

Motion Compensation for Face Recognition Based on Active Differential Imaging

Xuan Zou, Josef Kittler, and Kieron Messer

Centre for Vision, Speech and Signal Processing
University of Surrey, United Kingdom
{x.zou,j.kittler,k.messer}@surrey.ac.uk

Abstract. Active differential imaging has been proved to be an effective approach to remove ambient illumination for face recognition. In this paper we address the problem caused by motion for a face recognition system based on active differential imaging. A moving face will appear at two different locations in the ambient illumination frame and combined illumination frame and as result artifacts are introduced to the difference face image. An approach based on motion compensation is proposed to deal with this problem. Experiments on moving faces demonstrate that the proposed approach leads to significant improvements in face identification and verification results.

1 Introduction

The illumination problem in face recognition, is one of those challenging problems which remain to be addressed [8]. The variation in face appearance caused by illumination change can be much larger than the variation caused by personal identity [5]. The approaches to this problem can be divided into “Passive” and “Active”. In “Passive” approaches, attempts are made to process images which have already been affected by illumination variations. Such approaches are either based on illumination modelling, photometric normalisation, or the use of illumination-insensitive features. “Active” approaches usually involve additional devices (optical filters, active illumination sources or specific sensors) that actively obtain modalities of face images that are insensitive to or independent of illumination change. These modalities include face shape (depth map or surface normal) and thermal/near infrared images.

Active differential imaging is an important variant of active sensing to minimise the illumination problem. An active differential imaging system consists of an active illuminant and an imaging sensor. During the imaging process, two images are captured: the first one is taken when the active illuminant is on, the second one is taken when the illuminant is off. In the first image the scene is illuminated by the combination of both the ambient illumination and the active illumination, while in the second image the face is illuminated only by ambient illumination. Therefore the difference of these two images contains the scene viewed under active illumination only, which is completely independent of ambient illumination. Some sample images are shown in Fig. 1. Near-Infrared is often

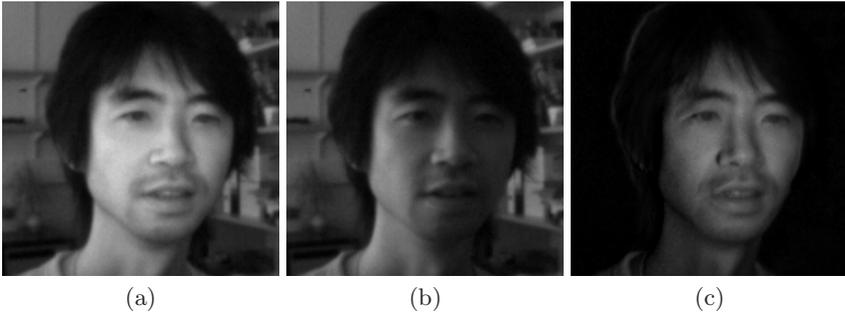


Fig. 1. Face images under combined illumination(a), ambient illumination(b) and their difference image(c), without any motion

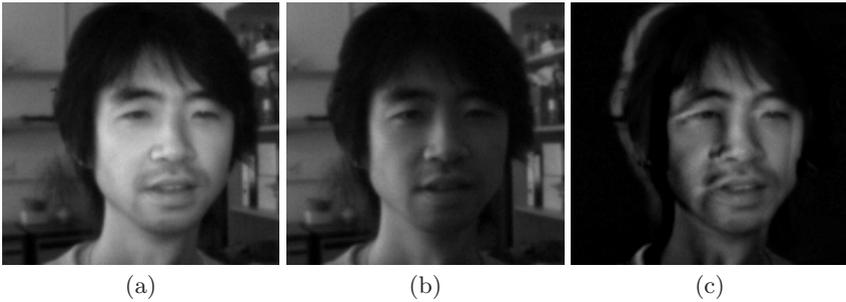


Fig. 2. Face images under combined illumination(a), ambient illumination(b) and their difference image(c), when the face is moving

chosen as the active illumination source for active differential imaging due to its invisibility, which makes the whole system unobtrusive.

Recently, the idea of active differential imaging has been applied to illumination invariant face recognition[9][10][1][6] and significant advantages on face recognition performance are reported for faces under varying illuminations. Specific sensors which can perform differential imaging have been proposed[3][7].

