RAPID STEREO-VISION ENHANCED FACE RECOGNITION

Sergey Kosov, Thorsten Thormählen, Hans-Peter Seidel

MPI Informatik, Saarbrücken, Germany

ABSTRACT

This paper presents a real-time face recognition system. The system uses a stereo camera to locate, track, and recognize a person's face. Our algorithm improves state-of-the-art monocular 2D object recognition techniques by additionally considering the facial 3D surface, which is relatively stable under different lighting conditions. First, faces are detected and their surfaces are reconstructed from the stereo images. Afterwards, a 3D face is composed by joining 2D image data and appropriate depth data. The 3D face is then decomposed into its principal components. The principal components are used to recognize a 3D face by comparing characteristics of the current face to those of known individuals in a database. The result is an efficient and accurate face recognition algorithm. To evaluate our approach, we compared its performance to a classical monocular face recognition algorithm and observed that the recognition rate increased on average by 7.7 percent.

Index Terms— Image analysis, object recognition, stereo vision

1. INTRODUCTION

Automatic face recognition has wide areas of application, including human-machine interaction, personalization of devises, data encryption, security, virtual reality, computer games, surveillance systems, or electronic commerce.

First face recognition algorithms appeared in the 60ies and used geometric relations between detected facial features in 2D images to identify a person [1]. At the beginning of the 90ies, a new approach was proposed by Turk and Pentland - the eigenface approach [2]. This approach is based on the principal components analysis (PCA), which was later refined by Belhumeur et al. [3] and Frey et al. [4].

Principal component analysis transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. By expressing the data in such a way, their similarities and differences can be observed.

Monocular 2D face recognition methods evaluate 2D image regions, which are not invariant under different lighting conditions, facial expressions, or make-up. Thus, they suffer from limited input data and insufficient information. In 2003, Tsalakanidou et al. [5] showed that depth information can be applied to improved face recognition. Recently, a number of face recognition methods using 3D information were proposed [6, 7]. Bronstein et al. [8] presented a recognition framework based on 3D geometric invariants of the human face. However, none of these 3D face recognition approaches is capable to track and recognize multiple persons in real-time. Wang et al. [9] described a real-time algorithm based on fisherfaces. However, their approach uses a morphological filter in combination with some heuristics to detect the closest face to the camera. Therefore, only a single face can be detected and analyzed.

Our PCA-based approach uses 3D depth information to significantly improve the rate of correct recognitions. Furthermore, it allows to track and recognize multiple persons in real-time. The depth information from the stereo images is estimated using a highly accurate and fast variational disparity solver. The 2D image intensity and depth information at each face point are transformed into PCA space. Thereby, two separate PCA transformations are calculated; one for the image intensities and one for the depth information. An incoming 3D face is transformed to PCA space and the Mahalanobis distance to each trained face is calculated. If the distance is small enough, the identifier of the person whose face has the smallest distance to the current one is returned, otherwise, if the distance is too large, the system classifies a person as unknown. If requested, this unknown person can be added to the PCA database. In our experiments we used a PCA database of 34 persons.

2. OVERVIEW

Figure 1 gives an overview of our real-time face recognition system. The input consists of a sequence of stereo image pairs and the output is a list of persons that are recognized in the current image as well as their positions in the 2D images and 3D scene. The system consists of three major building blocks: disparity estimation, face detection, and face recognition. Our approach for disparity estimation and face detection is summarized in the next section. Section 4 gives an introduction to the state-of-the-art PCA approach for face recognition and afterwards describes how we extended this approach by additionally taking the depth information into account. In Section 5 the approach is evaluated and the paper ends with a conclusion.

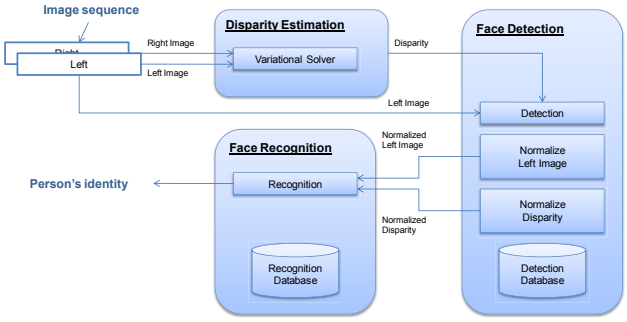

Fig. 1. System overview of our stereo-enhanced face recognizer.

