Krzysztof OKARMA, Aleksandra MIĘTUS
ZACHODNIOPOMORSKI UNIWERSYTET TECHNOLOGICZNY W SZCZECINIE,
ul. 26. Kwietnia 10, 71-126 Szczecin

Application of Image Based Rendering to improvement of face recognition using Principal Component Analysis

Ph.D. Eng. Krzysztof OKARMA

Assistant Professor in the Department of Signal Processing and Multimedia Engineering, West Pomeranian University of Technology, Szczecin. Graduated from Faculty of Electrical Engineering (1999) and Computer Science (2001), received his Doctor of Technical Sciences degree (2003) from Faculty of Electrical Engineering, where he is the vice-dean for student affairs. Author of over 90 publications related to signal and image processing and analysis.

e-mail: okarma@zut.edu.pl

M.Sc. Eng. Aleksandra MIĘTUS

Graduated from West Pomeranian University of Technology, Szczecin, Faculty of Electrical Engineering. Received her masters degree in electronics and telecommunications in June 2010. Specialising in multimedia engineering, media production and signal processing. She is a co-author of three publications and an alumni of International Telecommunication Union Youth Forum, Geneva 2009.

e-mail: amietus@gmail.com

Abstract

In the paper the application of Image Based Rendering as a supplementary method useful for PCA-based face recognition is discussed. Presented results are based on the synthetic images of human faces' side views obtained from 3D models and 300 faces taken from FERET database. Application of Image Based Rendering allows the use of en face images rendered as the output based on two side views so the recognition accuracy can be improved.

Keywords: Image Based Rendering, Principal Component Analysis, face recognition.

Zastosowanie metody Image Based Rendering do poprawy rozpoznawania twarzy metodą analizy komponentów głównych

Streszczenie

W artykule omówiono zastosowanie metody Image Based Rendering (IBR) jako techniki uzupełniającej, użytecznej przy rozpoznawaniu twarzy opartym na analizie komponentów głównych (PCA). Typowym zastosowaniem metody IBR jest szybka synteza obrazu o jakości porównywalnej z obrazem referencyjnym na podstawie informacji uzyskiwanych z rzeczywistej kamery zlokalizowanej w innym położeniu niż wirtualna kamera docelowa. Niezbędnym elementem do celów takiej syntezy jest również znajomość mapy głębokości obrazu referencyjnego. Uzyskiwane w taki sposób obrazy mogą być szczególnie użyteczne przy konieczności ich porównania ze wzorcami znajdującymi się w bazie, co jest typowe dla metod klasyfikacji i rozpoznawania wzorców, w tym obrazów. Przedstawione wyniki uzyskane zostały na podstawie syntetycznych obrazów twarzy obserwowanych z boku oraz 300 twarzy uzyskanych z bazy FERET. Jako reprezentatywna technika rozpoznawania twarzy, umożliwiająca dodatkowe wykorzystanie metody IBR, wybrana została metoda PCA, dla której uzyskano zauważalną poprawę skuteczności rozpoznawania twarzy z użyciem proponowanej metody. Zastosowanie metody IBR pozwala wykorzystać frontalne obrazy twarzy wyrenderowane nawet na podstawie obrazu z jednej kamery referencyjnej, co podnosi skuteczność rozpoznawania twarzy. Wykorzystanie obrazów z dwóch kamer bocznych wymaga precyzyjnego pasowania oraz kompensacji wpływu oświetlenia.

Słowa kluczowe: Image Based Rendering, analiza komponentów głównych, rozpoznawanie twarzy.

