# Dynamic Approach for Face Recognition using Digital Image Skin Correlation

Satprem Pamudurthy[1], E Guan[2], Klaus Mueller[1], and Miriam Rafailovich[2]
[1]Department of Computer Science, [2]Department of Material Science and Engineering
State University of New York at Stony Brook

Abstract.  With the recent emphasis on homeland security, there is an increased interest in accurate and non-invasive techniques for face recognition. Most of the current techniques perform a structural analysis of facial features from still images. Recently, video-based techniques have also been developed but they suffer from low image-quality. In this paper, we propose a new method for face recognition, called Digital Image Skin Correlation (DISC), which is based on dynamic instead of static facial features. DISC tracks the motion of skin pores on the face during a facial expression and obtains a vector field that characterizes the deformation of the face. Since it is almost impossible to imitate another person's facial expressions these deformation fields are bound to be unique to an individual. To test the performance of our method in face recognition scenarios, we have conducted experiments where we presented individuals wearing heavy make-up as disguise to our DISC matching framework. The results show superior face recognition performance when compared to the popular PCA+ LDA method, which is based on still images.

## 1.  Introduction

In this paper, we propose a face recognition method that is based on a feature tracking method we call *Digital Image Skin Correlation* (DISC). Unlike other feature tracking methods that require a set of sparse markers to be attached to the actor's face or have a sparse set of feature points, DISC uses the natural texture of the skin for tracking the facial motions. The distribution of skin pores on the face provides a highly dense texture, which we exploit for capturing fine-scale facial movements. By tracking an individual's skin pores during a facial expression, such as a subtle smile, we can generate a vector field that describes the individual's facial movement. We refer to these vector fields as facial deformation fields. Since facial movements are comprised of a complex sequence of muscle actions, which are unique to an individual, it is almost impossible to imitate an individual's facial expressions. This suggests that we can use the facial deformation fields for face recognition.

Face recognition is the most popular and successful application of image analysis and understanding. It has wide-ranging applications in a number of areas such as security, surveillance, virtual reality and human-computer interaction [16]. Face recognition has received renewed interest in recent times, which can be attributed to increased concerns about security, and the rapid developments in enabling technologies. Most of the existing methods for face recognition are based on still images and video. Image-based techniques are mainly interested in the shape, size and position of features such as eyes, nose and mouth. But the face of an individual is unique in other respects as well, such as in the way it moves during facial expressions. Video-based methods use both the temporal and spatial information for face recognition. But they suffer from low image quality and limited resolution. On the other hand, while high-resolution images are available to still-image-based methods, they only contain spatial information. Our method leverages

the strengths of both still-image and video-based methods, using high-resolution images to extract temporal information.

An important feature of our method is that it is completely based on affordable mainstream technologies. The only hardware we require is a high-resolution digital camera. In our experiments, we have observed that a digital camera with resolution greater than 4 mega pixels is sufficient to capture images that enable us to accurately track the skin pores on an individual's face [2] (see Figure 1 below). Nowadays, cameras with such resolutions are commonplace. Some newer cell phone cameras have resolutions as high as 5 mega pixels. The rapid explosion in digital camera technology means that images can now be acquired at hitherto unheard of detail. Our method exploits this detail to accurately track an individual's facial motions.
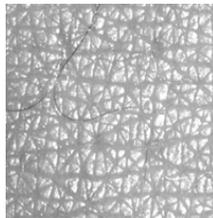


Figure 1. Skin texture 8x8mm. The skin pores are clearly visible

Face and facial expressions are the most visually distinguishing features of an individual and they cannot be easily, if at all, imitated. As such, an accurate and efficient feature tracking method such as DISC has great potential to be successfully applied to face recognition, and this paper presents some of the results that we obtained for face recognition with our method and compares them with still-image-based recognition.

The rest of the paper is organized as follows. Section 3 describes DISC. Section 4 describes our approach. Section 5 presents the results for our method and compares them with some standard algorithms. Section 6 concludes the paper.