3. DISPARITY ESTIMATION AND FACE DETECTION

To acquire stereo image pairs, we use two cameras with an image resolution of 640×480 pixels. The two cameras are mounted next to each other with a baseline distance of 20 cm and an angle of convergence of 9.5 degrees. After off-line calibration with a calibration pattern, we rectify the input images to standard stereo geometry and estimate a disparity map. Let the left picture be denoted by $I_l(x, y)$ and the right picture by $I_r(x, y)$. We then minimize the functional: $\iint_\Omega \left(I_l(x, y) - I_r(x - d(x, y), y)\right)^2 \, dS \; + \; \Psi((\nabla d(x, y))^2)$, where $d(x, y)$ is the disparity at pixel $(x, y)^\top$ and Ψ is a smoothness term. This term encourages neighboring pixels to have similar disparities. To find the minimum, we use a real-time variational approach [10].

The faces are detected with a real-time stereo-enhanced face detection algorithm [11]. For each detected face we extract the face region and the corresponding disparity map. Each region is scaled to a *normal* resolution of $N \times N$ pixels with $N = 100$ in our experiments. Afterwards, we perform a normalization of the image intensities and disparity values over the region. Thus, for each face region we extract two matrices with $N \times N$ elements. The first matrix contains the normalized image intensities, and the second matrix stores the normalized disparity values. Separately for each matrix, we reassemble the rows of the matrix into a $N^2 \times 1$ vector. Consequently, we obtain two $N^2 \times 1$ vectors, the intensity vector $\mathbf{x}$ and the disparity vector $\mathbf{d}$ for each detected face.

4. FACE RECOGNITION

For the face recognition task we want to construct a computational face model that contains the most relevant information about a face. The principal component analysis [12] allows us to analyze the variation between different faces and creates a face subspace, defined by an orthogonal basis of vectors. These vectors are the eigenvectors of the covariance matrix of the distribution, spanned by the training set of faces. This means that later a new face can be projected into the face subspace, i.e., represented by a linear combination of these eigenvectors. The recognition process is performed by representing a new face as a point in the face subspace and then computing the distances between this point and other points, which correspond to the faces of known individuals in the training set.

4.1. Training Database

In this work, we used a training set of $K = 34$ normalized faces with corresponding normalized disparity maps. Each image has a resolution of $N \times N$ pixels . Figure 2 shows the training database. Each intensity image k is represented by a $N^2 \times 1$ vector $\tilde{\mathbf{x}}_k$ and a $N^2 \times 1$ vector $\tilde{\mathbf{d}}_k$ representing the corresponding disparity map.

Fig. 2. Training database of intensity images and disparity maps. The face located in last row and rightmost column is the mean face, i.e., the average of the whole data set.

4.2. Principal Component Analysis

A principal component analysis is performed on the training set. In the following, this will be described in more detail for the intensity vectors $\tilde{\mathbf{x}}_k$, but the same procedure is also applied to the disparity vectors $\tilde{\mathbf{d}}_k$ in parallel.

First, we compute the average face $\bar{\mathbf{x}} = \frac{1}{K} \sum_k \tilde{\mathbf{x}}_k$. The average face $\bar{\mathbf{x}}$ and the corresponding disparity map $\bar{\mathbf{d}}$ are shown in Fig. 2. Then, the $N^2 \times N^2$ covariance matrix $\mathbf{C}$ describing the variation in our dataset is given by

$$\mathbf{C} = \frac{1}{K} \sum_{k=1}^{K} (\tilde{\mathbf{x}}_k - \bar{\mathbf{x}})(\tilde{\mathbf{x}}_k - \bar{\mathbf{x}})^\top \quad . \tag{1}$$

The covariance matrix $\mathbf{C}$ is a symmetric matrix and, thus, has a spectral decomposition of the form $\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, where $\mathbf{U}$ is an orthonormal matrix and $\mathbf{\Lambda}$ is a diagonal matrix (see [13]

for efficient algorithms that compute such a decomposition). The columns of $\mathbf{U}$ are the eigenvectors of $\mathbf{C}$ and the diagonal elements of $\mathbf{\Lambda}$ are the eigenvalues. The eigenvectors in $\mathbf{U}$ are sorted according to their corresponding eigenvalues starting with the largest eigenvalue. The eigenvectors are also called the principal components. The eigenvector with the highest eigenvalue is the principal component of the data set, i.e., it describes the most significant relationship between the data dimensions. The first three principal components for our data set are shown in Fig 3. Eigenvectors with small singular values contain less or no information about the variance in our dataset. Therefore, they can be dropped without losing information. The resulting face space spanned by the remaining eigenvectors will have fewer dimensions than the original one. As our covariance matrix $\mathbf{C}$ is generated only from $K = 34$ examples, a maximum of K singular values are not close to zero. Therefore, we keep only the first $\tilde{N}$ eigenvectors with $\tilde{N} < K$ and $\tilde{N} < N^2$. We used $\tilde{N} = 33$ in our experiments. The resulting matrix $\tilde{\mathbf{U}}$ becomes a $N^2 \times \tilde{N}$ matrix. PCA transformation of a $N^2 \times 1$ input vector $\mathbf{x}$ to a $\tilde{N} \times 1$ vector $\mathbf{y}$ in face space can now be performed with $\mathbf{y} = \tilde{\mathbf{U}}^\top(\mathbf{x} - \bar{\mathbf{x}})$. We can also transform our K training vectors $\tilde{\mathbf{x}}_k$ into PCA space with

$$\tilde{\mathbf{y}}_k = \tilde{\mathbf{U}}^\top(\tilde{\mathbf{x}}_k - \bar{\mathbf{x}}) \quad \forall\, k = 1 \ldots K \tag{2}$$