1. Introduction

Biometrics and applications of face recognition techniques seem to be one of the most promising areas of research in computer science not only as those related to the visitor identification or security systems but also for home use such as e.g. smile detection cameras. Regardless of the presence of many methods used for face recognition purposes such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and its modifications, described e.g. in [1], neural network approach etc., proper classification and recognition of human faces is often strongly dependent on the observation angle. Probably the most promising results are obtained for the en face views which are not always available. Nevertheless, using at least two cameras with additional hardware for the acquisition of depth information (being a simple 3D scanner), fast rendering of such en face view is possible using Image Based Rendering (IBR) approach proposed by Leonard McMillan in 1997 [2]. In the paper some experimental results of such face recognition using a set of 300 single en face images from FERET database and some images rendered from two side views of 3D models are presented. As the recognition method a simple PCA algorithm has been chosen in order to verify the usefulness of the IBR method.

2. Overview of Image Based Rendering

Assuming the availability of depth information together with each image captured by the reference camera (or cameras), the rendering of an image taken by a virtual (destination) camera located somewhere in space is possible, without creation of the 3D models of objects. Such approach is fundamental for the fast real-time rendering using the IBR method in contrast to the time consuming Geometry Based Rendering approach.

The main differences between the IBR based on warping equation used for mapping the reference image into the destination plane and some other methods is its speed. Widely used 3D scanning provides us with extremely high amount of information about a scanned model, so the computational complexity of further processing increases significantly. The scanned model usually requires some additional processing, often using advanced algorithms e.g. ray tracing, photon mapping, illumination modeling etc. Such approach leads to satisfactory results for still images but fast analysis of video frames is often impossible.

Alternatively, image synthesis based on photographs of a scene can be used but the main drawback is the necessary knowledge of depth maps. Nevertheless, the idea of warping can be used in a simplified way e.g. for stitching together a set of images into a panoramical view. The computational complexity of such an algorithm is low enough to implement it in some point-and-shoot cameras working in the fly. Another widely known application of the IBR is the free view-point 3D-TV which allows watching the broadcast from different angles.

The simple IBR method has some drawbacks that may be eliminated using some modifications and additional algorithms. One of the most evident ones is splatting used in order to avoid the presence of holes in the resulting image when the number of available reference pixels is insufficient. In fact, the resolution of the destination image is limited by the amount of useful data present in the reference image.

In the simplest version of the IBR algorithm only a single reference camera is present and further extension for multiple cameras is possible with an additional depth buffer for the reference image. Nevertheless, instead of performing such clear IBR with a 3D warping equation from multiple images and their corresponding depth maps (for each input image), only one intermediate representation called a layered depth image can be created [3]. In such a case, a single representation of the whole scene can be created so that one layered depth pixel stores a set of depth pixels located along one line of sight sorted in front to back order. Thus, when render is being processed the requested view can move away from the original view exposing new surfaces prior not visible.

The single reference camera IBR algorithm can be discussed assuming the pinhole camera model without any distortions, such that all the rays of light focus into one point in 3D space and can be described using the linear mapping function:

$$\bar{d} = \begin{bmatrix} a_i & b_i & c_i \\ a_j & b_j & c_j \\ a_k & b_k & c_k \end{bmatrix} \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = P \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{1}$$

where d represents a ray from the center of projection to the image plane, u and v denote the point coordinates in the image space. The columns of the mapping matrix P denote horizontal and vertical unit vectors and vector directed to the center of projection from the image corner, respectively. They can be calculated for specified coordinates of the reference and destination cameras (C_r and C_d), fields of view and resolutions of both images (for proper scaling). Using also the depth maps, the complete information about the scene geometry is available.

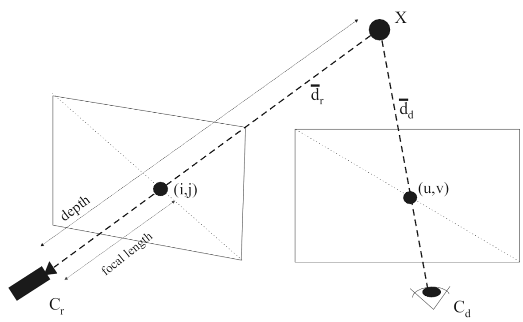

Fig. 1. Configuration of the reference and destination cameras
Rys. 1. Konfiguracja kamery referencyjnej i docelowej

