showed that recognition using surface matching from parallel profiles possesses high discrimination power, and also highlighted system sensitivity to absolute

gray level when range and intensity are considered jointly [10]. Yacoob and Davis [11] solved the related problem of face component labeling. Lapreste et al. [12] proposed a primal approach to face characterization from 3D images based on a structural analysis. Chua and Jarvis [13] proposed point-based features for free-form object recognition that could be used to match faces. Achermann et al. [14] also presented a system for face recognition using range images as input data, the results of their experiments show clearly that face recognition with range images is a challenging and promising alternative to techniques based on intensity.

Multimodal analyses seem to show promise in this domain. Recognition rates are improved by the combination of 3D and grey data, as reported by Beumier and Achery [10]. Wang et al. [15] proposes a face recognition algorithm based on both of the range and gray-level facial images. Chang et al. [16] designed a vector phase-only filter to implement a face recognition between range face (stored in the database) and intensity face (taken as the input), which is insensitive to illumination, but not scale and orientation invariant.

Since both infrared and range data are insensitive to variations caused by illumination, viewpoint, facial expressions and facial surface material changes, it is hoped that a combination of these two modalities may offer additional performance improvements for face recognition. Yet little multimodal experimental data of this sort exists. This paper presents a system that can semi-automatically produce a large dataset of integrated 3D model texture-mapped with IR data. As such, it offers a significant database building capability that can be used to good effect for large-scale face recognition trials from a limited database of experimental imagery.

2 Processing Method

The system described here takes as input multiple range and infrared images of the face, and produces a single 3D model with overlaid thermal sample values. The technical challenges in this task include interpolation of low-resolution IR values onto a high-resolution 3D mesh, registration of range views, and accommodation of some facial shape change between acquisitions. Our discussion focuses on two novel stages: mapping infrared data onto range data and view integration. The mapping stage assigns each range pixel an IR value and the integration stage combines two different view range images into one model.

2.1 Data Acquisition

Structured light acquisition systems use the projection of a known pattern of light (in our case, a laser stripe) to recover 3D coordinates [17]. Our acquisition proceeds as follows. A human subject is imaged in two poses corresponding to views offset 45 degrees (vertical rotation) on either side of frontal. Images are acquired roughly simultaneously from a Minolta Vivid 900 range camera and an Indigo Systems Merlin-Uncooled microbolometer array that senses long-wave infrared (LWIR) imagery. The cameras are placed side-by-side and standoff to the human subject is approximately two meters. Our LWIR camera is radiometrically calibrated but (other than

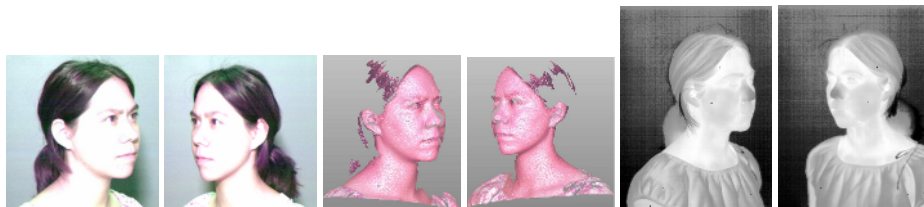


Fig. 1. Example images (color, range and infrared)

maintaining calibration during acquisition) we do not currently exploit the thermal calibration. After some trivial file manipulation, we have two 640x480 arrays of range and registered color intensity data, and two 320x240 arrays of infrared measurements.

2.2 Mapping Infrared Data onto Range Image

A. Spatial Transformation

A spatial transformation defines a geometric relationship between each point in the range/color and IR images. This is a 2-D image projection, since the cameras are assumed to be nearly coincident relative to the standoff. The goal is to obtain a mapping $(X(u, v), Y(u, v))$ between range image raster coordinates (u, v) and the corresponding position (x, y) in the IR image raster. $X(.,.)$ and $Y(.,.)$ are obtained through manual feature selection. Since the mapping will not in general take integer coordinates to integer coordinates, an interpolation stage is used to fill in destination raster values [20]. The form of the mappings $X(.,.)$ and $Y(.,.)$ is the affine transformation, with coefficients estimated from corresponding points (an assumption of affinity is appropriate given the standoff assumption above).

Rather than estimate a single affine coordinate map, we estimate maps independently within corresponding triangles identified in the image. The six coefficients a_{ij} are estimated from point triplet correspondences chosen manually. The more triangles into which the face is divided, the more precise the mapping will be. To infer the affine transformation, we need to provide at least three corresponding point pairs with the constraint on the point set in color image to consist of non-collinear points. When more than three correspondence points are available and when these points are known to contain errors, it is common practice to approximate the coefficients by solving an over-determined system of equations. However, it is unnecessary to use more than three point pairs to infer an affine transformation in our case since we can easily identify corresponding pairs with tolerable data errors.