## 2. Previous Work

Most of the early still-image-based methods were based on PCA. Eigenfaces [3] was one of the first successful face recognition methods. It can handle variations and blur in the images quite well, but it requires a large number of training images. Moghaddam et. al. [5] improved on the eigenface method by using a Bayesian probabilistic measure of similarity instead of Euclidean distance. PCA+LDA [4] is another popular method. It attempts to produce linear transformations that emphasize the difference between classes while reducing the difference within classes. Elastic Bunch Graph Matching [6] locates landmarks on faces and extracts Gabor jets from each landmark. These jets are used to form a face graph, which is used to compare for similarity. In general, appearance-based methods use multiple images per subject to deal with pose and illumination variations. Zhang et. al. [7] proposed a face recognition method using Harmonic Image Exemplars that is robust to lighting changes and requires only one training image per subject. They later extended this method to be pose-invariant by using Morphable Models to capture the 3D shape variations [8]. Some researchers have proposed 3D face matching techniques. Jain et. al. [9] proposed a face surface matching method that takes into account both the rigid and non-rigid variations. A number of other still-image methods based on neural networks, support vector machines and genetic algorithms have also been proposed for face recognition [16].

Earlier video-based face recognition methods used still-image methods for recognition after detecting and segmenting the face from the video. Wechsler et. al. [10] employed RBF networks for face recognition. McKenna et. al. [11] developed a

method using PCA. Instead of using a single frame for recognition, they implemented a voting scheme based on the results from each frame in the video. A disadvantage is that that the voting scheme results in increased computational cost. Choudhury et. al. [12] use both face and voice for person recognition. They use Shape from Motion to compute the 3D information for the individual's head so as to differentiate between a real person and an image of that person. Li and Chellappa's method [13] is perhaps the closest to our approach. They use Gabor attributes to define a set of feature points on a regular 2D grid. These feature points are tracked to obtain trajectories of the individual in the video, which are then employed to identify that individual. In their experiments, this method performed better than frame-to-frame and voting-based methods, even when there were large lighting changes. This method can track features in low-resolution videos, while our method can more accurately track features in high-resolution images. As high-resolution video capture devices become available in the future, our method may be further extended to be video-based.

## 3. DISC

Digital Image Skin Correlation (DISC) is based on a material science technique called Digital Image Speckle Correlation. DISC analyzes two images taken before and after the skin has deformed and yields a vector field corresponding to the deformation of the skin. It derives the displacement vectors by tracking skin pores in the digital images before and after deformation. As shown in Figure 1, human skin provides abundant features i.e. skin pores, that can serve as tracking markers.
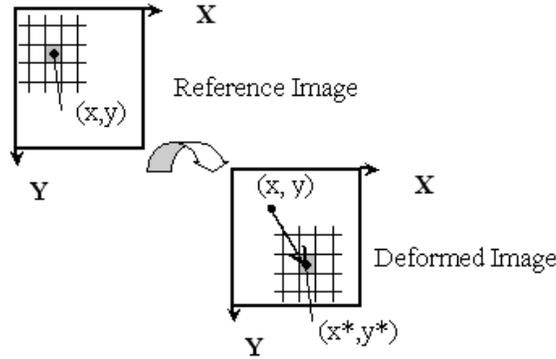


Figure 2. Schematic of DISC

As shown in Figure 2, two images of a face are taken, one before and one after deformation, producing a reference image and a deformed image. DISC divides these two images into smaller windows and compares each window in the reference image with it's neighboring windows in the deformed image. The windows are matched by computing the normalized cross-correlation coefficient between them, and the window with the highest coefficient is considered to be the corresponding window. The correlation is computed as

$$S\left(x,y,u,v,\frac{\partial u}{\partial x},\frac{\partial u}{\partial y},\frac{\partial v}{\partial x},\frac{\partial v}{\partial y}\right) = 1 - \frac{\sum I(x,y)*I^*(x^*,y^*)}{\sqrt{\sum I(x,y)^2 * \sum I^*(x^*,y^*)^2}}, \qquad (1)$$

where,

$$x^* = x + u,$$
$$y^* = y + v.$$

and $I(x,y)$ and $I^*(x^*,y^*)$ are the intensities within the window. The normalized cross-correlation function is computed over the entire window. The coordinate difference between the matched window pair gives the average displacement vector at the center of the reference window. More details on the tracking are available in [1,2].

