Low-resolution Person Recognition using Image Downsampling

Keiji Obara, Hiroki Yoshimura, Masashi Nishiyama and Yoshio Iwai
Graduate School of Engineering, Tottori University, Japan
nishiyama@eecs.tottori-u.ac.jp

Abstract

We propose a novel method for identifying individuals by resampling images using resolution inference to increase the performance of person recognition. The resolution of images dynamically changes the appearances of faces or whole bodies and low resolution decreases the performance of person recognition. To overcome the problem of resolution, we need to adequately normalize the sizes of images using resampling techniques before identification. In our preliminary experiment, we observe that downsampling high-resolution images to adjust to low-resolution images increases the identification performance while upsampling low-resolution images to adjust to high-resolution images decreases performance. The proposed method thus applies a downsampling technique to images of higher resolution by comparing the resolutions inferred for query and target images. We demonstrate that our method substantially improves the performance of person recognition on the publicly available datasets Multi-PIE and CUHK01 artificially degraded to low resolution.

1 Introduction

Advances in identification using person recognition technology has seen the development of convenient and non-intrusive biometric authentication systems. There are two major identification processes in person recognition technology: face recognition [1] using facial images and person re-identification [2] using whole-body images. Authentication systems based on person recognition technology frequently use cameras equipped with wide-angle lenses. The images acquired from the cameras have various resolutions, which affects the identification performance. The resolution dynamically affects the appearances of facial and whole-body images. The appearances of individuals clearly reveal the identities of individuals when images are of high resolution whereas different individuals tend to appear more similar at low resolution.

Resampling is widely used to normalize the sizes of images of different resolution as a preprocessing step before the identification process. The normalization process makes the dimensions of features extracted from query and target images uniform for the application of generic machine learning techniques. Although the sizes of normalized images are uniform, the identification performance decreases at low resolution. The effect of resolution that exists before resampling thus remains after resampling. There are existing methods that alleviate the effect of low resolution on face recognition [3, 4, 5, 6, 7] and person reidentification [8, 9]. The existing methods assume that target images are of high resolution in the synthesis with low-resolution query images and they do not sufficiently consider the problem of the low resolution of the target images. We cannot assume that target images are of high resolution in real-world applications. We must therefore consider that both query and target images are of low resolution.

We propose in this paper a novel method for identifying faces and whole bodies by inferring resolutions of both query and target images and resampling the images to adequate resolution for person recognition. Before designing our method, we investigate how the identification performance responds to changes in the resolutions of query and target images. We observe that downsampling high-resolution images to adjust to low-resolution images improves the identification performance. Our method thus applies a downsampling technique that adjusts high-resolution images to low-resolution images. To this end, we infer the resolutions of query and target images. We demonstrate the good identification performance of the proposed method on standard person recognition datasets artificially degraded to low resolution.

2 Related work

To tackle the problem of low resolution in face recognition, some existing methods [3, 4] exploit super-resolution techniques to upsample query images. Wong et al. [5] proposed a method of switching the parameters of identification algorithms learned from high-resolution target images using the resolution of the query images. Other existing methods [6, 7] synthesize low-resolution images by downsampling high-resolution target images and extract discriminative features from pairs of high-and low-resolution target images. Furthermore, other existing methods [8, 9] synthesize pairs of images of the two resolutions to extract discriminative features for person re-identification. Note that we assume the resolutions of the images are known in the simulation. The details of the simulation are described in the following section.

3 Analysis of the performance of person recognition at low resolution

3.1 Background

The process of degrading images from high resolution to low resolution is defined as

$$\boldsymbol{g} = \boldsymbol{B}\boldsymbol{H}\boldsymbol{f} + \boldsymbol{n}, \tag{1}$$

where vector $\boldsymbol{g}$ represents a low-resolution image, vector $\boldsymbol{f}$ a high-resolution image, matrix $\boldsymbol{B}$ resampling, and matrix $\boldsymbol{H}$ blur. Note that we assume $\boldsymbol{B}$ and $\boldsymbol{H}$ are shift invariant. We analyze the effect of resolution on the performance of person recognition. We conduct a simulation by degrading the images while changing

