Face Recognition Using Active Near-IR Illumination

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

A new approach to overcome the problem caused by illumination variation in face recognition is proposed in this paper. Active Near-Infrared (Near-IR) illumination projected by a Light Emitting Diode (LED) light source is used to provide a constant illumination. The difference between two face images captured when the LED light is on and off respectively, is the image of a face under just the LED illumination, and is independent of ambient illumination. In preliminary experiments with various ambient illuminations, significantly better results are achieved for both automatic and semi-automatic face recognition experiments on LED illuminated faces than on face images under ambient illuminations.

1 Introduction

Although face recognition has been an active research area for long time, several major problems remain to be solved. The effect of variation in the illumination conditions, which causes dramatic changes in the face appearance, is one of those challenging problems [14]. The performance of an automatic face recognition/verification system will degrade with the existence of illumination changes, because these changes cause problems for every stage of the system: from face detection/localization to the final recognition/verification stage.

Existing approaches addressing this problem fall into two main categories. The first category includes methods attempting to model the behaviours of the face appearance change as a function of illumination. However, the modelling of the image formation generally requires the assumption that the surface of the object is Lambertian, which is violated for real human face. In the other category, the goal is to remove the influence of illumination changes from face images or extract face features that are invariant to illumination. Various photometric normalization techniques have been introduced to pre-process face images, and a comparison of five photometric normalisation algorithms used in pre-processing stage for face verification on the Yale B database [5], the BANCA database [1] and the XM2VTS database [9] can be found in [11]. In PCA based face recognition the first several leading principal components are suggested to be influenced heavily by illumination and thus discarded to achieve illumination invariant face recognition [2]. Face shape (depth map or surface normal) [3] or face images in multiple spectra [8] are used in face recognition as illumination invariant features. However, face shape

BMVC 2005 doi:10.5244/C.19.24

acquisition always requires additional devices and is usually computationally expensive. The problem with using multi-spectral images is that although invisible spectral images can be invariant to visible illumination change, there can be big differences in the invisible spectra in real application environments. For example, the infrared component varies greatly between indoor and outdoor environments.

In this paper we present a completely different approach to address the illumination variation problem. Rather than studying passively the variation of illumination itself or attempting to extract illumination invariant feature, we actively create an active and invariant illumination condition for both gallery images and probe images. This way, there is no need to worry about changes to environmental illumination. In our system an LED lamp is used to provide active Near-IR illumination. Two face images are captured for every subject. The first capture is done when the LED lamp is on, and the other capture is done when LED is off. The difference of these two images is an image of the face illuminated only by the Near-IR illumination provided by the LED lamp, and is independent of environmental illumination. A nice property of Near-IR illumination is its invisibility, which ensures that the capture is non-intrusive.

The rest of the paper is organized as follows: A brief review of the previous applications of active Near-IR illumination in computer vision is presented in Section 2. Section 3 provides an overview of the whole hardware system. Section 4 describes the acquisition of a face database. In Section 5 we give the details and results of the recognition experiment performed on this face database. Finally conclusions are drawn in Section 6.

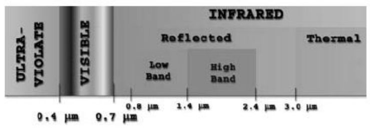
2 Active Near-IR Illumination

Active vision is not new in the computer vision area. In structure/coded light approaches, light patterns are projected onto object surfaces to facilitate 3D surface reconstruction. Active illumination is often used for shadow removal.

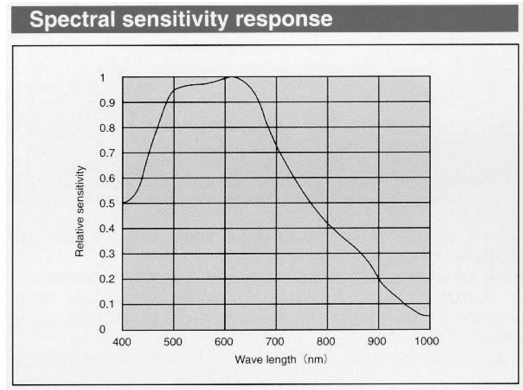
The Near-IR band falls into the reflective portion of the infrared spectrum, between the visible light band (0.3μm-0.6μm) and the thermal infrared band (2.4μm-100μm)(see Fig 1 a). Thus it has advantages over both visible light and thermal infrared. Firstly, since it can be reflected by objects, it can serve as active illumination source, in contrast to thermal infrared. Secondly, it is invisible, making active Near-IR illumination unobtrusive.

