의상 특징 기반의 동일인 식별

최유주 박선미 조위덕 김구진[O]

한독미디어대학원대학교 뉴미디어학부

경북대학교 전자전기컴퓨터학부

아주대학교 전자공학부

[O](교신저자) 경북대학교 컴퓨터공학과

yjchoi@kgit.ac.kr disvogue@knu.ac.kr chowd@ajou.ac.kr kujinkim@yahoo.com

Person Identification based on Clothing Feature

Yoo-Joo Choi Sun-Mi Park We-Duke Cho Ku-Jin Kim[O]

Dept. of Newmedia, Korean German Institute of Technology

School of EECS, Kyungpook National University

Division of Electrical & Computer Engineering, Ajou University

[O](corresponding author) Dept. of Computer Engineering, Kyungpook National University

요 약

비전 기반의 감시 시스템에서 동일인의 식별은 매우 중요하다. 감시 시스템에서 주로 사용되는 CCTV 카메라의 영상은 상대적으로 낮은 해상도를 가지므로 얼굴 인식 기법을 이용하여 동일인을 식별하기는 어렵다. 본 논문에서는 CCTV 카메라 영상에서 의상 특징을 이용하여 동일인을 식별하는 알고리즘을 제안한다. 건물의 주출입구에서 출입자가 인증을 받을 때, 의상 특징이 데이터베이스에 저장된다. 그 후, 건물 내에서 촬영한 영상에 대해 배경 차감 및 피부색 발견 기법을 이용하여 의상 영역을 발견한다. 의상의 특징 벡터는 텍스처와 색상 특징을 이용하여 구성한다. 텍스처 특징은 지역적 에지 히스토그램을 이용하여 추출된다. 색상 특징은 색상 지도의 옥트리 기반 양자화(octree-based quantization)를 이용하여 추출된다. 건물 내의 촬영 영상이 주어질 때, 데이터베이스에서 의상 특징이 가장 유사한 사람을 발견함으로써 동일인을 식별하며, 의상 특징 벡터 간의 유사도 측정을 위해서는 유클리디안 거리(Euclidean distance)를 사용한다. 실험 결과, 얼굴인식 기법이 최대 43%의 성공률을 보인 데 비해, 의상 특징을 이용하여 80%의 성공률로 동일인을 식별하였다.

Abstract

With the widespread use of vision-based surveillance systems, the capability for person identification is now an essential component. However, the CCTV cameras used in surveillance systems tend to produce relatively low-resolution images, making it difficult to use face recognition techniques for person identification. Therefore, an algorithm is proposed for person identification in CCTV camera images based on the clothing. Whenever a person is authenticated at the main entrance of a building, the clothing feature of that person is extracted and added to the database. Using a given image, the clothing area is detected using background subtraction and skin color detection techniques. The clothing feature vector is then composed of textural and color features of the clothing region, where the textural feature is extracted based on a local edge histogram, while the color feature is extracted using octree-based quantization of a color map. When given a query image, the person can then be identified by finding the most similar clothing feature from the database, where the Euclidean distance is used as the similarity measure. Experimental results show an 80% success rate for person identification with the proposed algorithm, and only a 43% success rate when using face recognition.

키워드: 동일인 식별, 의상 특징, 색상, 텍스처, 영상 분석, 감시 시스템

Keywords: person identification, clothing feature, color, texture, image analysis, surveillance system

*투고일: 2009.08.06 / 심사완료일: 1차 2009.11.30, 2차 2010.01.20 / 게재확정일: 2010.02.09

1. Introduction

In the field of content-based image retrieval, color and textural information are widely used for characterizing an image. Color information is an important cue for characterizing an entire image or the objects contained in the image, yet this information is invariably unstable due to a dependence on the illumination. Meanwhile, textural features can be used to represent the visual patterns in an image.