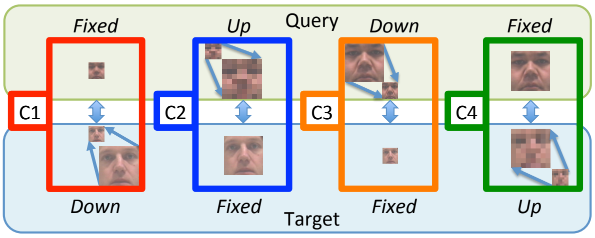

Figure 1. Experimental conditions of resampling for face recognition.

the parameters of $\boldsymbol{B}$, $\boldsymbol{H}$, and $\boldsymbol{n}$ in Equation (1). Note that we assume the resolutions of the images are known in the simulation. The details of the simulation are described in the following section.

3.2 Identifying individuals using high- and low-resolution images

We evaluate the identification performance while changing the query and target images from high resolution to low resolution. As the preprocessing of the identification process, we apply four resampling techniques:

- **C1**: Downsampling the target images to adjust to the query images,
- **C2**: Upsampling the query images to adjust to the target images,
- **C3**: Downsampling the query images to adjust to the target images,
- **C4**: Upsampling the target images to adjust to the query images.

We compare the performance while changing the resolution of the query images in C1 and C2 or the target images in C3 and C4. We use bilinear interpolation for upsampling and downsampling. We report the performances of face recognition in Section 3.2.1 and person re-identification in Section 3.2.2.

3.2.1 Performance of face recognition

We conduct experiments on face recognition for C1, C2, C3, and C4 as illustrated in Figure 1. We use the Multi-PIE [10] data set. There are 337 target images (one per person) and 674 query images (two per person). The target images are frontal views of faces acquired without a flash. The query images are acquired with a flash from the front or from the right at an angle of 15 degrees. The facial expression is neutral in all images. We align the positions of the pupils, mouth corners, and top of the nose labeled in [11]. We set the parameters of resolutions $\boldsymbol{B}_i (i = 1, \ldots, 5)$ as $128 \times 128, 64 \times 64, 32 \times 32, 16 \times 16$, and 8×8 pixels. Figure 2 (a) shows examples of facial images. We apply Gaussian blur $\boldsymbol{H}_1, \boldsymbol{H}_2$ of $\sigma = 1, 2$, and white Gaussian noise $\boldsymbol{n}_1, \boldsymbol{n}_2$ for a signal-noise ratio of 40 and 30 dB. We also check cases of no blur $\boldsymbol{H}_0 = \boldsymbol{I}$ and no noise $\boldsymbol{n}_0 = \boldsymbol{0}$. We use features for identification obtained by raster scanning facial images after applying the single-scale Retinex (SSR) algorithm [12]. We manually control the kernel parameter of SSR by adjusting to the resolution of images. We use a nearest-neighbor classifier to identify individuals.

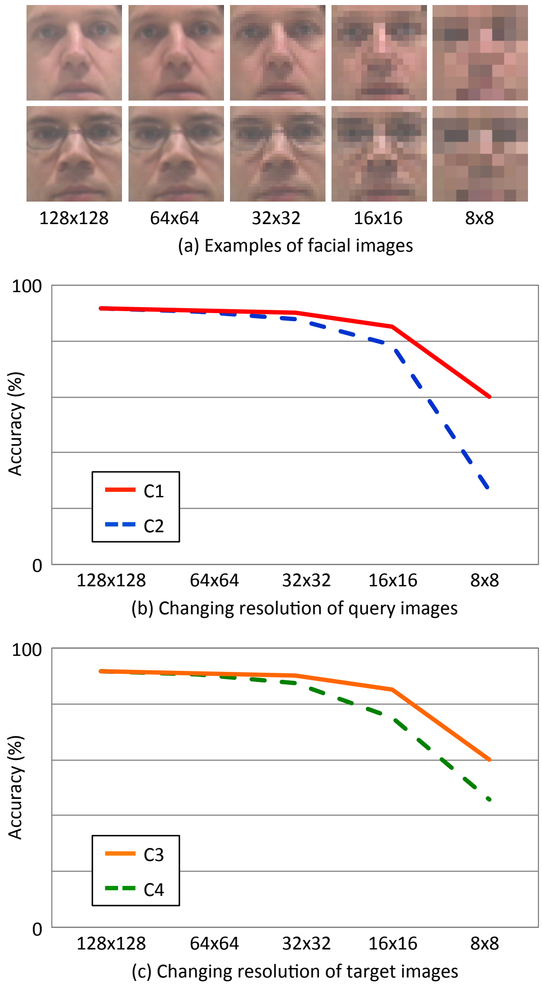

Figure 2. Identification performance of face recognition. We assume that the resolutions of facial images are known.

