Special issue: 3D image and video technology

Research Paper

Associating faces and names in Japanese photo news articles

Akio KITAHARA[1] and Keiji YANAI[2]
[1,2]*The University of Electro-Communications*

ABSTRACT

We propose a system which extracts faces and person names from news articles with photos on the Web and associates them automatically. The system detects face images in news photos with a face detector and extracts person names from news text with a morphological analyzer. In addition, the bag-of-keypoints representation is applied to the extracted face images for filtering out non-face images. The system uses the eigenface representation as image features of the extracted faces, and associates them with the extracted names by the modified *k*-means clustering in the eigenface subspace. In the experiment, we obtained the 66% precision rate at most regarding association of faces and names.

KEYWORDS

Web photo news, face detection, face recognition, eigenface, name extraction, associating names and faces

1 Introduction

Many commercial news sites exist on the Web, and they deliver many news articles every day. Web news sites are one of important news sources for many people as well as newspapers and TV news programs. The advantage of Web news is easy to search and store, while newspapers and TV news programs do not provide efficient search methods and need large physical or HDD space to store. Due to this useful feature, by accumulating Web news articles, we can use them as a personal news database.

Nowadays, most of Web news articles include photos as well as news texts. Photos can be understood intuitively with just a look without reading. Many of news photos contain person's faces related to the news articles. However, sometimes it is hard to identify persons in the news photos, especially in case that a news article includes multiple faces in its photo and multiple person names in its text. In this paper, we try to extract faces from news photos and person names from news texts, and associate them automatically. In the experiments, we used Yahoo! Japan News as a Web news source. The results are usable as a name-annotated face image database.

The rest of the paper is organized as follows. In Section 2, we present related work briefly. Section 3 describes overview of our system. Section 4 explains how to extract faces and names from news articles on the Web, and Section 5 describes how to associate faces and names. In Section 6, we describe the experimental result. In Section 7, we conclude this paper.

2 Related work

Face recognition processing is needed to extract faces from Web news photos. M. Turk et al. suggested the eigenface method which considered pixel values of a grayscale image to be elements of an image feature vector, compressed the vector by the principal component analysis (PCA), and assumed the compressed vector as an image feature for face recognition [1]. This method has been proved to perform well for recognizing frontal face images. We use the eigenface representation as image features of faces. The detail of the eigenface method is described in Section 4.3.

P. Viola et al. proposed a face detection technique to unify weak classifiers which used rectangular features of local regions by AdaBoost [2]. This method is imple-

Received October 14, 2009; Revised December 14, 2009; Accepted January 4, 2010.
[1)] *kitaha-a@mm.cs.uec.ac.jp*, [2)] *yanai@cs.uec.ac.jp*
DOI: 10.2201/NiiPi.2010.7.8

mented as a face detection module in the Open Computer Vision Library (OpenCV) [3], since it performs much faster than the other methods. In our system, we use this AdaBoost-based face detector module included in the OpenCV to extract face regions from Web news photos.

Person name extraction from texts is also needed for our system. This is a part of a well-known problem in NLP called named-entity detection. There are two kinds of approaches for named-entity detection; one is the dictionary-based approach such as [4], and the other is the machine-learning-based approach such as [5]. In this paper, we use the dictionary-based method.

The idea to associate faces and names in news articles were originally proposed by R. K. Srihari [6]. She proposed the PICTION system which extracts face candidates from a news photo and person names from a caption text of a news photo, and estimates correspondence using simple heuristic rules. Since it was about 20 years ago, there were no good methods to extract faces from images and person names from caption texts, and only a few examples were shows in the paper.

The most well-known face-name association systems in the literature is "Name-it" proposed by S. Satoh et al. [7]. The Name-it system focused on TV news programs. It used transcript texts and video caption texts extracted by video OCR as well as frame images extracted from a video. Since their target is TV news, it tracked faces of the same persons and selects the most frontal faces which are suitable for the eigenface-based face recognition [1].