Accordingly, a person identification method is presented based on the color and texture information of clothing. The proposed algorithm can be applied to vision-based surveillance systems, and used to authenticate arbitrary individuals. In a ubiquitous environment with the widespread use of intelligent surveillance systems, a function is needed that can automatically detect unauthenticated persons. Despite the development of many face recognition techniques [15] for this purpose, such techniques require high resolution images that clearly represent the characteristics of faces. However, most surveillance systems use CCTV cameras that are located at quite a distance from the individual subjects, meaning the resulting images have a low resolution, which makes it difficult to apply face recognition techniques. While several studies [1,5,12] have already explored person identification based on clothing features, the resulting methods either require an extensive amount of training data [1] or the recognition results are highly sensitive to an accurate segmentation of the region of the person in the image [5,12].

The proposed method uses background subtraction and skin color detection methods to detect the clothing area in an image. The textural feature of the clothing area is then computed by constructing a local edge histogram, while the color feature is extracted using octree-based quantization and a similar color integration technique. By measuring the similarity between the clothing feature vectors for query and target images, candidates with similar clothing to that in the query image are identified. In addition to online person identification in real time, the proposed algorithm can also be used for person image retrieval and indexing.

The remainder of this paper is organized as follows. Section 2 discusses related work, then Section 3 outlines the problem of person identification and presents a flowchart of the proposed algorithm. Section 4 describes the method used to detect a clothing area in an image, the construction of textural and color feature vectors for the clothing area are discussed in Sections 5 and 6, respectively, and the similarity measurement is covered in Section 7. Section 8 provides experimental results, and some final conclusions are given in Section 9.

2. Related Work

Image retrieval methods based on content features have already been actively researched for off-line image retrieval using a multimedia database, where these content-based image retrieval methods use visual information, such as color, texture, and shape, to retrieve similar images to a given query [9].

For example, Kim et al. [8] proposed an edge correlogram method for texture-based image retrieval. While correlograms are normally used to analyze the change in the spatial correlation of pairs of colors according to distance, Kim et al. use an edge correlogram to analyze object shapes. However, texture information by itself is not robust enough to find similar images, and even though the computation cost is lower, analyzing the spatial correlation of an image in real-time is still difficult due to the heavy computational cost of a correlogram.

Color-based image features have also been defined by applying various color analysis algorithms. As a simple and effective approach, histograms are generally used to represent the color information of an image, and the sum of L1 or L2 distances between matching bins in histograms used as the similarity between images [13]. Histogram intersection has been also widely used as an effective method to compare color features [11]. However, histogram-based approaches are unable to represent the spatial information of colors. Thus, CCV(Color Coherence Vector)-based methods [6] have been proposed as a way of representing the spatial information of colors, yet the real-time application of these approaches is difficult due to a high computation complexity.

More recently, content-based image retrieval technologies have been applied to intelligent surveillance systems for person identification or tracking a specific person from a video sequence [10]. In [10], the body part colors of the target models in a video are analyzed and compared with the person information in each video frame to detect the models in each frame. Yet, this approach assumes that the whole body of each person is captured.

Face recognition has been also actively investigated for person identification from a video. For example, Zhou et. al [16] proposed a time series state space model for robust face recognition in a sequence of video frames. However, the face image resolution has to be high enough for robust recognition.

Finally, the use of clothing for person identification has received some recent attention [1,5,7,12]. Kim et al. [7] suggested a method based on the textural features of clothing. While their method assumes the use of grayscale input images, an even better performance can be expected with the use of color images and the color features of the clothing. Anguelov et al. [1] suggested a probabilistic method for recognizing individuals in photo albums. Using facial and clothing information, this method combines several contextual cues, such as the lighting conditions and the time when the photo was taken, to construct a Markov Random Field. As a result, this method can be applied to images with a complicated background, and the performance is not highly sensitive to an accurate segmentation of the region of interest, although a huge amount of training data is required for successful recognition. Meanwhile, Gorras et al. [5] and Tous et al. [12] developed two similar methods that use the structural features of the clothing area. The structural feature is derived by interpreting the regions inside the clothing area based on the color and texture. However, the performance of these methods is highly dependent on accurate segmentation of the clothing area from the background and the segmentation of the regions inside the clothing area. Thus, in the case of an input image with a complicated background, high reflection, or shadows, it is difficult to segment the foreground accurately. Plus, clothing with complicated patterns also makes the accurate segmentation of the inside region very difficult, resulting in severe distortion of the structural feature.