However, despite its success in removing ambient illuminations for a still scene, the active differential imaging system is detrimentally affected by any motion of the subject in the field of view. A moving subject will appear at two different locations in the combined illumination frame(C -frame) and ambient illumination frame(A -frame), respectively, which results in a difference image with “artifacts” as shown in Fig. 2.

To the best of our knowledge, this problem has not been addressed before. In this paper, we propose a motion compensation technique developed to cope with this problem. The face capture system captures a C -frame and an A -frame alternately. The motion between two C -frames is estimated, and a “virtual” C -frame is computed by performing motion interpolation. The difference image between this “virtual” C -frame, and the A -frame captured between the above

two C -frames is used for recognition. The advantage of applying the proposed approach is proved by the improvement in face recognition results on a database containing moving face sequences.

The paper is organised as follows: A detailed look at the problem caused by motion is given in Section 2. Section 3 describes the proposed approach to deal with the motion problem. The information about a moving face database for our experiments is provided in Section 4. The details and results of the experiments carried out are presented in Section 5. Conclusions are drawn in Section 6.

2 Motion Problem of Active Differential Imaging

Two problems can be introduced by a motion for a capture system based on active differential imaging. The first one is motion blur for each captured image. However, this is a general problem for all imaging systems. Fortunately face recognition based on the commonly adopted subspace approaches is relatively robust to the degradation in resolution and consequently it is not seriously affected by motion blur. Therefore motion blur is not a major problem to be addressed. The second one is posed by the motion artifacts in the difference image. Since the C -frame and A -frame are captured at different times, the difference image will be quite different to what it is supposed to be due to the position change of the object. The displacement between object positions in two frames is directly related to the time interval between this two successive captures and the moving speed. As shown in Fig. 2, this problem is significant for a face capture system based on differential imaging. Artifacts are prominent especially around the edges, such as eyelids, nose and mouth. This paper focuses on the second problem.

An example is given below to show how the face similarity degrades due to displacement of the faces. A C -frame and an A -frame are taken for a still face with 60 pixels in inter-ocular distance as shown in Fig. 3. We manually shifted A -frame by 1 to 10 pixels from its original position horizontally, or vertically. The faces in the resulting difference images are registered with reference to the original position in the C -frame, cropped and normalised to 55×50 patches, and histogram equalised, as shown in Fig. 4. Fig. 5 shows that the similarities between

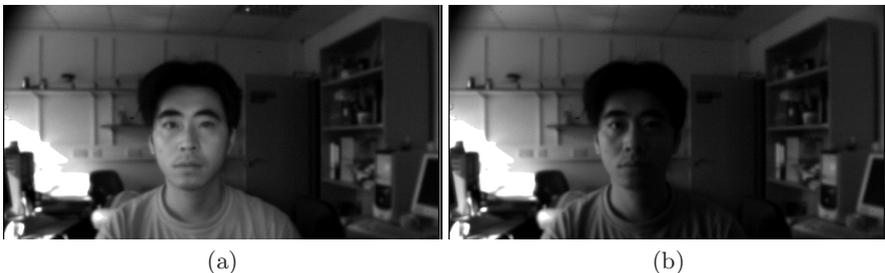


Fig. 3. Combined illumination frame (a) and ambient illumination frame (b)



Fig. 4. original face template (a), and resulting difference image when A -frame shifted $(-10, 7, 3, 0, 3, 7, 10)$ pixels horizontally (b) and vertically (c) from C -frame

the resulting difference images and the face template keep decreasing when displacement between the C -frame and A -frame in any direction (left, right, up or down) increases. The similarity is measured by the Normalised Correlation (NC) score in the original image space, Principal Component Analysis (PCA) subspace, and Linear Discriminant Analysis (LDA) subspace. PCA and LDA subspaces are built from a near-infrared face database with 2844 images of 237 people. NC in an LDA subspace usually gives much better results for face recognition than NC in the image space and PCA subspace, however, NC in the LDA subspace is much more sensitive to the displacement than in the other two spaces, according to Fig. 5.