The image coordinates on the left side of Fig. 1 (i and j) determine the ray passing from object X, through the reference image plane, to the reference camera with the focal length defined as a distance between the sensor on which the image is generated and the focus where the rays converge. The rays d_r and d_d can be then defined as

$$\bar{d}_r = P_r \cdot \bar{x}_r, \quad \bar{d}_d = P_d \cdot \bar{x}_d \tag{2}$$

and the mapping equation can be obtained:

$$C_r + P_r \cdot \bar{x}_r \cdot t_r = C_d + P_d \cdot \bar{x}_d \cdot t_d \tag{3}$$

where t_r and t_d are the scaling factors determining the length of the ray. Assuming that the rays having the same direction are equivalent regardless of the length of the three-space vector specifying this direction [2], the projective geometry instead of Euclidean can be used. In such a case the generalized disparity δ, related to the depth by a scaling factor (range), can be introduced and the simplified ray-to-ray mapping equation is received:

$$P_d \cdot \bar{x}_d = P_r \cdot \bar{x}_r + \delta(\bar{x}_r) \cdot (C_r - C_d) \tag{4}$$

After some transformations the following matrix form of the warping equation can be achieved:

$$\alpha \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \bar{a}_d & \bar{b}_d & \bar{c}_d \end{bmatrix}^{-1} \cdot \begin{bmatrix} \bar{a}_r & \bar{b}_r & \bar{c}_r & (C_r - C_d) \end{bmatrix} \cdot \begin{bmatrix} i \\ j \\ 1 \\ \delta(i,j) \end{bmatrix} \tag{5}$$

which is convenient for implementation as rational expressions

$$u = \frac{w_{11} \cdot i + w_{12} \cdot j + w_{13} + w_{14} \cdot \delta(i,j)}{w_{31} \cdot i + w_{32} \cdot j + w_{33} + w_{34} \cdot \delta(i,j)}$$
$$v = \frac{w_{21} \cdot i + w_{22} \cdot j + w_{23} + w_{24} \cdot \delta(i,j)}{w_{31} \cdot i + w_{32} \cdot j + w_{33} + w_{34} \cdot \delta(i,j)} \tag{6}$$

where w are the elements of the 4×3 warping matrix W, which may be precomputed for the specified parameters of the destination camera and is defined as:

$$W = \begin{bmatrix} \bar{a}_r(\bar{b}_d \times \bar{c}_d) & \bar{b}_r(\bar{b}_d \times \bar{c}_d) & \bar{c}_r(\bar{b}_d \times \bar{c}_d) & (C_r - C_d)(\bar{b}_d \times \bar{c}_d) \\ \bar{a}_r(\bar{c}_d \times \bar{a}_d) & \bar{b}_r(\bar{c}_d \times \bar{a}_d) & \bar{c}_r(\bar{c}_d \times \bar{a}_d) & (C_r - C_d)(\bar{c}_d \times \bar{a}_d) \\ \bar{a}_r(\bar{a}_d \times \bar{b}_d) & \bar{b}_r(\bar{a}_d \times \bar{b}_d) & \bar{c}_r(\bar{a}_d \times \bar{b}_d) & (C_r - C_d)(\bar{a}_d \times \bar{b}_d) \end{bmatrix} \tag{7}$$

Assuming the availability of the depth information, the obtained results depend on the resolution of the reference image(s) and geometrical configuration of objects on the scene concerning their visibility. Nevertheless, possible occlusion and exposure errors can be corrected using e.g. adaptive splatting. Some modified splatting techniques can also be used as described in our earlier paper [4].