To evaluate if an input face, given by an input vector $\mathbf{x}$, is likely to be among the faces in our database, we just have to calculate $\mathbf{y}$ and compare it with all transformed vectors $\tilde{\mathbf{y}}_k$ in our database. We then return the identity of the face with the smallest Mahalanobis distance $(\mathbf{y} - \tilde{\mathbf{y}}_k)^\top \tilde{\mathbf{\Lambda}}_y^{-1}(\mathbf{y} - \tilde{\mathbf{y}}_k)$ as the most likely match. Thereby the $\tilde{N} \times \tilde{N}$ matrix $\tilde{\mathbf{\Lambda}}_y$ is the reduced version of matrix $\mathbf{\Lambda}$ that contains only the $\tilde{N}$ largest eigenvalues. If all the distances are larger than a given threshold τ, it its likely that the face is not in the database, and it is marked as *unknown*. If requested, this unknown face can be added to the database.

This works already reasonably well, if only the intensity vectors are considered. However, even better results can be obtained, if we additionally employ the disparity information. Analogously to Eq. (2), the disparity vectors $\tilde{\mathbf{d}}_k$ of the training set are projected into their PCA space, given by $\tilde{\mathbf{V}}$ (calculated in the exact same manner as $\tilde{\mathbf{U}}$ before):

$$\tilde{\mathbf{e}}_k = \tilde{\mathbf{V}}^\top(\tilde{\mathbf{d}}_k - \bar{\mathbf{d}}) \quad \forall\, k = 1 \ldots K \quad . \tag{3}$$

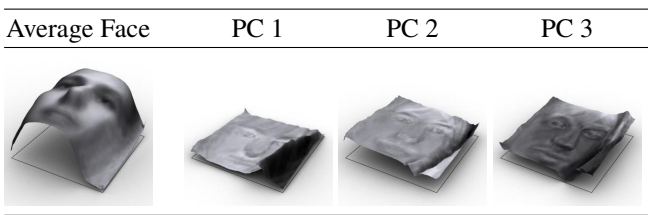

Fig. 3. Results of the PCA on the training dataset: The average face and the first three principal components.

To evaluate an input face, we now calculate the weighted Mahalanobis distances

$$\alpha\,(\mathbf{y} - \tilde{\mathbf{y}}_k)^\top \tilde{\mathbf{\Lambda}}_y^{-1}(\mathbf{y} - \tilde{\mathbf{y}}_k) + (1-\alpha)\,(\mathbf{e} - \tilde{\mathbf{e}}_k)^\top \tilde{\mathbf{\Lambda}}_e^{-1}(\mathbf{e} - \tilde{\mathbf{e}}_k)$$

for all $k = 1 \ldots K$ and return either the identity of the face with the smallest distance or *unknown*, if all distance are larger than τ. The weighting factor is chosen to be $\alpha = 0.75$ in our experiments.

5. EXPERIMENTS AND RESULTS

In order to evaluate our stereo-enhanced face recognizer we used leave-one-out cross validation. We randomly removed one face from our database and rebuild the PCA space with the remaining 33 faces. We then presented the algorithm the removed face as well as one other randomly selected face that is still in the database. In total we performed 2000 random trials. Thereby, the intensity image and the disparity maps were perturbed by Gaussian noise with variance σ^2. If the algorithm labelled the removed face as *unknown* and recognized the other random face with the correct identity, this trail was counted as a *correct* recognition; and as a *false* recognition otherwise. We repeated this experiment three times: 1) for the classical monocular PCA-based face recognition algorithm, 2) for a PCA approach that uses only the disparity information, and 3) for our stereo-enhanced face recognizer that combines intensity and disparity information. Furthermore, we repeated the experiment for different noise levels σ. Fig. 4 compares the recognition rates for the three approaches. On average, over all noise levels, the stereo enhanced method improves the recognition rate by 7.7 percent compared to the classical monocular approach.