As a face-name association system for Web news, the system proposed by T. L. Berg et al. is the first one. They proposed associating faces and person names in Web news articles with photos. Their methods employed a modified eigenface method which used kernel PCA and LDA as an image representation, and performed the modified k-means clustering to estimate associations between faces and names [8]. The method T. L. Berg et al. proposed achieved about 95% precision at most in terms of association between faces and names extracted from Web news. As another work for Web photo news, D. Ozkan et al. [9] proposed the system which employs a graph-based association method and SIFT-based face matching [10] instead of eigenface.

The objective of our work is associating faces and names in the Web photo news, which is the same as the objective of the works mentioned above. The methods employed in our work is basically based on the Berg's method. The major difference between theirs and ours is the news source. While the target of Berg's method was Web news written in English, our target is Japanese Web news. In addition, we propose using the bag-of-keypoints method to filter out incorrectly-detected faces from output of the face detector.

3 Overview of the proposed system

First of all, we mention the Web photo news from which we like to extract name-face associations automatically. The data source we assume is not just the Web news but the Web photo news such as Yahoo! Japan News. Being different from usual news articles, the main contents of articles of the Yahoo Photo News are photos. In addition to a photo, each article has a title and a short main text which explain the contents of the photo as shown in Fig. 2.

The processing of our system consists of two stages: an extracting stage and an associating stage.

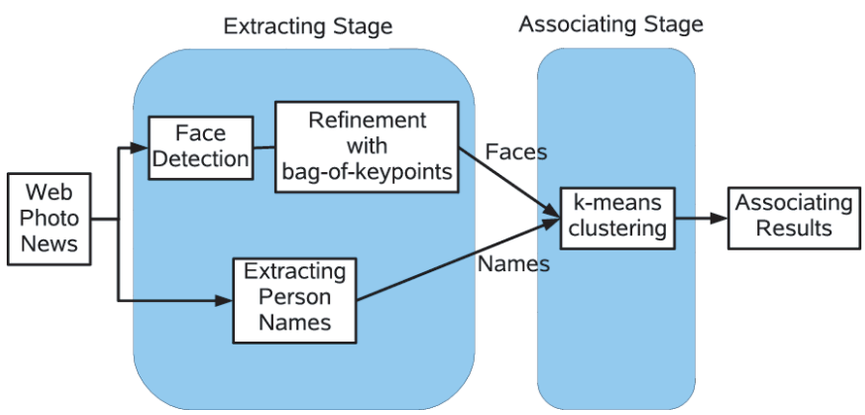

Fig. 1 The processing flow of the proposed system.

Fig. 2 An example of an article of the Yahoo! JAPAN Photo News (This article was published on September 16th, 2009).

In the extracting stage, the system detects face images for news photos with a face detector and extracts person names from news text with a Japanese morphological analyzer. Afterward, the bag-of-keypoints representation [11] is applied to the extracted face to filter out irrelevant face images.

In the associating stage, a candidate list is generated at the beginning. The candidate list records name candidates for each extracted face image. Next, each face is associated with the person name of the best candidate based on the candidate file. Eigenface which is effective for frontal face is used as image features, and the modified *k*-means method is used to estimate associations between faces and names. Finally, post-processing is carried out to resolve some conflicts from the association results. Fig. 1 shows the processing flow of the proposed system.

4 Extracting stage

4.1 Face detection

At first, we extract face regions from Web news photos. To do this, we use the Adaboost-based face detection module included in the Open Computer Vision Library (OpenCV). In our system, we handle only frontal face images since the OpenCV can extract only frontal face images.

4.2 Refinement of face images with bag-of-keypoints

The face detection function in the OpenCV runs very fast, but their output results usually contain some irrelevant non-face images. To filter out them, we apply the bag-of-keypoints representation and SVM [11]. The bag-of-keypoints representation became common recently in the research community of object recognition. It is proved that it has excellent ability to represent image concepts in the context of visual object categorization/recognition in spite of its simplicity.

The basic idea of the bag-of-keypoints representation is that an image is represented by a set of local image features. In our system, to sample a set of local image points and represent them, we use the Scale Invariant Feature Transform (SIFT) [10]. The resulting distribution of description vectors is then quantified by vector quantization against a pre-specified codebook, and the quantified distribution vector is used as an image feature vector of the image. The codebook is generated by the k-means clustering method based on the distribution of SIFT vectors extracted from many training frontal face images in advance. As a classifier to classify images associated with quantified vectors as face or non-face, we use a Support Vector Machine (SVM) classifier.