3. The proposed approach and its verification

When analyzing the results obtained by using many face recognition algorithms tested with some standard benchmark datasets such as FERET, AT&T (ORL), Yale etc., the relatively strong influence of the face orientation on the recognition accuracy is observed. For a significant improvement of results many sophisticated face recognition methods can be used as well as the application of stereo matching [5]. Nevertheless, in this paper we concentrate on the Principal Component Analysis and the possible compensation of the face orientation using the McMillan's IBR approach. The proposed approach can be supplementary for the method of the face posture estimation based on IBR database [6], which in fact does not use the warping equation.

In order to verify the usefulness of the Image Based Rendering two side views for several faces were rendered using 3D Studio Max from face models acquired by 3D scanning by Cyberware. The models are full, 360 degrees scans of male and female heads stored in *.PLY format (known as Polygonal File Format or Stanford Triangle Format).

Assuming that the human face is acquired from a side or both sides together with the depth map, the influence of face orientation can be reduced by the fast synthesis of the front view using the Image Based Rendering. Various techniques e.g. active stereo or structured light [7] can be additionally used for the depth map acquisition. Nevertheless, such methods are not the topic of this paper.

Verification of the recognition accuracy for the Principal Component Analysis was performed using 300 colour en face images (only one image of each person) from FERET dataset together with several images rendered by 3D Studio Max (front views) as the train dataset. In the experiments the synthetic side view images and the front view image (reconstructed with use of the IBR algorithm implemented in MATLAB environment) were used for testing purposes. Simplifying the calculations, the face recognition was performed after conversion to grayscale. The number of features was limited to 150, 50, 10 and 5. Because of the presence of only single image for each face in the dataset, the nearest neighbour approach was used for the classification utilizing the Euclidean distance. The obtained results are presented as the recognition places - the first place means perfect recognition as the proper face is the closest one in the feature space, other places indicate the position of the proper face sorted by Euclidean distances.

The idea of using Euclidean distance is illustrated in Fig. 2 for an exemplary image with only two principal components (two eigenvalues), where the proper image belonging to the very same individual as in the test image is marked as the white square, while the black square indicates the test image recognized as the equivalent one.

The experiments were performed using several synthetic face models, which were included into the FERET dataset. The results obtained for two representative ones are shown in Figs. 5 and 6. When analysing the results, the advantages of the proposed approach over the use of side images of an individual can be noticed, especially for the presented exemplary female face. In that case the application of the IBR method leads to proper recognition even for only 5 eigenvalues.

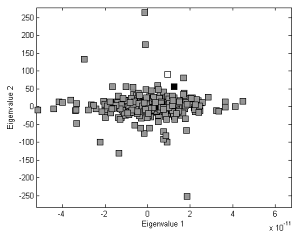

Fig. 2. Projection of the dataset on the space of two main eigenvalues
Rys. 2. Projekcja zbioru danych na przestrzeń dwóch głównych wartości własnych

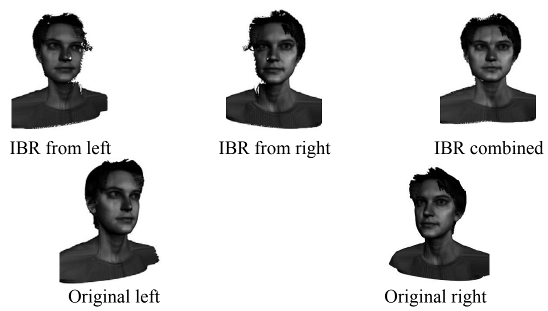

Fig. 3. Exemplary female face used in our experiments
Rys. 3. Przykładowa twarz kobieca użyta w eksperymentach

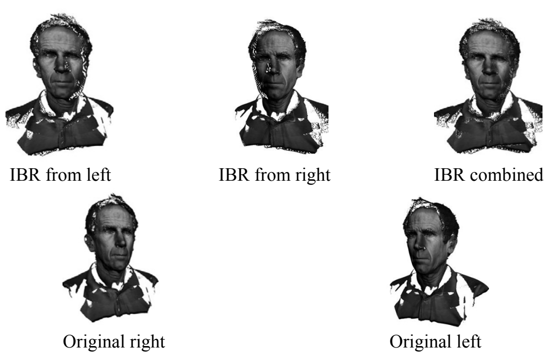

Fig. 4. Exemplary male face used in our experiments
Rys. 4. Przykładowa twarz męska użyta w eksperymentach