3. Overview of Proposed Algorithm

A flowchart of the proposed algorithm is shown in Figure 1. Consisting of two parts, the algorithm first constructs a database of authenticated people. Whenever an individual is authenticated at the building entrance, several images of that person are captured and feature vectors computed for the clothing areas. For each clothing area, a tuple is constructed that is composed of three components: the identification number of the person, plus texture and color feature vectors for their clothing. As such, for an authenticated person, several tuples that share the same identification number are added to the database.

The second part of the algorithm involves identifying the person in a given image. Inside a building, images of people are captured wherever a CCTV camera is installed, and the algorithm then identifies the person in the image by extracting the clothing feature and comparing it to those already in the database.

Thus, from a given image, the clothing area A is detected first and the textural feature V_T then extracted from the clothing area A. Next, the database is searched to find tuples with texture features within a distance of θ_T from V_T. From among these tuples, five are then selected with the closest color feature to that of A and within a distance of within θ_C. The values of θ_T and θ_C are thresholds that are tuned through experiments. The person that appears most frequently among the final five tuples is then the output.

The following restrictions on the input images are also assumed:

1) Between two captured images of a person, it is assumed that the subject has not changed their clothes.
2) Each input image contains a face and the upper half of the body.
3) In an image, different items of clothing have different visual characteristics as regards color or style.

4. Detection of Clothing Area

A clothing area is detected using the position of the face area as follows:

1. Extract the foreground and skin pixels.
2. Define skin blobs using a grid image [2].
3. Detect faces from skin blobs using the AdaBoost algorithm [3].
4. Detect clothing areas by finding the area that is apart from the face rectangle at a fixed distance in the lower direction. The size of the clothing area is decided by doubling the size of the face area.

The cameras are classified into two categories based on the height of the camera location, where the Type A cameras are low and capture the entrance images, while the Type B cameras are high on the ceiling and capture the indoor images. The viewpoint of the Type A cameras is also almost in front of the person, whereas the viewpoint of the Type B cameras is higher than the height of an average male adult. Thus, in the case of an image from a Type A camera, the distance between the face and the clothing area is 1/4 the height of the face blob, while for an image from a Type B camera, the clothing area is stuck to the face area. Fig. 2 shows front and ceiling camera images with the detected clothing areas (blue rectangle) and face areas (red rectangle).

5. Textural Feature Vector

Edge information is used to construct the feature vectors of the clothing area. The Sobel edge equation is applied to each pixel in the clothing area to produce an edge image, which is then vertically divided into 3 sections. For each section, a histogram is computed by accumulating the edge values in a vertical direction: $V_1 = (v_{11}, v_{12}, \cdots v_{1m})$, $V_2 = (v_{21}, v_{22}, \cdots v_{2m})$, and $V_3 = (v_{31}, v_{32}, \cdots v_{3m})$.

Meanwhile, for the whole edge image, a histogram is computed that represents the accumulation of the edge values in a horizontal direction: $H = (h_1, h_2, \cdots, h_n)$. The textural feature vector is then constructed as follows:

$$X = (v_{11}, \cdots, v_{1m}, v_{21}, \cdots, v_{2m}, v_{31}, \cdots, v_{3m}, h_1, h_2, \cdots, h_n).$$