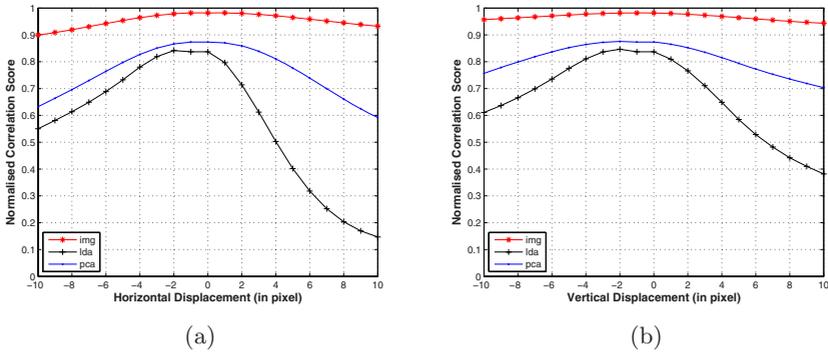


Fig. 5. NC score drops when A -frame shifts from C -frame horizontally (a) and vertically (b)

Therefore, the performance of a face recognition system based on active differential imaging will degrade when faces are moving. For a general-purpose camera, the time interval between two frames is 40ms (CCIR) or 33ms (EIA), which is so long that the motion effect maybe significant. A hardware solution is specific high speed sensors, such as the sensor developed by Ni[7], which can provide a capture speed of 100 images per second. However, due to the high price of custom designed devices, a software solution is always desirable.

The problem cannot be solved using only the difference image. First, simply applying subspace approaches does not work: as discussed above, none of the commonly used face subspace representations is both insensitive to this motion effect and discriminative enough for face recognition. Second, motion information cannot be recovered from the difference image to remove the motion effect. It is

also impossible to align the faces in the successive C -frame and A -frame because faces are in different illuminations in these two frames. We propose to use two nearest C -frames to obtain motion information and interpolate a virtual C -frame, the motion effect can then be removed from the difference image of the A -frame and the virtual C -frame.

3 Motion Compensation for Moving Face

Assuming the face is moving at the same speed in the same direction between t_i , the time when the first C -frame C_i is captured, and t_{i+2} , the time when the second C -frame C_{i+2} is captured. We can apply interpolation to obtain a virtual C -frame C'_{i+1} as “captured” at t_{i+1} , which is the same time when the frame A_{i+1} is captured. Therefore, the faces in the frame C'_{i+1} and frame A_{i+1} are exactly at the same location. As a result, the motion effect is removed in the difference image between the frame C'_{i+1} and frame A_{i+1} . This approach is illustrated in Fig. 6.

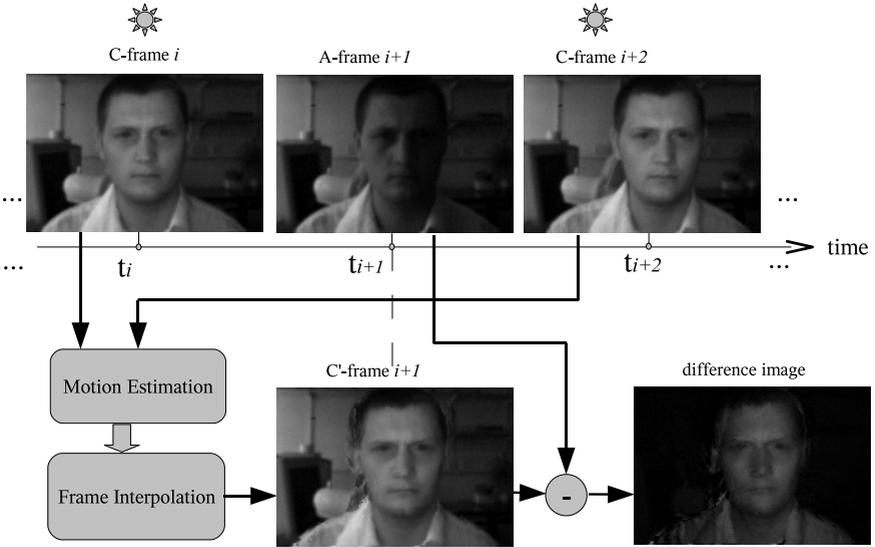


Fig. 6. Illustration of our proposed approach

The robust optical flow estimation method by Black and Anandan [4] is applied to obtain a dense motion field between two successive C -frames: C_i and C_{i+2} . For the image intensity function $I(x, y, t)$ defined on the region S , the estimation of optical flow field $U = \{(u_s, v_s) \mid s \in S\}$ can be treated as a minimization problem of a cost function $E(u, v)$ of the residual errors from the