Figure 2 (b) shows the identification performance of face recognition for C1 and C2 while Figure 2 (c) shows those for C3 and C4. We see that C1 and C3 outperform C2 and C4; i.e., downsampling to adjust to low resolution outperforms upsampling to adjust to high resolution in terms of the identification performance.

3.2.2 Performance of person re-identification

We conduct the experiments on person re-identification for C1, C2, C3, and C4 as illustrated in Figure 3. We use the CUHK01 [13] data set. There are 971 target images (one per person) and 971 query images (one per person). We use whole-body images resized to 60×160 pixels and captured by the same camera at different times. We set the parameters of resolutions $\boldsymbol{B}_i$ as $60 \times 160, 48 \times 128, 18 \times 48, 12 \times 32$, and 6×16 pixels. We use the same parameters of $\boldsymbol{H}_0, \boldsymbol{H}_1, \boldsymbol{H}_2, \boldsymbol{n}_0, \boldsymbol{n}_1$ as used in Section 3.2.1. Figure 4 (a) shows examples of whole-body images. We use feature vectors for identification using histograms of oriented gradients [14] in the CIE L*a*b* color space. We manually control the parameters of the histograms of oriented gradients by adjusting to the resolution of images. We use a nearest-neighbor classifier to identify individuals.

Figure 4 (b) shows the identification performance of person re-identification for C1 and C2 while Figure 4

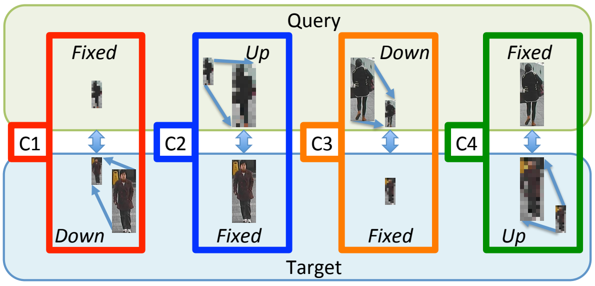

Figure 3. Experimental conditions of resampling for person re-identification.

(c) shows those for C3 and C4. We again see that C1 and C3 outperform C2 and C4. We believe that downsampling is useful for person re-identification as well as face recognition.

4 Person recognition by inferring resolution

4.1 Overview

We observe that downsampling provides better identification performance than upsampling. We thus apply downsampling to adjust to the lower resolution between query and target images. The proposed method infers the resolution of images before downsampling. An overview of our method is described below.

1. Synthesize training samples to infer resolution by degrading high-resolution images and give the parameters of resolution as labels.
2. Generate a classifier of resolution using a machine learning technique.
3. Given the query and target images, infer the labels of the resolution for each image using the classifier.
4. Compare the labels of the resolutions and apply downsampling to the images of higher resolution.
5. Apply a person recognition method to the downsampled images.

4.2 Feature space for inferring resolution

As a simple method for inferring resolutions, we can determine the resolution from the size of facial or whole-body regions detected from video image sequences. However, the relationship between the resolution and size frequently varies when changing the distances between the cameras and individuals or the parameters of the camera lens. To stably infer the resolution, we require features that are sensitive to the appearance variations of different resolutions but insensitive to differences among individuals, illumination changes, and pose changes. We compare the following feature spaces to determine features for inferring resolution.
Gray: Direct use of pixel values.
Sobel: Use of pixel values after applying a Sobel filter.
Magnitude: Use of the value of the power spectrum after applying a discrete Fourier transform.
We apply four-fold cross-validation to divide the individuals of training and test samples of resolution inference on Multi-PIE or CUHK01 datasets. We use the same parameters of resolution as used in Sections 3.2.1 and 3.2.2. We apply a linear support vector machine.