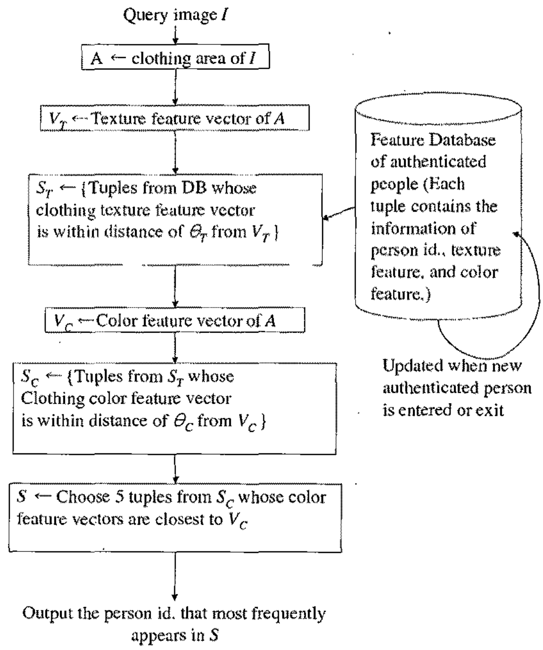

Fig. 1 Flowchart for person identification algorithm

Fig. 2 Automatically detected clothing areas (blue rectangles) in Type A (first row) and Type B (second row) camera images.

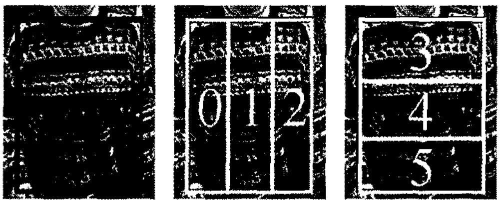

Fig. 3 Sub-regions of clothing area for color analysis.

6. Color Feature Vector

To define the color feature vector, quantized colors are used for an effective color analysis. In the proposed approach, up to 10 major colors are extracted for each clothing area by applying octree-based color quantization [4], which can deal with any image, regardless of the number of colors in the image, and can produce reasonable results with little degradation when converting from a high number of colors to a restricted number of colors.

To include spatial information on the colors, the clothing area is divided into six sub-regions, as shown in Fig. 3, and a histogram constructed for each sub-region based on the extracted quantized colors.

Let m be the number of major colors extracted from a clothing region when applying octree-based color quantization. The quantized color palette P consists of m extracted quantized RGB colors, $(RGB_1, RGB_2, \cdots, RGB_m)$. A color histogram C_i for each clothing sub-region with respect to m colors of P is defined as follows.

$$C_0 = (c_{01}, c_{02}, \cdots, c_{0m}), \qquad C_1 = (c_{11}, c_{12}, \cdots, c_{1m}),$$
$$C_2 = (c_{21}, c_{22}, \cdots, c_{2m}), \qquad C_3 = (c_{31}, c_{32}, \cdots, c_{4m}),$$
$$C_4 = (c_{41}, c_{42}, \cdots, c_{4m}), \qquad C_5 = (c_{51}, c_{52}, \cdots, c_{5m}),$$

where c_{ij} is the frequency of bin j in the histogram for sub-region i.

The quantized color palette $P = (RGB_1, RGB_2, \cdots, RGB_m)$ and color histogram $C = (C_0, C_1, C_2, C_3, C_4, C_5)$ are then used as the color feature vector of the clothing region.

7. Measuring Similarity between Feature Vectors

For the texture feature vectors, the Euclidean distance between two texture feature vectors, as defined in Section 5, is used to determine their similarity. Whereas, for the color similarity, the histogram bins are matched between two images. In the case of histograms based on a fixed color palette, they can be compared by computing the difference between the bins with the same index. However, with the proposed approach, since bins with the same index can represent different colors in different images, the bins also need to be matched between histograms of two images.

This bin matching between two histograms is achieved by finding the shortest Euclidean distance between the bin colors in the two histograms. Thus, multiple bins in the histogram for image 1 can be matched to one bin in the histogram for image 2, and vice versa. These multiple bins are then merged to match one bin, allowing the frequencies of multiple bins to be aggregated into one bin. Thereafter, the color similarity $C(I, J)$ of the two clothing areas is computed as follows:

$$C(I,J) = \sum_{i=0}^{m} D(C_{I,i}, C_{J,M(i)}) * \frac{MAX(H_{I,j}, H_{J,M(i)})}{MIN(H_{I,j}, H_{J,M(i)})},$$

where m means the number of color bins in histogram I, $D(C_{I,i}, C_{J,M(i)})$ represents the Euclidean distance between the color of bin i in histogram I and the color of the matched bin in histogram J, and $H_{I,i}$ and $H_{I,M(i)}$ represent the normalized frequency of bin i in histogram I and the normalized frequency of the matched bin in histogram J, respectively. Fig. 4 shows an example of matched bins, where bin 2 and bin 6 in histogram 1 are merged.