Therefore, our method is to manually select approximately ten feature points and obtain a Delaunay triangulation of the convex hull of the point set in both images. The feature points chosen include anatomically robust locations such as the pupil, eye corner, brows, nose tip, and mouth. Normally, the features are more difficult to obtain in the IR image. Only coordinates within the convex hull of the points chosen can be mapped to the range image coordinate system. Figure 2 shows a typical triangular decomposition of IR and range (depicted with the registered color) images for one subject.

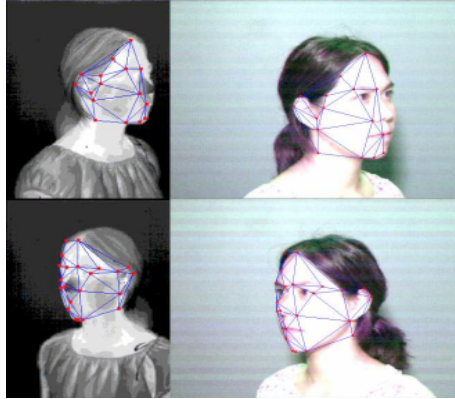


Fig. 2. Triangulation of color image of range data and grayscale image of IR data

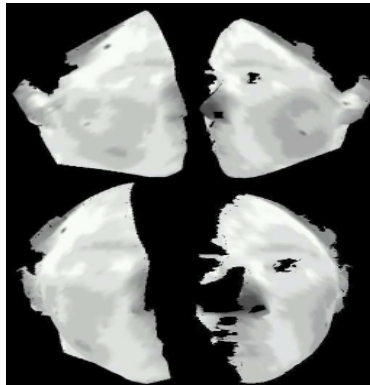


Fig. 3. Range face mapped with IR data

B. Temperature Interpolation

Once a mapping between range raster coordinates and corresponding IR pixels has been established, the IR measurements are used to “fill in” the range image mesh, causing the range image to gain another measurement (IR) at each pixel location. This requires an interpolation step. Figure 3 is a mapped result of the left-side face pose, rotated to see different views.

C. Initial Registration

We can estimate a rotation and translation that aligns the two objects roughly. Three non-collinear points are enough to compute the transformation since we can fix the 6 degrees of freedom in 3D space.

We manually select 3 points in the left and right pose range image respectively. The selection is not perfectly precise and need not be. We always select easily identified human face feature points to reduce the data error (eye corners, chin points, nose tip). Experience has shown that some guidelines should be followed when selecting points. Tangent edge points (jump edges) should not be picked since their positions are not reliably estimated. The triplet of points should not be nearly collinear because the transformation estimate may be ill-conditioned.

Before registration, we arbitrarily set the left-turn pose face surface fixed in 3D coordinate system; the right-turn pose face surface is moved to be in best alignment. We call former surface model shape, the latter data shape. As a result, let p_i be a selected point set on a data shape P to be aligned with a selected point x_i from a model point set X.

D. Modified ICP Registration

With the corresponding point set selected, we implement the quaternion-based algorithm for registration. It makes data shape P move to be in best alignment with model shape X.

Let $q_R = [q_0 \ q_1 \ q_2 \ q_3]^T$ be a unit quaternion, where $q_0 \geq 0$ and $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$. The corresponding 3x3 rotation matrix is given by

$$R(q) = \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 - q_0q_2) \\ 2(q_1q_2 + q_0q_3) & q_0^2 + q_2^2 - q_1^2 - q_3^2 & 2(q_2q_3 - q_0q_1) \\ 2(q_1q_3 - q_0q_2) & 2(q_2q_3 + q_0q_1) & q_0^2 + q_3^2 - q_1^2 - q_2^2 \end{pmatrix}$$

The translation component of the registration transform is denoted $q_T = [q_4 \ q_5 \ q_6]^T$. The complete registration state vector q is denoted $[q_R \ q_T]^T$. The mean square registration error (to be minimized) is

$$f(q) = \frac{1}{N_p} \sum_{i=1}^{N_p} \|x_i - R(q_r)p_i - q_t\|^2$$

Our goal is to minimize $f(q)$ subject to the constraint that the number of corresponding points is as large as possible. Besl and McKay[19] proposed an automatic surface registration algorithm called ICP which registers two surfaces starting from an initial coarse transformation estimate. This algorithm has been shown to converge fast but not necessarily towards the global optimal solution.

ICP is not useful if only a subset of the data point shape P corresponds to the model shape X or a subset of the model shape X, In our situation, assuming that one view is to be treated as the model shape and one as the data shape, these two shapes have only a narrow strip of overlapping area. ICP requires modification in our application. Another restriction of ICP is that the two surfaces are from rigid objects. However, the human face deforms non-rigidly continuously in our situation due to respiration and the body's unconscious balance control (subjects are standing when imaged). Again, this shows that ICP cannot be directly applied in our application.