By applying the bag-of-keypoint-based filtering, the precision of extracted face images were raised from about 70% to 90%.

4.3 Eigenface

To associate faces and names, we need to identify extracted face images. As a representation of extracted face images, we adopt the eigenface representation, which regards the PCA-based-compressed vector of a grayscale face image as a feature vector of the face image [1]. In advance, eigenface vectors are estimated from a set of training frontal face images, and the eigenface subspace spanned by some principal ones out of all the eigenface vectors are generated. In the experiments, we select the principal components corresponding to the several highest eigenvalues so that the cumulative percentage of the sum of eigenvalues of the selected components to the total sum of all the eigenvalues is more than 90%. The PCA compression is carried out by projecting the vector of a greasily face image into the eigen subspace. In our system, all the extracted face images are converted into compressed vectors on the eigen subspace.

4.4 Extracting person names

A Japanese morphological analyzer "ChaSen" [12] is used to extract person names from texts of Web news articles. Based on the output of "ChaSen", we extract person names with the following four patterns with respect to parts of speech.

Extracted person name

- last name + first name (e.g. Koizumi Junichiro)
- last name + position title (e.g. Koizumi Prime-Minister)
- last name + country + position title (e.g. Blair Great-Britain Prime-Minister)
- first name + · + last name (e.g. Michel · Jackson (rule for Western person's name))

In case that the 14 following nouns appear just after family names, it is regarded as position titles.

position titles

prime minister, president, minister, prince, princess, emperor, empress, supervisor, head, coach, CEO, governor, representative, secretariat

The names of journalists who wrote the article texts, and the names of the photographer who took the photo sometimes are included in the news article texts. To ignore them, we drop the names extracted from the last line of article texts. This processing is effective only for Yahoo! Photo News which is the target of our system.

5 Associating stage

5.1 Candidate list

At the associating stage, first of all, the candidate list is generated. At first, we ignore all the articles from which only faces, only names or none of both are extracted, since our system associates names and faces extracted within the same article.

Next, for the news articles from which both faces and person names are extracted, we generate the candidate list which records all the name candidates for each extracted face image.

5.2 Modified *k*-means clustering

Each face in the candidate list is associated with one of the candidate person names by the modified *k*-means clustering. The value of *k* is assumed to be the number of kinds of person names. In short, each cluster corresponds to one unique person name. Therefore, we call a cluster as "a name cluster" in this section. For one-to-one correspondence pairs which mean that only one name and one face are extracted from the same article, the name cluster to which the face is assigned is fixed and not changed during the loop of the *k*-means clustering. That is, the differences between the normal *k*-means and the modified *k*-means are that (1) a face will be assigned with one of the clusters corresponding to candidate names, and (2) one-to-one correspondence pairs are fixed and unchanged. The modified *k*-means clustering is carried out according to the following procedure:

modified *k*-means clustering

1. Decide *k* cluster centers
Each cluster center is chosen from eigenface vectors of all the face images that have the person name associated with the cluster. If a one-to-one correspondence regarding a name and a face in the candidate list is found, the vector of the face is assigned to the cluster center of the name. If not, a center of a name cluster is selected randomly from all the vectors of the faces which have the same name in their name candidates. The cluster centers are assigned in the ascending order of the total number of the candidate names so that centers of all name clusters can be decided.

2. Associate faces to name clusters
Each face except one among one-to-one correspondence pairs is assigned to the nearest one of the name clusters corresponding to its candidate names in the eigenface subspace. The eigenface representation is used to represent face images, and we use the Euclid distance as a metric in the eigenface subspace. In case that the distance exceeds the predefined threshold, the face is not assigned to any clusters.

3. Calculate cluster center
The center vector of each cluster is calculated again by averaging all the vectors of the assigned faces.