Finally, the clothing similarity $S(I, J)$ based on the texture and color feature vectors is defined by the weighted sum of the texture similarity $X(I, J)$ and color similarity $C(I, J)$.

$$S(I,J) = X(I,J) / MAX_X + C(I,J) / MAX_C,$$

where MAX_X and MAX_C are weighted factors to normalize the texture and color similarities, respectively.

8. Experimental Results

In experiments, the proposed approach based on clothing similarity was applied to distinguish authenticated persons from non-authenticated persons.

Five pictures of 42 different items of clothing worn by 13 different people were captured using two CCTV cameras A and B. Camera A was located at the main gate of the building and took pictures of a person when they were authenticated based on an identification card or fingerprints. Five images were then selected from the sequence of images captured by the main gate CCTV camera (Type A). Meanwhile, another five images were selected from the images captured by a CCTV camera (Type B) located in an indoor corridor.

Let the main gate and indoor images be denoted as $A_s(i)$ and $B_s(i)$, respectively, where i is the serial number of the clothing included in the image and s is the serial number of the image(1-5). Given one image of clothing i captured by the indoor camera as the query image, five identical clothing images were retrieved from $A_s(j)$, where $1 \le s \le 5$, $1 \le j \le 42$. If most of the 5 retrieved images included the same clothing, this was counted as a success for identifying an authenticated person. When using the proposed texture and color features, the rate of success in identifying authenticated and non-authenticated people was 80.0% and 68.57%, respectively.

Fig. 5 shows a detected person with an authenticated pass captured by the camera at the main gate, while Fig. 6 shows an automatically detected authenticated and non-authenticated person.

To compare these results with the performance of face recognition, a 2D PCA (Principal Component Analysis) based on a face recognition technique [14] was applied to the face data set derived from the clothing images. As the clothing images were taken using thirteen people, thirteen different faces were used for the test. The composition of the training and test data set is shown in Fig. 7. The images in each column in Fig. 7 were taken of one person. The data set is explained in Table 1 and the recognition success rate is shown in Table 2. The Type A camera images were used as the training set, while the Type B camera images were used as the test set.

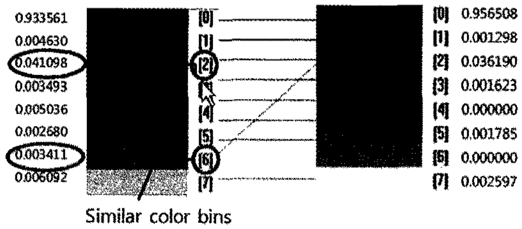

Fig. 4 Matching and similar color bins between histograms.

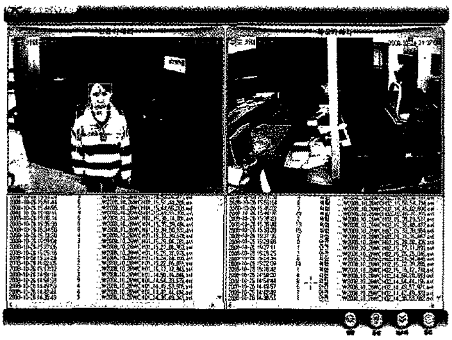

Fig. 5 Authenticated person captured by camera at main gate.

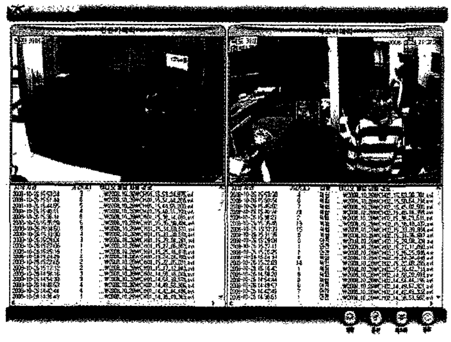

Fig. 6 Automatically detected authenticated and non-authenticated persons from images captured by camera B.

Table 1 Data used in face recognition.

