A Novel Pose Invariant Face Recognition Approach Using A 2D-3D Searching Strategy

Nicholas Dahm and Yongsheng Gao
School of Engineering, Griffith University
Brisbane, QLD, Australia
{n.dahm, yongsheng.gao}@griffith.edu.au
Queensland Research Laboratory, National ICT Australia
Brisbane, QLD, Australia
{nicholas.dahm, yongsheng.gao}@nicta.com.au

Abstract—Many Face Recognition techniques focus on 2D-2D comparison or 3D-3D comparison, however few techniques explore the idea of cross-dimensional comparison. This paper presents a novel face recognition approach that implements cross-dimensional comparison to solve the issue of pose invariance. Our approach implements a Gabor representation during comparison to allow for variations in texture, illumination, expression and pose. Kernel scaling is used to reduce comparison time during the branching search, which determines the facial pose of input images. The conducted experiments prove the viability of this approach, with our larger kernel experiments returning 91.6% - 100% accuracy on a database comprised of both local data, and data from the USF HumanID 3D database.

Keywords-Face Recognition; Pose Invariant; 2D-3D

I. INTRODUCTION

Face recognition is an increasingly popular biometric identification technique, currently being adopted for security and law enforcement. It excels in these applications, as its passive and non-intrusive nature allows constant monitoring, while not interrupting personnel. Despite the fact that face recognition has matured into an extensive research field, there are several major challenges prevailing, including handling changes in lighting, expression, facial hair and facial pose. Facial pose is a particularly difficult challenge, and is still an open research problem [1].

Recent technological advances have made it increasingly easy to capture 3D textured data of the entire head. Using 3D textured data, algorithms can overcome problems with facial pose. While this may increase accuracy, it generally requires that a subject sits still while an imaging device circles them. Systems that require full scans for recognition can no longer be classified as non-intrusive, as subjects must interrupt their normal duties to use them.

In this paper, we present a novel face recognition approach that utilises 3D data to overcome changes in facial pose, while remaining non-intrusive. To achieve this, we use 3D textured head models in the gallery, as gallery data is generally taken from cooperative subjects (e.g. identification photos, mugshots). For the probes, we use 2D images, which can be taken from passive cameras such as ceiling-mounted surveillance cameras. Together this gives us a cross-dimensional approach, combined with a non-intrusive nature.

II. RELATED WORK

A considerable amount of work on 3D face recognition looks at using 3D shape without 2D texture [2]. While this prevents textural inconsistencies from altering the recognition accuracy, it discards what has been proven (by 2D face recognition) to be useful identification information.

2D face recognition techniques commonly make use of known structural information to fit Active Shape Models (ASMs) [3] or similar variants [4][5]. Wiskott et al. [5] characterise a face by Gabor jets at a set of keypoints previously identified by their Elastic Bunch Graph Matching (EBGM) algorithm, which uses an ASM-like mesh. These Gabor jets are insensitive to lighting and texture variations, as well as small rotations, making their framework considerably robust.

The introduction of laser head and face scanners has eased the capturing of detailed facial structure. Blanz and Vetter [6] used this technology to create a generic 3D head shape, which could be morphed to fit a subject from a 2D facial image. This gives the ability to obtain images from arbitrary angles around the head, however any parts occluded in the 2D image can only be estimated or guessed. This restricts the allowable pose variations.

III. CROSS DIMENSIONAL COMPARISON

Our proposed approach is cross-dimensional, comparing 2D probe images against a gallery of textured 3D head models. In order to create meaningful recognition scores, we must determine a way to compare 2D against 3D. Our approach creates 2D rendered views of the 3D model from different angles, which are then compared against the 2D probe. Each rendered view is created by deforming the 3D model’s texture with the 3D shape information, as shown in Figure 1.
We use the term render here, as the lighting on the rendered view is not possible in the real world. The reason it is not possible is due to the lighting configuration on the laser scanner. The scanner that is used to create the 3D models was a Cyberware® 3030PS Head and Face Laser Scanner [7]. The scanhead circles around the subject, capturing vertical scanlines of the subject’s head and face. To generate adequate lighting, the scanhead contains its own light sources, which circle the subject with it. This means that each vertical line of the subject’s face has its own unique light source, generally perpendicular to the skin.

In addition to this unrealistic lighting on the render, there are two other issues with laser scanners. Firstly, the skin surface appears more rough and grainy. Secondly, the skin texture appears to have a different shade altogether, not simply in terms of brightness and contrast.