4. Judge repeating or finishing
If no cluster centers change at all, the processing is finished. Otherwise, return to 2.

5.3 Post-processing

As a result of the modified *k*-means clustering, two or more faces extracted from the same certain news article might be associated with the same person name. To resolve such conflict, if two or more faces extracted from the same news article are associated with the same person name, only the nearest face to the center of the person name cluster is kept, and all the other faces are removed from the cluster. After this processing, the output is regarded as the final output of face-name associations.

6 Experimental setting

We collected news articles with photos and texts from Yahoo! JAPAN Photo News since March 6, 2005 to August 31, 2006 for one and half year. In the experiments, we used them as experimental data.

As pre-processing, to obtain the eigenface space, we selected 100 relevant face images by hand from all the extracted face images, normalized them to 50×50, and applied PCA in advance.

When we apply the modified k-means, we set a threshold regarding the distance between a face image and the center of a cluster which decides if the face image is joined into the name cluster, or not. As a distance, we used the Euclidean distance. In the experiment, we set the threshold as the square root of the following four kinds of values: (1) none, (2) 6,000,000, (3) 4,000,000, and (4) 2,000,000.

Regarding face-image-filtering with the bag-of-keypoints method, we made experiments with four kinds of the size of the codebook: 100, 300, 500, and 800. By the experiments we select the best size among them and use it for all the other experiments.

To evaluate results, we use the precision rate for all of the experiment results. The precision rate is defined as follows:

$$Precision = \frac{the\ number\ of\ relevant\ face\ output\ images}{the\ \ number\ of\ all\ face\ output\ images} \quad (1)$$

As an evaluation measure for these kinds of work, we have not only the precision rate but also the recall rate. However, since the number of faces included in news articles on the Web is enormous, the more photo news articles we fetch from the Web, the more face images we can get apparently. Therefore, we give priority to improving the precision rate.

7 Experimental results

7.1 Results of face extraction

Fig. 3 and Fig. 4 show the results of face images extracted from the news photos of Yahoo! Photo News. The total number of extracted face images is 38,650. As shown in Fig. 4, some output images were irrelevant images, which included necks of the persons who wore neckties, waists of the persons who put on uniforms, and upper half of circles. Since it was difficult to investigate whether all the extracted images are correct as faces, instead, we chose 100 images randomly and examined them. As a result, the precision rate of face extraction is estimated to be 78%.

Fig. 3 Relevant faces extracted from news photos. (All the faces were extracted from news photos of the Yahoo! JAPAN Photo News.)

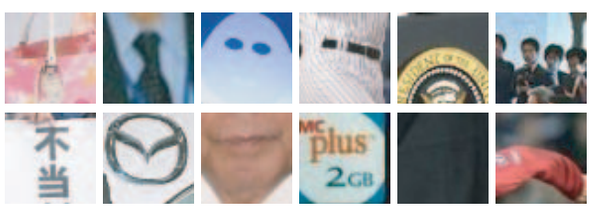

Fig. 4 Irrelevant faces extracted from news photos.

Table 1 Result of filtering non-face.

code words size	faces	precision
100	23,124	93%
300	26,405	95%
500	25,699	97%
800	26,006	98%

7.2 Results of face filtering

The estimated precision rate 78% was not enough for associating faces and names. Then, we filter out irrelevant images from face images extracted by the AdaBoost-based face detector. The results that non-face images are excluded by the bag-of-keypoints representation are shown in the Table 1. In case that the size of codebook was 800, the best precision rate was obtained, so that we used this size for all the other experiments. Introducing the bag-of-keypoint-based face filtering raised the precision rate from 78% to 98%, which has proved to be very effective.

7.3 Results of name extraction

13,579 kinds of unique person names were extracted in the experiments. In the same way as evaluation of extracted faces, we evaluated 100 kinds of names randomly selected from all the extracted names. As a result, the estimated precision rate is 78%.

Some examples of extracting person names correctly and incorrectly are shown as follows. The following three incorrectly-extracted names resulted from failure of separation of words by the morphological analyzer.

Table 2 Results of Association.