Data set	Number of persons	Images per person	Total number of images
Training set	13	5	65
Test set	13	5	65

Table 2 Success rate of face recognition (number in parentheses represents number of eigenvectors used in test)

Dimension of feature vector	40*(3)	40*(6)	40*(10)	40*(20)
Success rate (%)	41.53	43.07	35.38	33.84

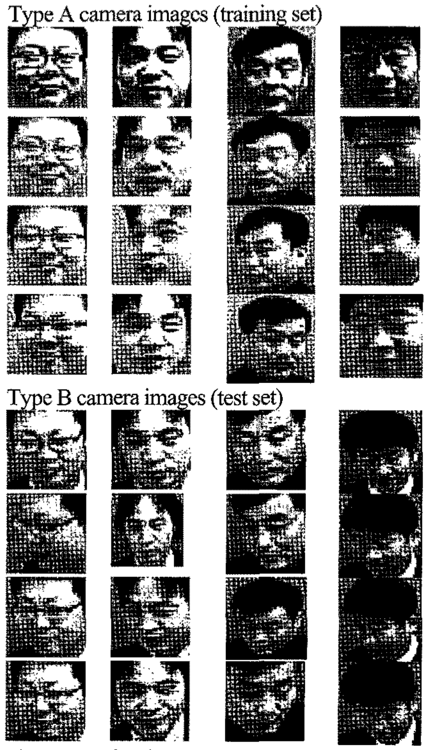

Fig. 7 Face data images

9. Conclusion

This paper presented a person identification method based on clothing. Textural feature vectors of the clothing area are constructed based on local edge histograms, which are efficient in representing spatial information, while color features are extracted based on octree-quantization and the integration of similar colors.

The similarity between two feature vectors is measured by the Euclidean distance, and the final success rate of person identification shown to be up to 80%.

Acknowledgements

This research was supported by the Foundation of Ubiquitous Computing and Networking Project(UCN), Ministry of Knowledge Economy (MKE) 21st Century Frontier R&D Program in Korea and a result subproject UCN 10C2-T3-12T.

References

[1] D. Anguelov, K. Lee, S. B. Gokturk and B. Sumengen, "Contextual Identity Recognition in Personal Photo Albums," *In Proc. of IEEE CVPR*, pp. 1-7, 2007.
[2] Y. J. Choi, K. J. Kim and W. D. Cho. "Grid-based Approach for Detecting Head and Hand Regions," *In Proc. of ICIC 2007, CCIS 2*, pp. 1126-1132, 2007.
[3] Y. Freund and R. E. Schapire. "Decision-theoretic Generalization of Online Learning and an Application to Boosting," *In Computational Learning Theory: Eurocolt'95*, pp. 23-37, 1995.
[4] M. Gervautz and W. Purgathofer. "A Simple Method for Color Quantization: Octree Quantization," *Graphics Gems,* Morgan Kaufmann, 1990.
[5] A. Gorras, F. Tous, J. Llads and M. Vanrell. "High-level Clothes Description Based on Color-texture and Structural Features," *Lecture Notes in Computer Science*, 2652, pp. 108-116, 2003
[6] X. Hongli, X. De and G. Yong. "Region-based Image Retrieval Using Color Coherence Region Vectors," *In Proc. of the 7th International Conference on Signal Processing*, pp. 761-764, 2004.
[7] K. J. Kim, S. M. Park and Y. J. Choi, "Clothing Identification Based on Edge Information," *In Proc. of APSCC*, IEEE Asia-Pacific Services Computing Conference, pp.876-880, 2008.
[8] N. W. Kim, T. Y. Kim and J. S. Choi. "Edge-based Spatial Descriptor for Content-based Image Retrieval," *Lecture Notes in Computer Science*, 3568, pp. 454-464, 2005.
[9] M. S. Lew, N. Sebe, C. Djeraba and R. Jain. "Content-based Multimedia Information Retrieval: State of the Art and Challenges," *ACM Transactions on Multimedia Computing, Communications & Applications*, 2(1), pp. 1-19, 2006.
[10] D. Ramanan, D. A. Forsyth and A. Zisserman. "Tracking People by Learning Their Appearance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1), pp. 65-81, 2007.
[11] M. J. Swain and D. H. Ballard. "Color Indexing," *International Journal of Computer Vision*, 7(1), pp. 11-32, 1991.
[12] F. Tous, A. Borras, R. Benavente, R. Baldrich, M. Vanrel and J. Llads. "Textual Descriptors for Browsing People by Visual Appearance," *Lecture Notes in Computer Science*, 2504, pp. 419-429, 2002.
[13] J. Z. Wang, J. Li and G. Wiederhold. "Simplicity: Semantics-sensitive Integrated Matching for Picture Libraries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9), pp. 947-963, 2001.
[14] J. Yang, D. Zhang, A. F. Frangi and J. Y. Yang, "Two-Dimensional PCA:A New Approach to Appearance-Based Face Representation and Recognition," *IEEE Trans. on PAMI*, 26(1), pp. 131-137, 2004.
[15] W. Zhao, R. Chellappa, P. J. Phillips and A. Rosenfeld. "Face Recognition: A Literature Survey," *ACM Computing Survey*, 35(4), pp. 399-458, 2003.
[16] S. Zhou, V. Krueger and R. Chellappa. "Probabilistic Recognition of Human Faces from Video," *Computer Vision and Image Understanding*, 91(1-2), pp. 214-245, 2003.