DISC shows that the distribution of skin pores on the face provides a natural and highly dense texture for tracking. It eliminates the need for cumbersome markers to be attached to an individual's face, which will not provide tracking information at the skin's high pore-resolution in any case.

## 4. Our Approach

In order to recognize an individual, we capture two high-resolution images of the subject's face – one with a neutral expression and the other with a subtle smile. Following, we use DISC to compute the deformation field that represents the way the subject's face deforms. Then, instead of directly using the deformation field as the feature for recognition, we generate two scalar fields from each vector field, which are then used in the recognition process. We call these scalar fields *projection images* since they are obtained by projecting each vector in the deformation field onto the X- and Y-axes, respectively. Alternatively, we could also separate the deformation field into magnitude and orientation fields, and possibly use their gradients as features. We have employed the projection images for the experiments presented in this paper since they have yielded satisfactory results.

We obtain two projection images from each vector field, one corresponding to projections onto the X-axis and the other onto the Y-axis. We could use either of them for recognition or use both and weigh the results. Figure 3 below shows some typical projection images, in this case images obtained from identical twins. We observe strong differences in the projection images (right), while the still images (left) are very similar.
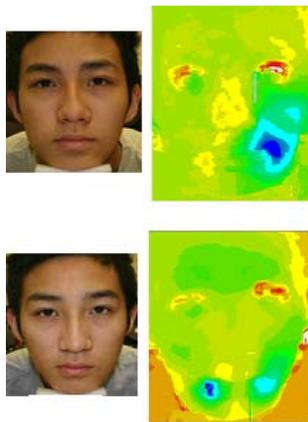


Figure 3. Two projection images (right) of identical twins (left)

In the recognition procedure, we compare the projection image of the candidate subject with the projection images stored in our database. However, before we can compare the projection images of two individuals, we need to deal with the variations in geometry. Even if the images are of the same view, we might have to scale, rotate or shift them in order to align the images. The first step would be to align the eyes and then adjust the aspect ration so as to align the mouth. But then other facial features may still be misaligned. The solution to dealing with different facial geometries is to warp both projection images to an average geometry. Before we can align or warp the projection images, we need to find the fiducial points for registration. Though we use the information derived from the deformation fields for

classification, we do not disregard the still images entirely. Instead, we use them to identify the fiducial points such as the corners of the eye, mouth and nose bridge. We then use this information to warp the projection images.

We use a nearest-neighbor classifier for recognizing the subject's face. For each projection image in the database, we align it with the subject's projection image and compute the similarity between them. The measure of similarity we use is the normalized cross-correlation coefficient. Normalized cross-correlation computes the similarity between two functions, which in our case are the projection images. If *I(x,y)* and *I\*(x,y)* are the two projection images under consideration, where *(x,y)* are the image coordinates, the normalized cross-correlation between these projection images is given as:

$$S = 1 - \frac{\sum I(x,y) * I^*(x, y)}{\sqrt{\sum I(x,y)^2 * \sum I^*(x, y)^2}}, \tag{2}$$

The distance between the projection images is computed as

$$d = 1 - s \quad . \tag{3}$$

The subject is recognized by finding the projection image in the database that is closest to that of the subject.