4. Conclusions

The images denoted as "combined" in Figs. 3 and 4 were rendered using both original side view images while the other two with the use of single image only. The differences are clearly visible in the quality sense (the use of splatting would be necessary) but the result of "combined IBR" is not necessarily the best recognized image. It is caused mainly by the imperfect matching and changing light conditions.

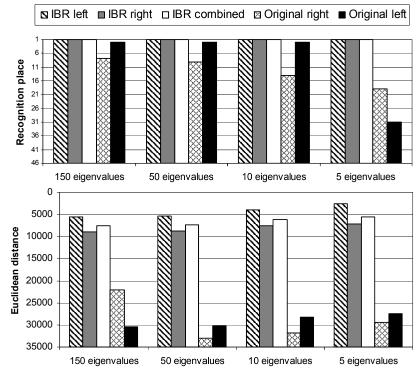

Fig. 5. Recognition results and Euclidean distances for exemplary female face
Rys. 5. Wyniki rozpoznawania i odległości euklidesowe dla przykładowej twarzy kobiecej

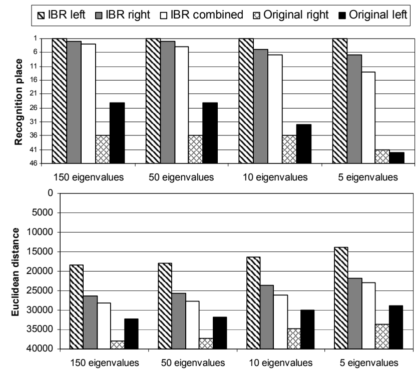

Fig. 6. Recognition results and Euclidean distances for exemplary male face
Rys. 6. Wyniki rozpoznawania i odległości euklidesowe dla przykładowej twarzy męskiej

Images rendered with warping equation, are generally better recognized than the side views of individuals but use of the simple PCA algorithm is insufficient for proper face recognition. Some hybrid methods that combine feature extraction, like EBGM, supported by feature reduction algorithms like PCA or LDA would lead to better results so their support by the IBR method can be an interesting direction of further research.

5. References

[1] Nowosielski A.: Face recognition using DCT and LDA. Advances in Soft Computing, Computer Recognition Systems. pp. 799 – 806, Springer-Verlag, Berlin Heidelberg, 2005

[2] McMillan L.: An image based approach to three-dimensional computer graphics, Ph.D dissertation, Chapel Hill, 1997.

[3] Shade J., Gortler S., He L., Szeliski R.: Layered depth images. Proc. International Conference SIGGRAPH'98, pp. 231 – 242, 1998.

[4] Okarma K., Miętus A., Tecław M.: Vector median splatting for image based rendering. Lecture Notes in Computer Science vol. 6375, pp. 150 – 157, Springer-Verlag, Berlin Heidelberg, 2010.

[5] Castillo C.D., Jacobs D.W.: Using stereo matching with general epipolar geometry for 2D face recognition across pose. IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 31 issue 12, pp. 2298 – 2304, 2009.

[6] Sengupta K., Lee P., Ohya J.: Face posture estimation using eigen analysis on an IBR (image based rendered) database. Pattern Recognition vol. 35 no. 1, pp. 103 – 117, 2002.

[7] Ribera R.B., Kim T., Kim J., Hur N.: Dense depth map acquisition system for 3DTV applications based on active stereo and structured light integration. Lecture Notes in Computer Science vol. 5879, pp. 499 – 510, Springer-Verlag, Berlin Heidelberg, 2009.

otrzymano / received: 21.09.2010
przyjęto do druku / accepted: 01.11.2010 *artykuł recenzowany*