〈저자소개〉

최유주

- 1989년 이화여자대학교 전자계산학과(이학사)
- 1991년 이화여자대학교 전자계산학과(이학석사)
- 2005년 이화여자대학교 컴퓨터학과(공학박사)
- 1991년1월~1993년6월 한국컴퓨터주식회사 기술연구소 주임연구원
- 1994년5월~1999년10월 포스데이타주식회사 기술연구소 주임연구원
- 2005년9월~2010년 서울벤처정보대학원대학교 컴퓨터응용기술학과 조교수
- 2009년3월~현재 한독미디어대학원대학교 뉴미디어학부 조교수
- 관심분야 : 컴퓨터 그래픽스, 가상현실, HCI, 컴퓨터 비전, 의료영상처리 등

박선미

- 2005년 경북대학교 한문학과(문학사)
- 2007년 경북대학교 전자전기컴퓨터학부(공학석사)
- 2008년~현재 경북대학교 전자전기컴퓨터학부 박사과정
- 관심분야 : 컴퓨터 그래픽스, GPGPU, 패턴인식 등

조위덕

- 1987년 한국과학기술원 전기 및 전자공학과 박사
- 1983년~1991년 금성전기(현LG전자) 기술연구소 DSP 연구실장
- 1991년~2003년 전자부품연구원(KETI) 시스템연구본부 본부장
- 1993년 미국 TCSI/Berkeley PCG Group 공동개발연구원
- 1994년 미국 TTP/Cambridge GSM Division 공동개발연구원
- 현 지식경제부 21세기 프론티어사업 (재)유비쿼터스컴퓨팅(UCN)사업단당
- 현 아주대학교 유비쿼터스시스템연구센터장
- 현 아주대학교 전자공학부 교수
- 관심분야 : U 라이프케어서비스디자인, U city 서비스디자인, 모바일앱디자인, 스마트디바이스개발(스마트베드, 스마트 미러, 인간행동측정기), 스마트 CCTV

김구진

- 1990년 이화여자대학교 전자계산학과 학사
- 1992년 한국과학기술원 전자계산학과 석사
- 1998년 포항공과대학교 컴퓨터공학과 박사
- 1998년~2000년 Dept. of Computer Sciences, Purdue University, PostDoc
- 2000년~2002년 아주대학교 정보및컴퓨터공학부 BK21 조교수
- 2002년~2003년 Dept. of Mathematics and Computer Science, University of Missouri St. Louis, Visiting Assistant Professor.
- 2004년~2007년 경북대학교 컴퓨터공학과 조교수
- 2009년~현재 경북대학교 컴퓨터공학과 부교수
- 관심분야 : 컴퓨터 그래픽스, 컴퓨터 애니메이션, 곡면 및 기하모델링 등