A. Gabor Representation

In order to overcome the appearance issues detailed in section III, a representation that is insensitive to changes in lighting and texture must be used. Building on existing techniques, we represent images in the Gabor domain, similar to the work on EBGM [5]. Representation in the Gabor domain also makes the comparison less sensitive to viewing angle changes, which will aid the searching in section III-C.

We use 40 Gabor kernels [8][5], where each kernel is a wavelet of one mother wave, sampled over 5 frequencies \( \nu = 0 \ldots 4 \) and 8 orientations \( \mu = 0 \ldots 7 \). The parameter vector \( \tilde{k}_j \) of each wavelet is formed using equation 1.

\[
\tilde{k}_j = \left( \frac{k_{jx}}{k_{yy}} \right) = \left( \frac{k_{x} \cos \varphi_{\mu}}{k_{x} \sin \varphi_{\mu}} \right), k_{x} = 2^{-\frac{j+2}{2}}, \varphi_{\mu} = \mu \frac{\pi}{8}
\]

(1)

The parameter vector can then be used to create a wavelet. Each wavelet \( \psi_0 \) to \( \psi_99 \) can be thought of as an image with complex valued pixels, or two images representing the real and imaginary components. Each pixel in the wavelet can be determined using equation 2. In this equation, \( x^2 \) represents \( \|\tilde{x}\|^2 \) but is shown in condensed format for readability.

\[
\psi_j(\tilde{x}) = \frac{k_{x}^2}{\sigma^2} \exp\left( -\frac{k_{x}^2 x^2}{2\sigma^2} \right) \left[ \exp(i\tilde{k}_j \tilde{x}) - \exp\left( -\frac{\sigma^2}{2} \right) \right]
\]

(2)

Once the set of wavelets is constructed, they can be convoluted with an image \( I \). Each wavelet is centered across each pixel in the image, then the convolution result is recorded for that pixel. When convoluted with all wavelets, each pixel in the image returns a single complex number for each wavelet. This array of complex numbers forms a Gabor jet \( J \). Each jet \( J \) is a set of complex numbers \( J = \{ J_j : j \in \{0,1, \ldots, 39\} \} \) where \( J_j \) represents the coefficients of an individual wavelet. Equation 3 calculates the coefficient for a single wavelet on a single image pixel where \( \tilde{x} = (x, y) \).

\[
J_j(\tilde{x}) = \int I(\tilde{x}') \psi_j(\tilde{x} - \tilde{x}') d^2 \tilde{x}'
\]

(3)

The convolution operation has a computational complexity of \( O(n^2) \), so instead we use Fast Fourier Transforms and multiplication in the frequency domain to achieve a complexity of \( O(n \log n) \).

B. Comparison

The Gabor domain representation shown in section III-A transforms an image into a vector of \( n \) pixels, where each pixel is a jet of 40 complex values. Each complex number in a jet contains real and imaginary components, which may be combined or used independently. If used independently, the image is represented by a vector of \( m = 80n \) values. A better way is to combine the real and imaginary parts to form a magnitude value [8][5]. Magnitude is calculated by: \( \text{magnitude} = \sqrt{x_{\text{real}}^2 + x_{\text{imag}}^2} \). This gives a vector of \( m = 40n \) values to compare.

The similarity function shown in equation 4 has been shown to achieve better results than others, such as Euclidean distance.

\[
\psi_j(\tilde{x}) = \frac{k_{x}^2}{\sigma^2} \exp\left( -\frac{k_{x}^2 x^2}{2\sigma^2} \right) \left[ \exp(i\tilde{k}_j \tilde{x}) - \exp\left( -\frac{\sigma^2}{2} \right) \right]
\]

(4)
\[
\text{Similarity}(P, R) = \frac{P \cdot R}{\|P\|\|R\|} = \sqrt{\sum_{i=0}^{m} P_i R_i^2 \sum_{i=0}^{m} P_i^2 \sum_{i=0}^{m} R_i^2}
\]

where \(P\) and \(R\) represent the Probe image and Rendered image respectively, each containing \(m = 40n\) magnitude values.

One clear advantage of using magnitude with this similarity function is that it gives us phase invariance [5]. This prevents small alignment errors from creating large inconsistencies in accuracy.

C. Searching Strategy

The steps so far outline the process of obtaining a single rendered view and comparing it against a probe. Our system makes no assumption about the pose in the probe, and instead searches for the ideal match on the 3D model. It does this by iteratively creating rendered views from different pitch and yaw angles, then comparing them to the probe.

We can visualise our search space as a rectangular point grid, with pitch and yaw axes taken at intervals of a given precision (usually \(5^\circ\)). Since Gabor is not too spatially sensitive, the search space is smooth, allowing for efficient searching. Unfortunately, a single similarity value does not give any clues as to an intelligent search direction, so we instead use a branching search algorithm.