	ALL		OVER_3		BoK-ALL		BoK-OVER_3	
threshold[2]	faces	precision	faces	precision	faces	precision	faces	precision
(1) none	9,329	34%	9,108	32%	7,003	42%	6,802	40%
(2) 6,000,000	6,480	42%	6,027	46%	4,947	42%	4,578	47%
(3) 4,000,000	3,853	40%	3,125	44%	2,852	43%	2,282	46%
(4) 2,000,000	1,281	54%	379	62%	960	44%	223	66%

The morphological analyzer uses its own word dictionary to extract person names. Person names not shown in the dictionary tends to fail to be extracted. In the third case of failure of name extraction, two person's names were extracted as one person name. This problem might be solved by adding person names which frequently appear in the Web news articles to the dictionary.

Some results of name extraction

Relevant names

Koizumi Junichiro, Matsui Hideki,
President Bush, Bobby Valentine
(小泉純一郎, 松井秀喜, ブッシュ米大統領,
ボビー・バレンタイン)

Irrelevant names

Sawa Ichiro (correct name:Aisawa Ichido),
Toda Megumi (correct name: Toda Erica),
Koichi・Suzuki (correct name: Saito Koichi, Suzuki Yoshizumi)
(澤一郎 (相澤一郎), 戸田恵 (戸田恵梨香),
浩一・鈴木 (齋藤浩一・鈴木義純))

7.4 Results of association of faces and names

Table 2 and Fig. 5 shows the results of associating faces and names, which include the number of associations and their precision rate evaluated by random sampling of 100 associates for four kinds of evaluation strategies and four kinds of thresholds. In the table, "ALL" and "OVER_3" mean all the association results and the association results of names associated with more than three faces, respectively. Both are the results without bag-of-keypoint-based face filtering. In addition, "BoK-ALL" and "BoK-OVER_3" mean the results with bag-of-keypoint-based (BoK-based) face filtering. We made experiments with four kinds of thresholds shown in the table, which have been already explained in Section 6.

Table 2 shows that the strictest threshold (4) with BoK-based face filtering gave the best association result among all the results. Although the best precision rate 66% was obtained, the number of associations was only 223 in the same setting, which was only about one forty-second of the result of threshold (1) without BoK. In case of the threshold (4) without BoK, only three face images were associated with the person name ranked in the 21st in terms of the number of associated face images.

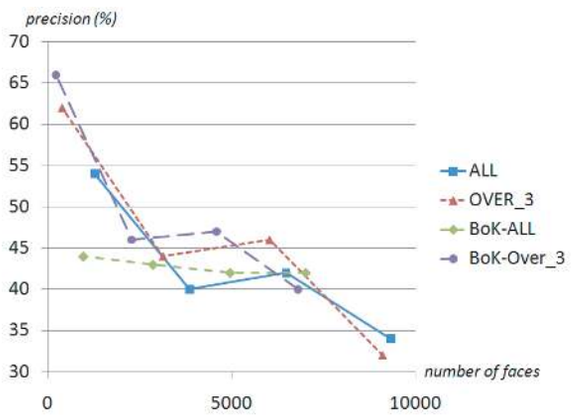

Fig. 5 Results of association.

Compared with "ALL", "OVER_3" tends to better in the precision rate. This implies that the precision rates of names with the less faces were relatively degraded.

As described in Section 7.2, the precision rate of detected faces are improved much by the bag-keypoint-based face filtering. However, regarding the results of association of faces and names, there was no prominent improvement between "non-BoK" and "BoK". This is because the precision rate of extracted names still remains relatively low. Therefore, we need to improve the name extraction part in our system.

Fig. 5 shows a graph where the X-axis and the Y-axis represent the number of faces and precision, respectively. There are four lines corresponding to "ALL", "OVER_3", "BoK-ALL" and "BoK-OVER_3". In this graph, upper lines means better results in terms of

trade-off between the number of faces and their precision. In this sense, "OVER_3" and "BoK-OVER_3" are better than "ALL" and "BoK-ALL".

To clarify the "trade-off" relation between the number of associations and the precision rate, we show a graph on the results of BoK-OVER3 in Fig. 6. The stricter the threshold became, the less associations were obtained while the more the precision rate increased.