In [13] infrared patterns are projected to the human face to solve the correspondence problem in multi-camera 3D face reconstruction. Dowdall and et. al performed face detection on Near-IR face images [4]. They claimed the face skin under active Near-IR illumination yielded much smaller variation as compared to the variation across human races and dynamic visible light illumination conditions. Skin region is detected based on the fact that skin has different responses to the upper band and the lower band of near-IR illumination. Morimoto and Flickner [10] proposed a multiple face detector which deployed a robust eye detector, exploiting the retro-reflectivity of the eyes. One Near-IR light set is used to provide bright pupil image, whilst another setting is used to generate dark pupil image, while keeping similar brightness in the rest of the scene. The pupils are very prominent and easy to detect in the difference image. Similar eye detectors using active illumination are used in [7] for 3D face pose estimation and tracking.

Although active Near-IR illumination has been widely used in face processing as detailed above, the novel idea advocated in this paper is to use it to provide constant illumi-



(a)



(b)

Figure 1: (a) The EM spectrum [4]. (b) The spectral response of KP-M3RP camera

nation for face recognition.

3 System overview

The hardware of our face capture system consists of an LED light source, a camera, the frame grabber card and a computer. The LED lamp in our system has a set of 42 LEDs with adjustable output and a photocell. The LED beam angle is 30 degrees. The peak output wavelength is 850nm, and it lies at the invisible and reflective light portion of the Electromagnetic spectrum.

The sensor we use is the HITACHI KP-M3RP CCD camera, which is a monochrome device with Near-IR sensitivity (see Fig 1 b). The peak sensitivity lies at 620nm and the sensitive spectrum ranges from 400nm to 1000nm. To ensure the difference of the images taken when the LED lamp is on and off represents the face illuminated just by the LED lamp, the Automatic Gain Control (AGC) in the camera is disabled and the aperture is tuned to ensure that the camera will have a linear response to the scene radiance, and that there is no saturation in the captured images. If the images are unsaturated, the face image under the combined LED and ambient illumination is exactly the addition of the face images obtained separately under just LED illumination and under just the ambient illumination, respectively.

The LED lamp is attached close to the camera so that the reflective component of the Near-IR light from the eyes will be projected straight into the camera. This allows us to obtain face images with prominent bright pupils. A ring with 4 fluorescent lamps is used to capture face images under varying illumination conditions. This is not a part of the envisage practical system but just for the purpose of capturing a dataset on which comparison with the LED illuminated face can be made.

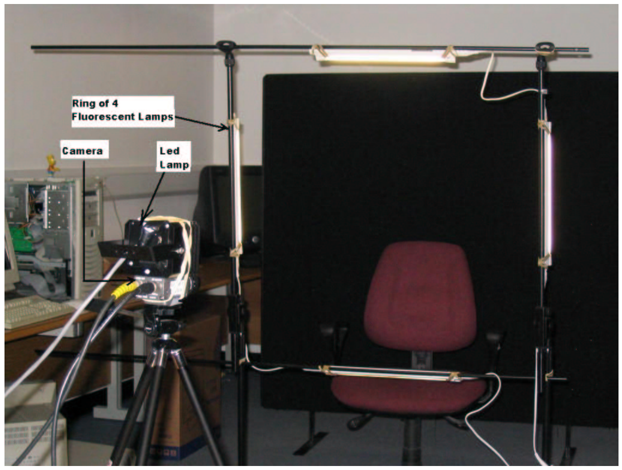


Figure 2: A picture of face capture system

4 Face Database Acquisition

A database of face images of 18 subjects has been captured indoor. This database contains two subsets: ambient faces (face images under ambient illumination) and LED faces (face images under LED illumination). Two capture sessions have been conducted with a time interval of several weeks. For each session, 4 different illumination configurations are used with light sources directed individually from left, bottom, right and top. 6 recordings were acquired for each illumination configuration. For each recording a face image under ambient illumination only and one under combined ambient and LED illumination are captured. Therefore, we have 18*2*4*6 =864 ambient faces and the same amount of ambient plus LED faces.