data conservation constraint and the spatial coherence constraint. The robust formulation presented in [4] is as below:

$$E(u, v) = \sum_{s \in S} \left[\lambda_D \rho_D(I_x u_s + I_y v_s + I_t, \sigma_D) + \lambda_S \left(\sum_{n \in \mathcal{G}_s} \rho_S(u_s - u_n, \sigma_S) + \sum_{n \in \mathcal{G}_s} \rho_S(v_s - v_n, \sigma_S) \right) \right] \quad (1)$$

where I_x , I_y and I_t are the partial derivatives of $I(x, y, t)$. Both ρ_D and ρ_S are Lorentzian function:

$$\rho(x, \sigma) = \log\left(1 + \frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right) \quad (2)$$

λ_D and λ_S are weights for the data conservation term and spatial coherence term, respectively. σ_D and σ_S are the parameters controlling the shape of Lorentzian function and the threshold for outliers. Using Lorentzian function instead of the quadratic function used in Least Square Estimation, the influence of the outliers for the data conservation constraint and the spatial coherence constraint can be reduced. A coarse-to-fine strategy is employed to cope with large motions.

If u and v represent the horizontal and vertical motion between frame C_i and C_{i+2} , then the motion between C_i and C'_{i+1} are $u \frac{t_{i+1}-t_i}{t_{i+2}-t_i}$ and $v \frac{t_{i+1}-t_i}{t_{i+2}-t_i}$ based on linear interpolation. C'_{i+1} can be warped from C_i based on:

$$C'_{i+1}(p) = \sum_{s=0}^3 d^s C_i(p_o^s) \quad (3)$$

where $C'_{i+1}(p)$ is the grey value of a pixel p with coordinates (m, n) in frame C'_{i+1} , $\{p_o^s\}_{s=0,\dots,3}$ are the 4 nearest neighbors of the original subpixel location (m_o, n_o) of p in the frame C_i with $m_o = m - u(m, n) * \frac{t_{i+1}-t_i}{t_{i+2}-t_i}$, $n_o = n - v(m, n) * \frac{t_{i+1}-t_i}{t_{i+2}-t_i}$. $\{d^s\}_{s=0,\dots,3}$ are the weights related to the distances between (m_o, n_o) and $\{p_o^s\}_{s=0,\dots,3}$. Fig. 7 illustrates the interpolation process.

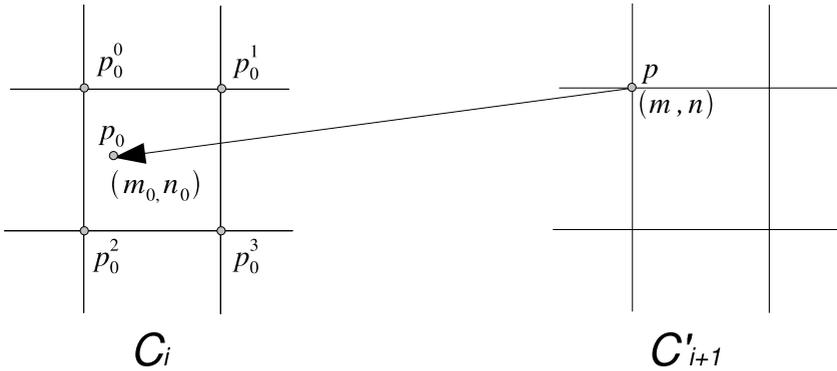


Fig. 7. Illustration of the interpolation process

4 Capture System and Database Capture

An experimental face capture system based on active differential imaging was used to capture moving face data for our experiments. The image sensor has a high resolution of 755×400 pixels. A face database of 37 subjects was captured in an indoor environment near the window with sunlight coming in. Each subject sat 1 meter away from the camera, and was asked to keep moving his/her face. Two sessions of data were recorded for each subject, with 29 *C*-frames and 29 *A*-frames captured continuously for each session. Ambient illumination was mainly from left for one session and from right for the other. Another 6 difference images for each subject were captured when the subject sat still. These images will serve as gallery images for the identification and verification experiments.