## 5. Experiments

In order to validate the effectiveness of our framework and test its robustness, we compared the ability of our approach with that of still-image-based approaches in the task of recognizing people with heavy make-up. For this purpose, we have created two small databases, since the standard face recognition databases do not contain the kind of images and tracking data we require. The first of these databases consists of ten subjects, with at least three still images per subject. The second database consists of the same ten subjects, but here we store only two images, one with a neutral expression and the other with a subtle smile, as well as the corresponding projection images. The first database is used for the still-image methods, while the second database is employed for our approach. In our experiment, we used the projection images corresponding to the X-axis. We used a Canon EOS 300D/Digital Rebel camera to capture the still-images. This camera can capture images with a maximum resolution of 6.3 mega pixels. Figure 4 shows some of the subjects in our database and the corresponding X-projection images.
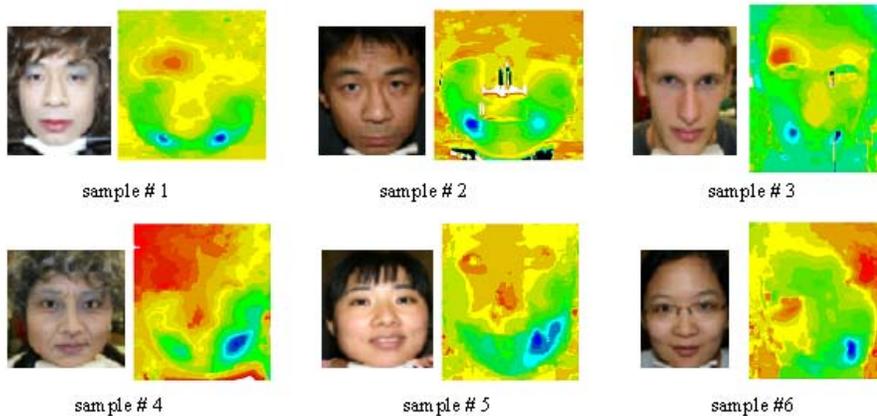


Figure 4. Some of the subjects in our database. The projection images have been color-mapped so as to emphasize the variations

To compare our approach to still-image-based recognition, we used the PCA+LDA algorithm from the CSU Face Identification Evaluation System (CSU FaceId Eval) [15] as our baseline algorithm. CSU FaceId Eval is a collection of standard face recognition algorithms and statistical systems for comparing face recognition algorithms used in the FERET [14] tests. It includes four distinct algorithms – Principle Component Analysis (PCA) [3], a combined Principle Component Analysis and Linear Discriminant Analysis (PCA+LDA) [4], Bayesian Intrapersonal/Extrapersonal Classifier (BIC) [5] and Elastic Bunch Graph Matching (EBGM) [6]. For both of the approaches, we employed normalized cross-correlation as the similarity measure. The distance is computed as

$$d = 1 - s \tag{4}$$

For our experiment, we had a professional make-up artist disguise two of our subjects. We then applied the two approaches in an attempt to recognize them. Table 1 (on the next page) shows the distances we obtained with the PCA+LDA algorithm. The table also shows the query image (the image of the subject wearing make-up, that is presented to the algorithm), the target image (the image of the same subject stored in the database) and the top matches reported by PCA+LDA. From the results in Table 1, we observe that PCA+LDA failed to recognize the individual. For the first individual, the target image is not even among the top three matches and for the second individual the target image is only the second best match. Table 2 shows the distances we obtained with our method. We observe that in both the cases, our method clearly recognized the individual, while this is not easy to do even by visual inspection. Note, for our approach, we used the corresponding projection images for recognition and not the images shown in the first column of Table 2.

## 6. Conclusions

In this paper, we propose a novel method for face recognition that is based on tracking the dense texture of skin pores naturally present on the face instead of placing cumbersome and less dense markers on the person's face. Our approach combines the strengths of both the still-image methods and video-based methods. It requires two subsequent high-resolution images of the individual performing a subtle facial motion, such as a faint smile. These images may be acquired by any higher-end consumer-grade camera. A central part of our approach is the deformation field that is computed for each individual using our Digital Image Skin Correlation (DISC) method. During the recognition process, projection images are generated from these deformation fields and a distance metric to projection images of individuals stored in the database is computed. The initial results presented in this paper verify the potential of our method to recognize faces accurately, even with heavy make-up, where the tested existing popular still-image method had difficulties.

This paper shows that our skin pore tracking-based face recognition technique is a promising way to accurately recognize faces. Our future work will seek to compute 3-dimensional deformation fields, to make our approach pose-invariant. We are also exploring the use of higher-level analysis methods. To this end, we are characterizing the deformation fields using vector derivatives and flow analysis techniques. Also, to more effectively evaluate the performance of our method in a variety of scenarios, we are planning to considerably expand our database. Another goal of ours is to optimize the DISC algorithm and its implementation, which will reduce the computation time down to seconds and also allow us to compute the deformable fields faster.