During each recording, the subject was asked to sit as still as possible, without talking, in front of the lighting ring, and face frontally the camera. Just one of those four lamps was switched on, and the LED lamp was turned on and off. We captured face images before and after the LED lamp was turned on to obtain a set. Usually it took 1-2 video frames time for the LED lamp output to become stable. Assuming little movement of the subject between the takes, we can then compute the difference image of the above two images to obtain the face under just LED illumination, which is independent of the ambient illumination. However, the face is in exactly the same pose in both ambient face image and LED face image within the same recording, which means illumination conditions are the only distinctive factor in the image formation. Hence the difference in the face recognition results on the two sets of images respectively are solely due to the difference in illumination condition rather than any other factors like variations in pose and expression.

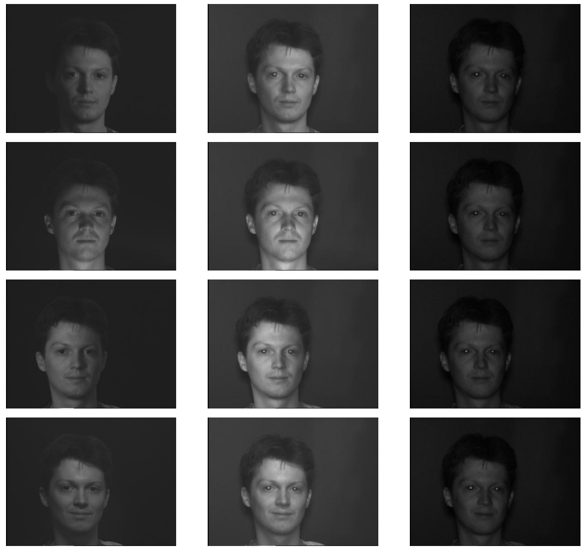


Figure 3: Ambient faces (the left column), combined illumination faces (the middle column) and LED illuminated faces (the right column) under 4 different illumination configurations. The ambient illumination change caused significant differences in the appearance of the whole face. All important facial features look very different in different illumination conditions. Ambient faces and LED faces are relatively dark because we have to adjust the aperture for the camera to avoid the saturation of the combined illuminated faces.

Figure 4: Ambient face images for the two different sessions. Top row for each subject are images captured in session 1, and bottom row for session 2

Figure 5: LED face images for the two different sessions. Top row for each subject are images captured in session 1, and bottom row for session 2

5 Experiments and Results

5.1 Face Localisation

For all face images, we manually marked the eye centres as ground truth positions, and performed automatic eye centre localization. Different automatic localization algorithms were used for ambient faces and LED faces respectively. For ambient faces, we used the algorithm based on Gaussian Mixture Model (GMM) face feature detector and enhanced appearance model [6], which has been trained on 1000 images from BANCA face database. A simple correlation-based localization algorithm has been applied to LED faces. The reason we used a different approach for LED faces is that usually bright pupils can be found in LED faces and they can serve as a strong features for eye localization. We computed the correlation between the histogram equalized image patch in search window and the histogram equalized mean eye template. Figure 6 shows the automatic detection performance on both data sets. From the localisation errors shown on Figure 6, it is evident that the illumination variations directly lead to the poor performance on ambient faces. With the help of the bright pupils and the consistency in LED illumination, the simple correlation-based approach gives much better results.

Face images are registered according to the manually marked or automatically detected eye centre positions, then cropped and sampled to the same size (55*50). Histogram equalization is applied subsequently. It has been proved in our previous experiments that histogram equalization helps for face verification even on faces under controlled illumination. Fig 7 shows some samples of faces after the histogram equalization has been performed. The resulting images are then projected to an LDA subspace. Since currently there is no Near-IR face database available to us to build an LDA subspace