5 Experiments

To show the advantage brought by the proposed motion compensation approach, face recognition experiments are conducted on two sets of difference face images for comparison. The first set contains the original difference face images between every pair of *C*-frame and *A*-frame, without considering the motion issue. For the second set, the proposed approach is applied to obtain the difference images without motion effect. Faces in both sets are geometrically normalised based on the manually marked eye positions in the corresponding *C*-frame, cropped to

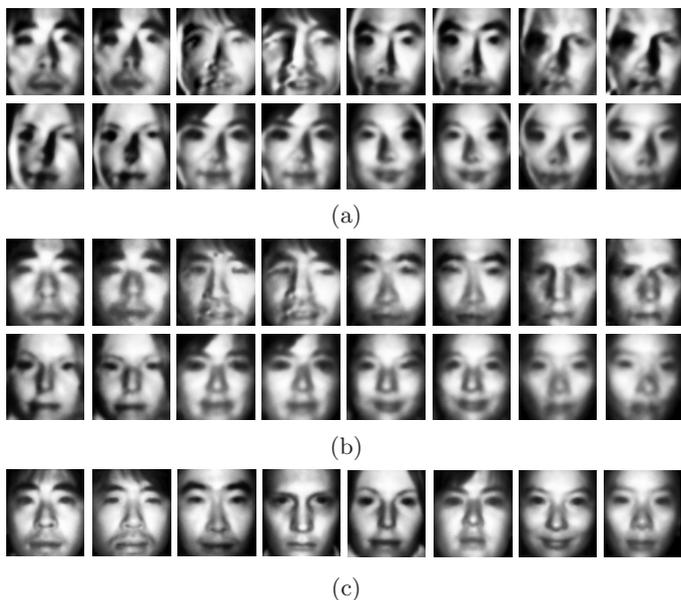


Fig. 8. Faces in difference images without motion compensation (a) and with motion compensation (b), and corresponding template images(c) for subjects in (a) and (b)

55*50 image patches, and photometrically normalised using histogram equalisation. Examples for faces in both sets are shown in Fig. 8 and it can be seen that the motion artifacts have been removed after motion compensation. Each set contains 2072 faces(37 subjects \times 2 sessions \times 28 difference images).

5.1 Improvement in Face Similarity to Template

A histogram of the motion between every two frames for the whole database is shown in Fig. 9.(a). For every pair of C -frames, the motion value recorded in the histogram is the average length of all those motion vectors of the motion foreground pixels (pixels with motion vector magnitude above a threshold 0.8). Since faces are normalised based on the eye positions for recognition, the same absolute displacement will have different influence for faces with different inter-ocular distances. Therefore relative motion, which is the motion value over the inter-ocular distance, is applied here to measure motion.

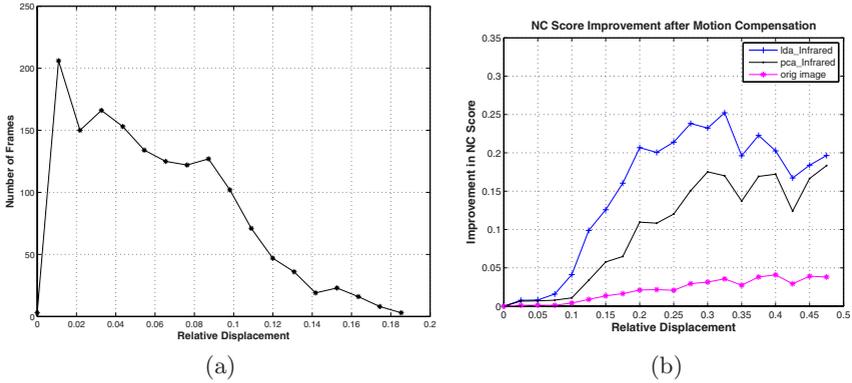


Fig. 9. (a)Histogram of the relative motion in all sequences. (b)Average improvement in NC score after using motion compensation in terms of the relative motion.