“Closest points” that are “too far apart” are not considered to be corresponding points and marked as invalid so they have no influence during the error minimization. This is accomplished through an “outlier detection” phase. We define a threshold dynamically. In each ICP step, we “trust” the previous step’s result and make use of the mean square distance calculated from that step as the threshold for the current step. This method can prevent introduction of unlikely corresponding point pairs while giving a good registration quickly.

In ICP, a good starting configuration for the two surfaces P and X is essential to a successful convergence. However, the range of successful starting configurations is quite large which does not impose difficult constraints to the operator when entering a pose estimate for P and X . Fortunately, it is fairly easy to manually select three corresponding points in each view to obtain a tolerable data error. The initial registration not only gives a useful approximation of registration but also provides an approximate average distance between the corresponding point pairs in the two images. Specifically, we can use the mean square distance calculated from the mean square objective minimization as our first threshold for modified ICP algorithm.

The modified ICP algorithm is defined as follows:

- Input: Two face surfaces P and X containing respectively N_P and N_X vertices, an initial transformation $q_0 = (R_0, t_0)$ which registers P and x approximately; and a mean square distance computed in initial registration using the default threshold T .
- Output: A transformation $q=(R, t)$ which registers P and X .
- Initial Configuration: Apply the transformation (R_0, t_0) to P .
- Iteration: Build the set of closest point pairs (p, x) . If their distance exceeds T , discard the pair. Find the rigid transformation (R, t) that minimizes the mean square objective function. Update R and t . Set $T = f(q)$. Repeat until convergence of $f(q)$.

We use a kd-tree data structure to facilitate nearest-neighbor searches in the ICP update step. In order to verify the registration quality and terminate the iterative algorithm, the mean-square distance is generally used. We can also use the number of corresponding point pairs as a sign to stop the iterations. Figure 4 (left) shows the result of the registered range images of a human face scanned in two different poses. Figure 4 (right) shows the registration result of the two face surfaces mapped with IR data.

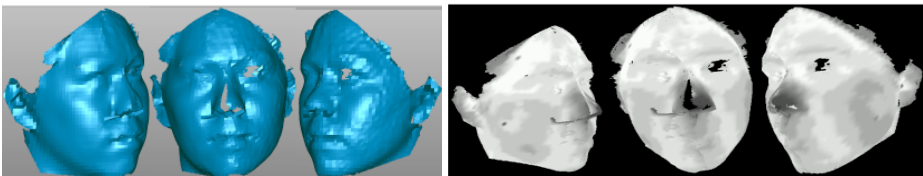


Fig. 4. Registered face surfaces (left) and registered face surfaces mapped with IR data (right)

E. Mesh Fusion

There are several methods to integrate registered surfaces acquired from different views. We propose a new mesh fusion algorithm that is particularly simple and useful in our human face applications. It erodes the overlapping surface of the data face shape until the overlap disappears, then constructs a frontier mesh region to connect them.

Due to the complexity of the human face surface, we expect that there will be multiple disjoint regions of overlap between the data and model meshes. Schutz et al. [18] proposed a mesh fusion algorithm which can deal with such problems. Our approach is simpler and relies on the distinctive nature of the registered meshes arising from our sensing set-up. We preserve the model mesh as a continuous region without holes while maximizing the face area it can cover by carefully selecting feature points which construct the convex hull (mentioned in Section 2.2A). The model mesh remains intact while the data mesh is eroded in the overlapping region. Vertices in the data mesh that are within a threshold distance of a vertex in the model mesh are removed; this process continues until no vertices are removed. The threshold value is determined empirically, and in our case a 5 to 10mm value works well. The result, as depicted in Figure 5, is a pair of faces with a curvilinear frontier between them.

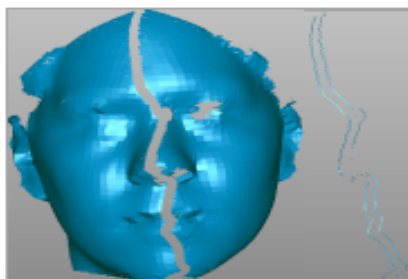


Fig. 5. Gap and frontiers

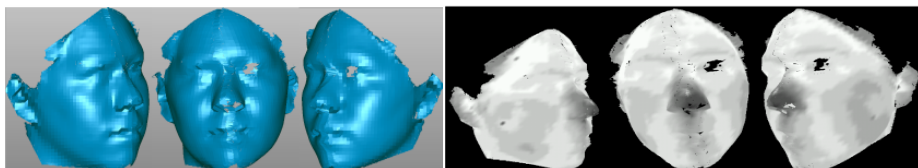


Fig. 6. Mesh integration results

The frontier is a distinguished set of vertices. Any point inside the convex hull of either mesh whose left/right adjacent pixel is eroded is labeled as a frontier point. Holes in the image due to missing data are not considered. These vertices are placed

in a linked list. The gap enclosed by the two frontiers is filled with triangles. The frontier list of the two face surfaces is sorted in incremental y coordinate order. Figure 6 illustrates the mesh fusion result as shaded 3D mesh data and seamlessly integrated IR overlays.