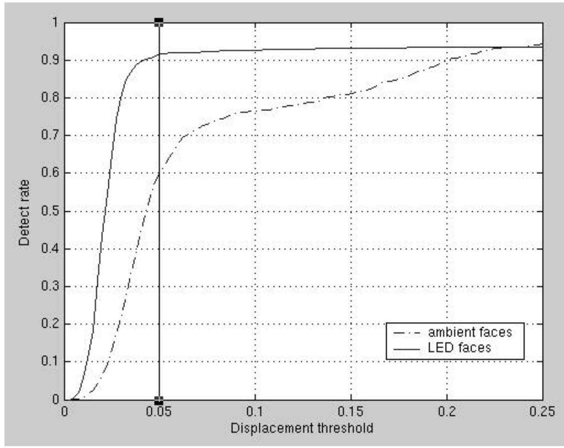


Figure 6: The automatic eye center detection results for LED faces and ambient faces. The detection rate is defined as the percentage of faces in which the automatically detected positions of both eye centers fall within a certain distance from the corresponding ground truth positions. The ratio of the radius of this region and the ground truth of the inter-ocular distance in this image is defined as Displacement Error. Empirically the detection rates with Displacement Errors below 5% are important [6].

specifically for Near-IR faces, the LDA subspace we used is constructed from the PCA projections of all the 2360 face images of 295 subjects in XM2VTS face database and is supposed to be a subspace focusing on discrimitive information among subjects.

5.2 Recognition Experiments and Results

In the above LDA subspace, several different face recognition tests have been carried out on the 4 subsets: manually registered datasets and automatically registered datasets of LED faces and ambient faces. A machine learning toolbox named WEKA [12] developed by University of Waikato has been used to perform experiments on the above datasets. We applied four typical classifiers: Radial Basis Function (RBF) neural network with Gaussian kernel, Adaboost.M1 with Decision Tree J48 as a weak classifier, Support Vector Machine (SVM) with polynomial kernel, and Nearest Neighbor(NN). Pairwise classification is used for SVM to solve multi-class problem.

In the first experiment we measured the face recognition error across different sessions within the same subset. Data in one session serves as the training set and the other one serves as the test set. So the training set for each subject contains 24 images with 6 images in each of the 4 illumination conditions. The same amount images are used as test set. Table 1 shows the error rate obtained using four different classifiers for each test and the average error for each classifier. Although different configurations of different classifiers can give different test results, it is consistently shown for all classifiers that the test results on LED faces are much better than on ambient faces, regardless of the way the faces were registered. The advantage that LED faces offer over ambient faces is significant. So in general LED face data is more readily separable for most type of classifiers. The best

Figure 7: Resulting images after the histogram equalization is performed for manually and automatically registered ambient faces (top 2 rows) and for corresponding LED faces (bottom 2 rows). It is obvious that data from LED faces exhibits much less variation as compared to the data from ambient images. Bright pupils are prominent in the LED face data. There are localisation errors in some automatically registered faces.

Table 1: Error in face recognition across different sessions (in percentage)

Classifiers	Ambient Faces		LED Faces		Test Protocol	
	Manual	Auto	Manual	Auto	Training Set	Test Set
RBF	11.34	35.42	*2.08*	*23.84*	Session 1	Session 2
	3.94	41.67	*1.16*	*30.79*	Session 2	Session 1
	7.64	38.55	*1.64*	*27.32*	average	
Adaboost	14.35	25.46	*1.39*	*6.48*	Session 1	Session 2
	10.65	26.62	*3.01*	*12.5*	Session 2	Session 1
	12.50	26.04	*2.20*	*9.49*	average	
SVM	2.78	12.50	*0*	*1.85*	Session 1	Session 2
	0.23	9.26	*0*	*4.63*	Session 2	Session 1
	1.51	11.88	*0*	*3.24*	average	
NN	3.01	*20.60*	*0*	20.88	Session 1	Session 2
	1.16	28.94	*0*	*25.46*	Session 2	Session 1
	2.09	24.74	*0*	*23.17*	average	

Table 2: Error in face recognition across manually and automatically registered data (in percentage)

Classifiers	Ambient Faces	LED Faces	Test Protocol	
			Training Set	Test Set
RBF	40.28	*11.34*	Manu. Session 1	Auto. Session 2
	32.87	*13.66*	Manu. Session 2	Auto. Session 1
	36.58	*12.50*	average	
Adaboost	40.28	*6.94*	Manu Session 1	Auto. Session 2
	38.66	*10.65*	Manu. Session 2	Auto. Session 1
	39.47	*8.79*	average	
SVM	25.67	*5.56*	Manu Session 1	Auto. Session 2
	21.07	*6.94*	Manu. Session 2	Auto. Session 1
	23.37	*6.25*	average	
NN	28.37	*6.25*	Manu Session 1	Auto. Session 2
	24.54	*9.72*	Manu. Session 2	Auto. Session 1
	26.40	*7.99*	average	

results on manually registered data are achieved using SVM or NN on LED faces, with zero error rate. The best result on automatically registered data is an average error of 3.24% also obtained using SVM on the LED faces.