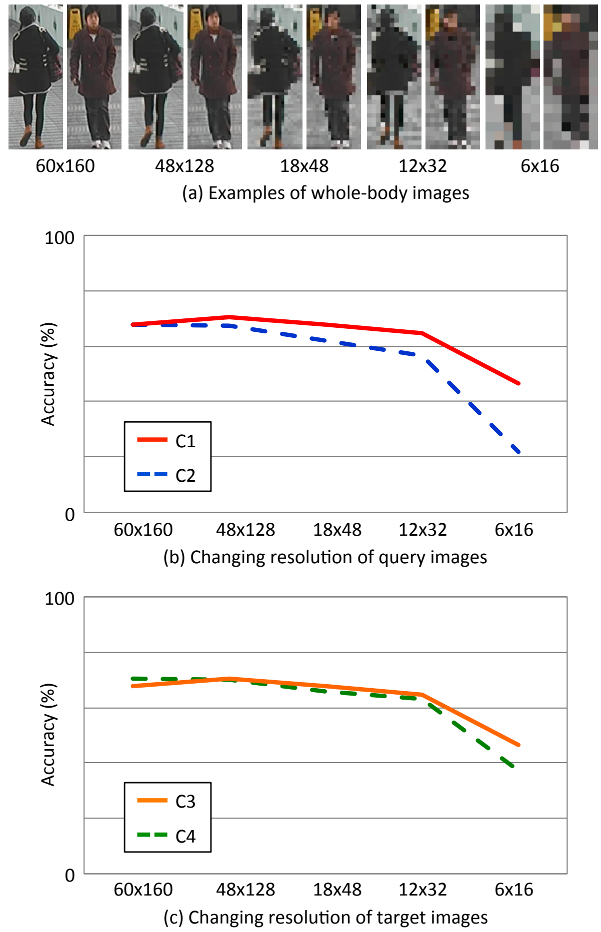

Figure 4. Identification performance of person re-identification. We assume that the resolutions of whole-body images are known.

Table 1 gives the accuracy of resolution inference. We see that the Sobel features provide better performance than the gray and magnitude features.

4.3 Evaluation of identification performance

We evaluate the identification performance achieved using resolution inference by comparing results obtained with the following methods.
Super-resolution: Identifying individuals from the images upsampled following [15] and using the resolution inference.
Random: Identifying individuals from images randomly downsampled.
Our method: Identifying individuals from images downsampled using the resolution inference.
We apply two-fold cross-validation to divide individuals of training samples of resolution inference and test images of identification on Multi-PIE and CUHK01 datasets. We use the same parameters of resolution and classifier of individuals as used in Sections 3.2.1 and 3.2.2. The Sobel features are used for resolution inference.

Table 1. Accuracy (%) of resolution inference.

Image	Gray	Sobel	Magnitude
Face	83.1	**90.0**	66.1
Whole-body	38.1	**59.3**	**59.9**