3 Summary and Conclusions

The system described in this paper has been used to process several sets of multimodal imagery of experimental subjects acquired in a data collection protocol. Inspection of these results suggests that IR detail and overall 3D shape of the face are well preserved, and that the range image integration step works reasonably well. However, spurious range points are not always eliminated by the filtering procedure, missing data due to the lack of range points on the eyeball yields a model with holes, and radiometrically calibrated IR data is currently not incorporated into the model. This is the focus of current research. Results to date suggest that this approach to creation of synthetic head models with IR attributes, which can then be rendered to produce IR images from any viewpoint, offers a potentially valuable source of data to multimodal face recognition systems.

References

1. Jain, A., Bolle, R. and Pankanti, S., *Biometrics: Personal Identification in Networked Society*, Kluwer Academic Publishers, 1999.
2. Zhao, W., Chellappa, R., Rosenfeld, A. and Phillips, J. "Face Recognition: A Literature Survey", Univ. of MD Tech. Rep. CFAR-TR00-948, 2000.
3. Adini, Y., Moses, Y. and Ullman, S. "Face Recognition: The Problem of Compensating for Changes in Illumination Direction", Proc. ECCV, A:286-296, 1994.
4. Wilder, J., Phillips, P.J., Jiang, C. and Wiener, S. "Comparison of Visible and Infrared Imagery for face Recognition", Proc. Int. Conf. Autom. Face and Gesture Recog., 192-187, 1996.
5. Wolff, L., Socolinsky, D. and Eveland, C. "Quantitative Measurement of Illumination Invariance for Face Recognition Using Thermal Infrared Imagery", Proc. Workshop Computer Vision Beyond the Visible Spectrum, Kauai, December 2001.
6. Cutler, R. "Face Recognition Using Infrared Images and Eigenfaces", website <http://cs.umd.edu/rgc/face/face.htm>, 1996.
7. Socolinsky, S. and Selinger, A., "A Comparative Analysis of face Recognition Performance with Visible and Thermal Infrared Imagery", Tech Rep., Equinox Corp., 2001.
8. Selinger, A. and Socolinsky, D. "Appearance-Based Facial Recognition Using Visible and Thermal Imagery: A Comparative Study", Proc. Int. Conf. Pattern Recognition, Quebec City, 2002.
9. Gordon, G. "Face Recognition based on Depth Maps and Surface Curvature", Proc. SPIE 1570, 234-247, 1991.
10. Beumier, C. and Acheroy, M., "Automatic Face Verification from 3D and Grey Level Clues", Proc. 11th Portuguese Conference on Pattern Recognition (RECPAD 2000), Sept. 2000.

11. Yacoob, Y. and Davis, L. "Labeling of Human Face Components from Range Data," CVGIP 60(2):168-178, Sept. 1994.
12. Lapreste, J., Cartoux, J. and Richetin, M. "Face Recognition from Range Data by Structural Analysis", NATO ASI Series v. F45 (Syntactic and Structural Pattern Recognition), Springer, 1988.
13. Chua, C. and Jarvis, R. "Point Signatures: A New Representation for 3D Object Recognition", Int. J. Comp. Vision 25(1):63-85, 1997.
14. Achermann, B. and Jiang, X. and Bunke, H., "Face Recognition using Range Images", Proc. International Conference on Virtual Systems and MultiMedia '97 (Geneva, Switzerland), Sept. 1997, pp. 129-136.
15. Wang, Y., Chua, C. and Ho, Y. "Facial Feature Detection and Face Recognition from 2D and 3D Images", Pattern Recognition Letters 23(10):1191-1202, August 1991.
16. Chang, S., Rioux, M. and Domey, J. "Recognition with Range Images and Intensity Images", Optical Engineering 36(4):1106-1112, April 1997.
17. Beumier, C. and Acheroy, M., "Automatic Face Authentication from 3D Surface", Proc. 1998 British Machine Vision Conference, Sept. 1998, pp. 449-458.
18. Schutz, C., Jost, T. and Hugli, H. "Semi-Automatic 3D Object Digitizing System Using Range Images", Proc. ACCV, 1998.
19. Besl, P.J. and McKay, N.D., "A Method for Registration of 3-D Shapes", IEEE Trans. on PAMI 14(2):239-256, February 1992.
20. Wolberg, G., Digital Image Warping, Wiley-IEEE Press, 1990.