Fig. 7 shows all the face images associated with four person names out of the top five ones regarding the number of associated faces, and Table 3 shows the precision rates of the results of the five names. Note that although "Koizumi Junichiro" and "Prime Minister Koizumi" represent the same person, our system cannot identify them as the same person since it has no knowledge about the Japanese prime minister. However, they can be unified as the same person by hand easily as post-processing or by introducing visual-based or textual-based person identification methods. The precision ratio 100% was obtained for "Abe Shinzo", since he never shows his emotions on his face and his facial expression always stays unchanged.

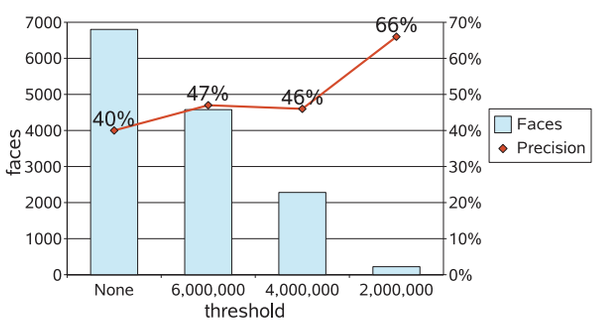

Fig. 6 The precision rates for each threshold.

8 Conclusions and future work

In this paper, we proposed the system which extracts faces from news photos and person names in Japanese news article texts and associated them. In the experiments, 26,006 face images have been extracted in the precision rate 98% with the AdaBoost-based face

Table 3 The number of associated face images and the precision rates of the five names shown in Fig. 7.

person name	faces	precision
Koizumi Junichiro	98	72.4%
Prime Minister Koizumi	55	40.0%
President Bush	42	64.3%
Zico coach*	36	66.7%
Abe Shinzo	21	100%

*..former head coach of the Japanese national soccer team.

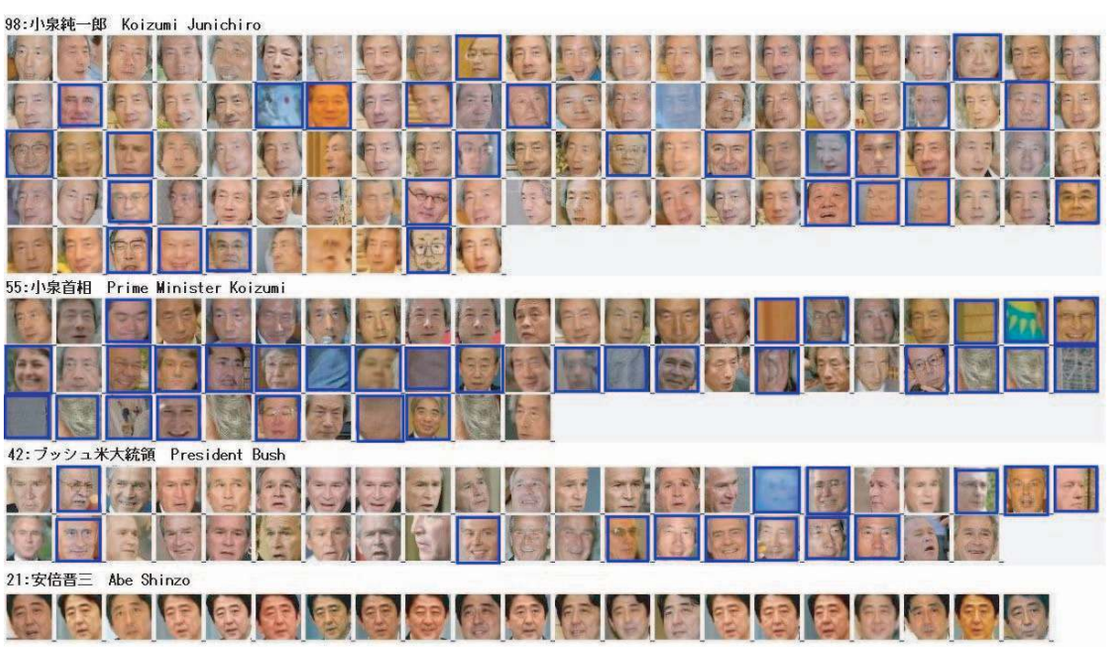

Fig. 7 Extracted faces for the four person names. The faces with blue boxes represent faces assigned wrong names. All the faces were extracted from news photos of the Yahoo! JAPAN Photo News.

detector and the bag-of-keypoint-based filtering, and 13,579 kinds of unique person names have been extracted in the precision rate 78% with a Japanese morphological analyzer "ChaSen" and a name extraction method based on predefined patterns. After extracting faces and names, they are associated by the modified *k*-means clustering in the eigenface space. In the experiments, the 66% precision rate was obtained at most.