Table 1. Distances obtained with the LDA algorithm

| Query Image | Target Image | Best Match | 2nd Best Match | 3rd Best Match |
|---|---|---|---|---|
|  |  |  |  |  |
| Distance | 0.83 | 0.32 | 0.45 | 0.73 |
| Query Image | Target Image | Best Match | 2nd Best Match | 3rd Best Match |
|  |  |  |  |  |
| Distance | 0.56 | 0.47 | 0.56 | 0.925 |

Table 2. Distances obtained with our DISC algorithm. Note, comparisons were done on the projection images and not the images shown here

| Query person | Target person | Best Match | 2nd Best Match | 3rd best Match |
|---|---|---|---|---|
|  |  |  |  |  |
| Distances | 0.077 | 0.077 | 0.709 | 0.797 |
| Query person | Target person | Best Match | 2nd Best Match | 3rd Best Match |
|  |  |  |  |  |
| Distances | 0.088 | 0.088 | 0.709 | 0.789 |

## 7. Acknowledgements

## 8. References

1.  H.A. Bruck, S.R. McNeil, M.A. Sutton and W.H. Peters, "Digital Image Correlation Using Newton-Raphson Method of Partial Differential Correction," *Experimental Mechanics*, pp. 261-267, September 1989.
2.  E. Guan, S. Smilow, M. Rafailovich and J. Sokolov, "Determining the Mechanical Properties of Rat Skin with Digital Image Speckle Correlation," *Dermatolog*, vol. 208, no. 2, pp. 112-119, 2004.
3.  M.A. Turk and A.P. Pentland, "Face Recognition Using Eigenfaces," In Proceedings, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586-591, June 1991.

4.  W. Zhao, R. Chellappa, and A. Krishnaswamy, "Discriminant Analysis of Principal Components for Face Recognition," In Proceedings, *International Conference on Face and Gesture Recognition*, pp. 336, 1998.

5.  B. Moghaddam, C. Nastar, and A. Pentland, "A Bayesian similarity measure for direct image matching," In Proceedings, In Proceedings, *International Conference on Pattern Recognition*, pp. 350–358, 1996.

6.  L. Wiskott, J. Fellous, N. Kruger and C. Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Tranactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775-779, 1997.

7.  L. Zhang and D. Samaras, "Face Recognition Under Variable Lighting using Harmonic Image Exemplars," In Proceedings, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 19-25, 2003.

8.  L. Zhang and D. Samaras, "Pose Invariant Face Recognition under Arbitrary Unknown Lighting using Spherical Harmonics," In Proceedings, *Biometric Authentication Workshop (in conjunction with ECCV 2004)*, 2004

9.  X. Lu, A.K. Jain, "Deformation Analysis for 3D Face Matching," To appear in, *IEEE Workshop on Applications of Computer Vision*, 2005

10. H. Wechsler, V. Kakkad, J. Huang, S. Gutta, and V. Chen, "Automatic Video-based Person Authentication using the RBF Network," In Proceedings, *International Conference on Audio and Video-Based Person Authentication*, pp. 85-92, 1997.

11. S. McKenna and S. Gong, "Recognizing Moving Faces," In H.Wechsler, P.J. Phillips,V. Bruce, F.F. Soulie, and T.S. Huang, editors, *Face Recognition: From Theory to Applications*, NATO ASISeries F, vol. 163, 1998.

12. T. Choudhury, B. Clarkson, T. Jebara and A. Pentland, "Multimodal Person Recognition using Unconstrained Audio and Video," In Proceedings, *International Conference on Audio- and Video-Based Person Authentication*, pp. 176-181, 1999.

13. B. Li and R. Chellappa, "Face Verification through Tracking Facial Features", *Journal Optical Society of America,* vol. 18, no. 12, pp. 2969-81, 2001.

14. P.J. Phillips, H.J. Moon, S.A. Rizvi, and P.J. Rauss, "The FERET Evaluation Methodology for Face Recognition Algorithms", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090-1104, 2000.

15. R. Beveridge, "Evaluation of Face Recognition Algorithms", *http://cs.colostate.edu/evalfacerec.*

16. W. Zhao, Chellappa R, Philips PJ, Rosenfeld A, "Face Recognition: A Literature Survey", *ACM Computing Surveys*, vol. 35, no. 4, pp. 399-458, 2003.