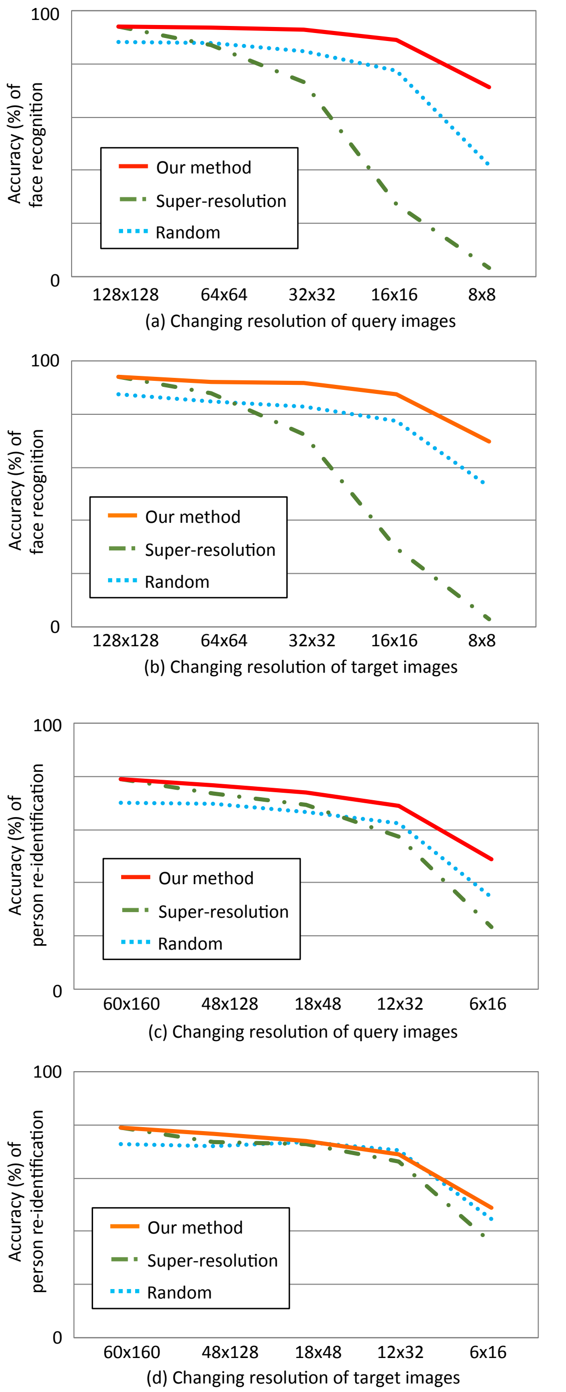

Figure 5. Identification performance. Our method applies downsampling to images after automatically inferring resolutions.

Figure 5 (a) and (b) shows the performance of face recognition while 5 (c) and (d) shows the performance of person re-identification. We see that our method is superior to existing methods when the query and target images are of low resolution.

5 Conclusion

We proposed a method for inferring the resolutions of images and downsampling images of higher resolution as a preprocessing step in person recognition. We demonstrated that image downsampling increased the identification performance of face recognition and person re-identification. We intend to evaluate the performance of the proposed method on real-world datasets as future work.

References

[1] S.Z. Li and A. Jain. *Handbook of Face Recognition.* Springer, 2nd edition, 2011.

[2] S. Gong, M. Cristani, S. Yan, and C.C. Loy. *Person Re-Identification.* Springer, 2014.

[3] D. Jia and S. Gong. Multi-modal tensor face for simultaneous super resolution and recognition. In *Proceedings of International Conference on Computer Vision*, pages 1683–1690, 2005.

[4] J. S. Park and S. W. Lee. An example-based face hallucination method for single frame, low resolution facial images. *IEEE Transactions on Image Processing*, 17(10):1806–1816, 2008.

[5] Y. Wong, C. Sanderson, S. Mau, and B.C. Lovell. Dynamic amelioration of resolution mismatches for local feature based identity inference. In *Proceedings of 20th International Conference on Pattern Recognition*, pages 1200–1203, 2010.

[6] S. Shekhar, V. M. Patel, and R. Chellappa. Synthesis-based recognition of low resolution faces. In *Proceedings of International Joint Conference on Biometrics*, pages 1–6, 2011.

[7] W.W.W. Zou and P.C. Yuen. Very low resolution face recognition problem. *IEEE Transactions on Image Processing*, 21(1):327–340, 2012.

[8] X.Y. Jing, X. Zhu, F. Wu, X. You, Q. Liu, D. Yue, R. Hu, and B. Xu. Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 695–704, 2015.

[9] X. Li, W. S. Zheng, X. Wang, T. Xiang, and S. Gong. Multi-scale learning for low resolution person re-ident -ification. In *Proceedings of IEEE International Conference on Computer Vision*, pages 3765–3773, 2015.

[10] R. Gross, I. Matthews, J.F. Cohn, T. Kanade, and S. Baker. Multi-PIE. *Image and Vision Computing*, 28(5):807 – 813, 2010.

[11] L.E. Shafey, C. McCool, R. Wallace, and S. Marcel. A scalable formulation of probabilistic linear discriminant analysis: Applied to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7):1788–1794, 2013.

[12] D. J. Jobson, Z. Rahman, and G. A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing*, 6(7):965–976, 1997.

[13] W. Li, R. Zhao, and X. Wang. Human reidentification with transferred metric learning. In *Proceedings of 11th Asian Conference on Computer Vision*, pages 31–44, 2013.

[14] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 886–893, 2005.

[15] S. Villena, M. Vega, S.D. Babacan, R. Molina, and A.K. Katsaggelos. Bayesian combination of sparse and non-sparse priors in image super resolution. *Digital Signal Processing*, 23(2):530 – 541, 2013.