Thanks to the bag-of-keypoint-based face filtering we proposed in this paper, we obtained the 98% precision rate in terms of face detection. On the other hand, the precision rate of name extraction remains 78%. This is partly because the proposed system prepares only four kinds of patterns to extract person names from news article texts. Names without titles or family names without first names were not extracted, since the patterns rely on position titles or combination of first and family names. The other reason of relatively low precision rate regarding name extraction is that a Japanese morphological analyzer "ChaSen" we used in the experiments cannot extract person names except names registered in its dictionary. Therefore, the system was not able to extract uncommon Japanese names and names of non-Japanese persons.

For future work, we plan to introduce more sophisticated named entity detection methods to extract person names from Web news articles. And we will also introduce visual-based or textual-based person identification methods as post-processing to merge identical names such as "Junichiro Koizumi" and "Prime Minister Koizumi".

References

[1] M. Turk and A. P. Pentland, "Eigenfaces for recognition," *Cognitive Neuroscience*, vol.3, no.1, pp.71–96, 1991.

[2] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. of IEEE Computer Vision and Pattern Recognition*, vol.1, pp.511–518, 2001.

[3] Open Source Computer Vision Library, *http://sourceforge.net/projects/opencvlibrary/*.

[4] I. Ide, R. Hamada, S. Sakai, and H. Tanaka, "Semantic analysis of television news captions referring to suffixes," in *Proc. of Intl. Workshop on Information Retrieval with Asian Languages*, pp.37–42, 1999.

[5] M. Asahara and Y. Matsumoto, "Japanese named entity extraction with redundant morphological analysis," in *Proc. of Human Language Technology Conference*, pp.8–15, 2003.

[6] R. K. Srihari, "PICTION: A system that uses captions to label human faces in newspaper photographs," in *Proc. of the 9th National Conference on Artificial Intelligence*, pp.80–85, 1991.

[7] S. Satoh, Y. Nakamura, and T. Kanade, "Name-It: Naming and detecting faces in news videos," *IEEE Multimedia*, vol.6, no.1, pp.22–35, 1999.

[8] T. L. Berg, A. C. Berg, J. Edwards, M. Maire, R. White, Y-W. Teh, E. Learned-Miller, and D. A. Forsyth, "Names and faces in the news," in *Proc. of IEEE Computer Vision and Pattern Recognition*, pp.848–854, 2004.

[9] D. Ozkan and P. Duygulu, "A graph based approach for naming faces in news photos," in *Prof. of IEEE Computer Vision and Pattern Recognition*, pp.1477–1482, 2006.

[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol.60, no.2, pp.91–110, 2004.

[11] G. Csurka, C. Bray, C. Dance, and L. Fan, "Visual categorization with bags of keypoints," in *Proc. of ECCV Workshop on Statistical Learning in Computer Vision*, pp.59–74, 2004.

[12] ChaSen, *http://chasen.naist.jp/hiki/ChaSen/*.

Akio KITAHARA

Akio KITAHARA received B. Eng. and M. Eng. from the University of Electro-Communications in 2007 and 2009. He is currently working for Automotive System Division, Toshiba Corporation. His research interests include image recognition and Web multimedia mining.

Keiji YANAI

Keiji YANAI received B. Eng., M. Eng. and Dr. Eng. degrees from the University of Tokyo in 1995, 1997 and 2003, respectively. From 1997, he was a research associate of Department of Computer Science, the University of Electro-Communications. From 2006, he is an associate professor of Department of Computer Science, the University of Electro-Communications. His recent research interests include image recognition, multimedia processing and Web multimedia mining.