Each iteration of our searching algorithm subdivides the current search space into a number of possible regions. The points in the center of each region are tested, and the regions with the best similarity survive to the next round, while all others are discarded. This process continues until the searching algorithm finds a single point on the grid, which represents the most likely pitch and yaw angles for the probe. Since the searching process uses the same mechanism as the comparison, there is no need for any final comparison to be done. The search can be sped up by using a smaller initial kernel, then increasing it as the search progresses. As the kernel size increases, it trades speed for accuracy, as we have done in section IV-B.

IV. EXPERIMENTAL RESULTS

In order to test the viability of this design, we developed a complete prototype system. The system takes in a 2D photo and a 3D model, searches for the best matching angle, and gives a similarity measurement.

A. Dataset

The dataset required for our experiments consists of three parts. The first is 2D photos of the probe subjects taken with a standard camera. Secondly, we require 3D scanned models of the probe subjects to allow us positive matches. Finally, a number of other 3D scanned models are needed to ensure that the database is sufficiently large. These additional models serve as negative matches.

The Griffith University Computer Vision and Image Processing Laboratory (Griffith CVIP Lab) provided 2D photos and 3D scanned models of the probe subjects. This contained data of 4 subjects, each with 3 probe images and a 3D head model. An extra 22 random 3D models were taken from the USF-HumanID Database [9], each of a different subject. Unfortunately since the USF-HumanID Database did not have 2D photos of the same people in the 3D head models, we could not use these for positive matches. Overall, this gave us 12 probe photos of 4 different subjects to compare against a 3D model database of 26 different subjects. Each of these probe photos was within \(\pm 80^\circ\) yaw and \(\pm 10^\circ\) pitch from the directly frontal pose. There was a marginal roll in some probes, however more than \(\pm 5^\circ\) is unlikely.

B. Experimental Setup

There were two main configuration options for our prototype. The first was the angle interval, which determined the final precision for both yaw and pitch. We conducted primary experiments on both \(5^\circ\) and \(1^\circ\) precisions.

The second option was the Gabor profile, of which we had 2 possibilities. Profile 1 used a 32x32 kernel size, along with a 32x32 image size for comparison. Profile 2 used a 64x64 kernel size, with a 64x64 image size. Both profiles used a 16x16 kernel size and 16x16 image size during searching, then switched to their respective larger sizes for the final comparison.

C. Results

Our primary experiments were conducted on both profiles, using both \(5^\circ\) and \(1^\circ\) angle precisions. The \(5^\circ\) precision experiments performed exceptionally well. Profile 1 obtained 91.6\% 1st rank matches and 8.3\% 4th rank matches, while Profile 2 obtained a perfect 100\% 1st rank matches.

The \(1^\circ\) precision experiments performed slightly worse, however no matches were lower than 3rd rank. This time, Profile 1 obtained only 75\% 1st rank matches, with 25\% 2nd rank matches, while Profile 2 dropped to 91.6\% 1st rank matches and 8.3\% 3rd rank matches.

In extended experiments, we tested our system against the three different probes shown in figure 2, each containing some occlusion. Probes 1 and 3 were taken with standard reading glasses on. Probe 2 was taken with sunglasses on, which created visible reflections, including one large specular reflection in the right eye.

Despite the occlusion in these images, the system was able to find high rank matches. These experiments were conducted at \(5^\circ\) precision against the full 26 head models. Profile 1 returned 2nd, 5th and 1st rank matches respectively while Profile 2 returned 2nd, 2nd and 1st rank matches respectively.
V. Conclusion

In this paper we have proposed a novel approach to face recognition between 2D and 3D data. Our approach is based on the current trend of identification in security, where gallery data is taken while subjects are cooperative, but probe data is unconstrained. By combining the pose invariance given by a 3D model, with the insensitive nature of Gabor jets, we build a robust framework for conducting face recognition across varying pose, illumination and texture.

The results from our experiments show that our approach is viable and can be quite effective, with our larger kernel experiments returning 100% and 91.6% respectively. Results from the occlusion experiment were particularly interesting, with our design giving high rank matches, despite such a large specular reflection. While these results show the viability of our approach, a larger test set is required to thoroughly gauge its true effectiveness.

VI. Future Work

In future experiments, we plan to adapt this approach to use 2.5D (range) images. This will allow us to test against larger, publicly available datasets. Since frontal range images have little information for the side of the head, we anticipate that the accuracy will suffer significantly at angles of greater than 60°.

REFERENCES

[9] “USF HumanID 3D Face Database, Courtesy of Sudeep Sarkar, University of South Florida, Tampa, FL.”