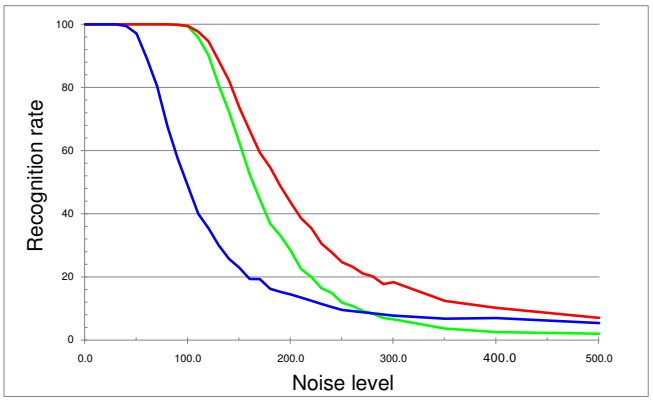

Fig. 4. Recognition rates (in percent) over noise level σ for the monocular recognition approach (green), an approach that only uses the disparity information (blue), and our stereo enhanced method (red).

Another important feature of our stereo enhanced method is that it can not be tricked by a photo print of a person. An

Fig. 5. Eliminating false-positives (marked by a red cross) with the stereo-enhanced face recognizer.

example is given in Fig. 5. In Tab. 1 timings for the different steps of our algorithms are given. Furthermore, the supplemental video[1] shows a live demo of our stereo-vision enhanced face recognition system. Multiple persons can be recognized in real-time.

Step	Mono [msec]	Stereo [msec]
Estimate disparity map	-	91
Face detection	86	152
Face recognition	1	2
Total	87	245

Table 1. Timings for an Intel® Core™ 2 Quad CPU Q9550 with 2,83 GHz.

6. CONCLUSION AND DISCUSSION

In this paper a PCA-based face recognition algorithm is extended by additionally considering the disparity map generated by a stereo camera. The system runs in real-time and increases the number of correctly recognised faces by 7.7 percent on average. This result is coherent with the findings of other researchers who have reported similar improvements (e.g. [7]). However, in our case a real-time algorithm was employed where the disparity estimation errors are slightly larger compared to the best off-line methods (cp. [10]). Thus, this paper has shown that even in a real-time framework it is beneficial to use additional information provided by a stereo camera system. An important feature of our system in practise is that it can differentiate between real faces and photos of faces (because the photo has a flat surface). This makes it more difficult to fake an identity and is especially important, if the application of face recognition software is data encryption, security, or electronic commerce.

7. REFERENCES

[1] W. W. Bledsoe, "The model method in facial recognition," Tech. Rep. 15, Panoramic Research Inc., Palo Alto, CA, 1964.

[1]The video can be downloaded at:
http://www.mpi-inf.mpg.de/homepage/skosov/icip2010xvid.avi

[2] M.A. Turk and A.P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Computer Vision and Pattern Recognition*, 1991, pp. 586–591.

[3] Peter N. Belhumeur, Joe;o P. Hespanha, and David J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 711–720, 1997.

[4] Brendan J. Frey, Antonio Colmenarez, and Thomas S. Huang, "Mixtures of local linear subspaces for face recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1998, pp. 32–37.

[5] F. Tsalakanidou, D. Tzovaras, and M. G. Strintzis, "Use of depth and colour eigenfaces for face recognition," *Pattern Recogn. Lett.*, vol. 24, no. 9-10, pp. 1427–1435, 2003.

[6] Akihiro Hayasaka, Koichi Ito, Takafumi Aoki, Hiroshi Nakajima, and Koji Kobayashi, "A robust 3d face recognition algorithm using passive stereo vision," *IEICE Transactions*, vol. 92-A, no. 4, pp. 1047–1055, 2009.

[7] T.-H. Sun, M. Chen, S. Lo, and F.-C. Tien, "Face recognition using 2D and disparity eigenface," *Expert Syst. Appl.*, vol. 33, no. 2, pp. 265–273, 2007.

[8] Alexander M. Bronstein, Michael M. Bronstein, and Ron Kimmel, "Three-dimensional face recognition," *Int. J. Comput. Vision*, vol. 64, no. 1, pp. 5–30, 2005.

[9] J.-G. Wang, E.T. Lim, X. Chen, and R. Venkateswarlu, "Real-time stereo face recognition by fusing appearance and depth fisherfaces," *J. VLSI Signal Process. Syst.*, vol. 49, no. 3, pp. 409–423, 2007.

[10] Sergey Kosov, Thorsten Thormählen, and Hans-Peter Seidel, "Accurate real-time disparity estimation with variational methods," in *Proc. International Symposium on Visual Computing*, 2009, pp. 796–807.

[11] Sergey Kosov, Kristina Scherbaum, Kamil Faber, Thorsten Thormählen, and Hans-Peter Seidel, "Rapid stereo-vision enhanced face detection," in *Proc. IEEE International Conference on Image Processing*, 2009, pp. 1221–1224.

[12] Howard Anton, *Elementary Linear Algebra*, John Wiley and Sons Inc., 1998.

[13] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery, *Numerical Recipes in C++: The Art of Scientific Computing*, Cambridge University Press, 2002.