Another observation one can make from Table 1 is that test results on manually registered data are always better than on automatically registered data. This is not surprising because false detections of the eye positions will lead to a normalized image which is completely different from the reference face image (see Figure 7). Generally we believe the better the eye detection performance is, the smaller difference we can find between test results on manually registered data and automatically registered data. However, due to the nature of classifiers, the difference can also vary greatly among different configurations of different classifiers. For example, average test error is 25.68% worse for LED face if we use RBF on automatically registered data than on manually registered data, but just 3.24% worse if we use SVM.

The second experiment reports the results of face recognition tests across manually registered data and automatically registered data. This test corresponds to a practical application scenario for automatic face recognition. The manually registered data serves as training set and automatically registered data as test set. Table2 shows the test errors using four different classifiers for each test and the average errors for each classifier. Again, test errors on LED faces are much smaller than on ambient faces. Best results are obtained using SVM, with a 6.25% test error on LED faces.

6 Conclusion and Future Work

We proposed in this paper a novel way to address the illumination problem in face recognition by using active Near-IR illumination. Active Near-IR illumination provides a constant invisible illumination condition and faciliates the automatic eye detection by introducing bright pupils. In our preliminary experiments, the actively illuminated faces show better separability for all classifiers than faces under varying ambient illumination. The proposed active Near-IR illumination approach to face recognition is promising for face recognition. Further work will include acquisition of a larger database under both controlled and uncontrolled environment, and extension of current experiments to more tests including cross-illumination tests.

References

[1] E. Bailly-Bailliere and etc. The banca database and evaluation protocal. In *International Conference on Audio- and Video-based Person Authentication*, 2003.

[2] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs fisherfaces : Recognition using class specific linear projection. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 19:711–720, 1997.

[3] K. W. Bowyer, K. Chang, and P. Flynn. A survey of approaches to three-dimensional face recognition. In *Proceedings of International Conference on Pattern Recognition*, 2004.

[4] J. Dowdall, I. Pavlidis, and G. Bebis. Face detection in the near-ir spectrum. *Image Vis. Comput.*, 21:565–578, 2003.

[5] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 23(6):643–660, 2001.

[6] M. Hamouz, J. Kittler, J. K. Kamarainen, P. Paalanen, and H. Kalaviainen. Affine-invariant face detection and localization using gmm-based feature detector and enhanced appearance model. In *Proceedings of the Sixth International Conference on Automatic Face and Gesture Recognition*, pages 67–72, May 2004.

[7] Qiang Ji. 3d face pose estimation and tracking from a monocular camera. *Image and Vision Computing*, 20:499–511, 2002.

[8] S.G. Kong, J. Heo, B. Abidi, J. Paik, and M. Abidi. Recent advances in visual and infrared face recognition - a review. *Computer Vision and Image Understanding*, 2004.

[9] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. Xm2vtsdb: The extended m2vts database. In *International Conference on Audio- and Video-based Person Authentication*, pages 72–77, 1999.

[10] C. H. Morimoto and M. Flickner. Real-time multiple face detection using active illumination. In *Proceedings of the Fourth International Conference on Automatic Face and Gesture Recognition*, 2000.

[11] J. Short, J. Kittler, and K. Messer. A comparison of photometric normalisation algorithm for face verification. In *Proceedings of the Sixth International Conference on Automatic Face and Gesture Recognition*, pages 254–259, May 2004.

[12] I. H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques with JAVA Implementations*. Morgan Kaufmann, 1999.

[13] I. A. Ypsilos, A. Hilton, and S. Rowe. Video-rate capture of dynamic face shape and appearance. In *Proceedings of the Sixth International Conference on Automatic Face and Gesture Recognition*, pages 117–122, May 2004.

[14] W. Zhao, R. Chellappa, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35:399–458, December 2003.