According to Fig.9.(b), NC score in LDA subspace is improved by motion compensation more significantly than in PCA subspace and original image space. When motion is too small, applying motion compensation brings a little improvement. But for moderate motion, the improvement is significant. When motion is too large the improvement decreases because large motion tends to introduce large pose change which has a negative influence on motion estimation.

5.2 Face Identification Experiments

Face identification experiments are carried out using Nearest Neighbor classifier based on full face, left/right half of the face, and the fusion of left/right halves of the face. The similarity is measured by NC score in the respective LDA subspaces for full face, left face, or right face. For the fusion case, the Sum rule[2] is applied to fuse the similarity score in the LDA subspace for left face and right face. As

Table 1. Face Identification Rank-1 Error Rates(%) on Moving Faces

	with motion compensation	without motion compensation
full face	12.07	31.71
left face	14.23	32.09
right face	10.28	26.93
fusion of left/right faces	6.13	14.96

Table 2. Verification Half Total Error Rates(%) on Moving Faces

	with motion compensation	without motion compensation
full face	7.88	14.13
left face	8.69	15.63
right face	8.72	15.33
fusion of left/right faces	6.42	10.95

shown in Table 1, after motion compensation the errors decrease to less than half of the errors achieved without performing motion compensation. Applying fusion technique gives the best identification result with an error rate 6.13%.

5.3 Face Verification Experiments

For the first round of verification experiment, the data of one session is used for evaluation, and the other session for testing. Then evaluation and testing sessions are switched for the other round of testing. Every test image is used to make claim to all 37 identities in the gallery. So in total there are 1036 instances of true claims and 1036*36 instances of false claims during evaluation and the same numbers of true claims and false claims in the testing stage. The average verification errors are reported in Table 2. Again, significant improvement is achieved by applying motion interpolation. Fusion has the best performance with the lowest error rate 6.42%.

6 Conclusion and Future Work

Motion causes problems for face recognition systems based on active differential imaging. The similarities between probe images and the template decrease due to the artefact in the captured face image caused by motion. In this paper we proposed an approach based on motion compensation to remove the motion effect in the difference face images. A significant improvement has been achieved in the results of face identification and verification experiments on moving faces. Since the motion artifact also introduces difficulties in automatic face localisation, we are now investigating the performance boost of the system in the fully automatic operation scenario.

Acknowledgement

The support from the Overseas Research Students Awards Scheme (ref. 2003040015) is gratefully acknowledged.

References

1. Hizem, W., Krichen, E., Ni, Y., Dorizzi, B., Garcia-Salicetti, S.: Specific sensors for face recognition. In: Proceedings of IAPR International Conference on Biometric (2006)
2. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(3), 226–239 (1998)
3. Miura, H., et al.: A 100frame/s cmos. active pixel sensor for 3d-gesture recognition system. In: Proceeding of IEEE Solid-State Circuits Conference, pp. 142–143. IEEE Computer Society Press, Los Alamitos (1999)
4. Black, M.J., Anandan, P.: The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding* 63(1), 75–104 (1996)
5. Moses, Y., Adani, Y., Ullman, S.: Face recognition: the problem of compensating for the illumination direction. In: Eklundh, J.-O. (ed.) *ECCV 1994. LNCS*, vol. 801. Springer, Heidelberg (1994)
6. Ni, Y., Krichen, E., Salicetti, S., Dorizzi, B.: Active differential imaging for human face recognition. *IEEE Signal Processing letter* 13(4), 220–223 (2006)
7. Ni, Y., Yan, X.L.: Cmos active differential imaging device with single in-pixel analog memory. In: Proc. IEEE Eur. Solid-State Circuits Conf., pp. 359–362. IEEE Computer Society Press, Los Alamitos (2002)
8. Zhao, W., Chellappa, R., Rosenfeld, A.: Face recognition: A literature survey. *ACM Computing Surveys* 35, 399–458 (2003)
9. Zou, X., Kittler, J., Messer, K.: Face recognition using active Near-IR illumination. In: Proceedings of British Machine Vision Conference, pp. 209–219 (2005)
10. Zou, X., Kittler, J., Messer, K.: Ambient illumination variation removal by active Near-IR imaging. In: Proceedings of IAPR International Conference on Biometric, January 2006, pp. 19–25 (2006)