Headgear Recognition by Decomposing Human Images in the Thermal Infrared Spectrum

Brahmastro Kresnaraman*, Yasutomo Kawanishi*, Daisuke Deguchi†, Tomokazu Takahashi‡, Yoshito Mekada§, Ichiro Ide* and Hiroshi Murase*
*Graduate School of Informatics, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan
Email: brahmastrok@murase.is.i.nagoya-u.ac.jp, {kawanishi, ide, murase}@i.nagoya-u.ac.jp
†Information Strategy Office, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan
Email: ddeguchi@nagoya-u.jp
‡Faculty of Economics and Information, Gifu Shotoku Gakuen University, 1-38 Naka-uzura, Gifu 500-8288, Japan
Email: ttakahashi@gifu.shotoku.ac.jp
§Graduate School of Computer and Cognitive Science, Chukyo University, Tokodachi, Kaizu-cho, Toyota 470-0393, Japan
Email: y-mekada@sist.chukyo-u.ac.jp

***Abstract*—Surveillance systems play a critical role in security and surveillance. A surveillance system with cameras that work in the visible spectrum is sufficient for most cases. However, problems may arise during the night, or in areas with less than ideal illumination conditions. Cameras with thermal infrared technology can be a better option in these situations since they do not rely on illumination to observe the environment. Furthermore, in our daily lives, it is common for humans to wear headgears such as glasses, masks, and hats. In surveillance, such headgears can be a hindrance to the identification of a person, and hence pose a certain degree of risk. This is not ideal in areas where the identity of a person is important, for example, in a bank. Therefore, in this paper we propose a headgear recognition method using an innovative decomposition approach on thermal infrared images. The decomposition method is based on Robust Principal Component Analysis, a modification of the popular Principal Component Analysis. The proposed method performs decomposition on a human image and isolates headgears in the image for recognition purposes. Experiments were conducted to evaluate the capability of the proposed method. The results show a positive outcome when compared with other methods.**

I. Introduction

In recent years, we often see security cameras in various locations, especially in commercial or public buildings. These cameras are part of surveillance systems, to maintain security in certain areas-of-interest. These systems are very useful for crime investigations. Furthermore, the presence of a security surveillance system may also deter criminal acts, since it is likely for the cameras to capture the identity of the person-of-interest. However, as a tool to prevent incidents from happening, it is only relatively adequate. In practice, it is difficult for the system to perform its duty in real-time. In common practice, there is usually a dedicated person observing the camera feed. This causes an inconsistent level of security due to fatigue, and other human errors. Automation of some tasks can be very beneficial to achieve a consistent and reliable security surveillance systems. Some of the tasks that can be automated includes, but not limited to, face recognition.

Additionally, surveillance systems generally work continuously without stopping for 24 hours a day. In most cases, unless there are sources of light nearby, areas monitored by the system are dark during the night. Normal cameras that work in the visible spectrum need illumination to be able to capture the scene. Without sufficient illumination, loss of information from the environment is inevitable. In this case, the usage of thermal infrared cameras are a better option. Thermal infrared cameras work in the thermal infrared spectrum. They produce images by capturing infrared radiation from objects, whose intensity depends on the temperature of the said object. This is fundamentally different with normal cameras that capture visible light to produce images. Nighttime surveillance is made possible by utilizing these thermal infrared cameras. Fig. 1 shows an example of a scene captured in the visible and

Fig. 1: Examples of images captured by cameras in different spectra. Left: An image captured in the visible spectrum. Right: An image captured in the thermal infrared spectrum.

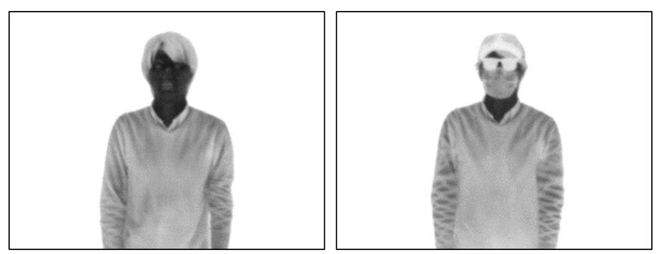

Fig. 2: Examples of images in thermal infrared spectrum. Left: An image of a person without a headgear. Right: An image of a person with headgears (hat, glasses, and mask).

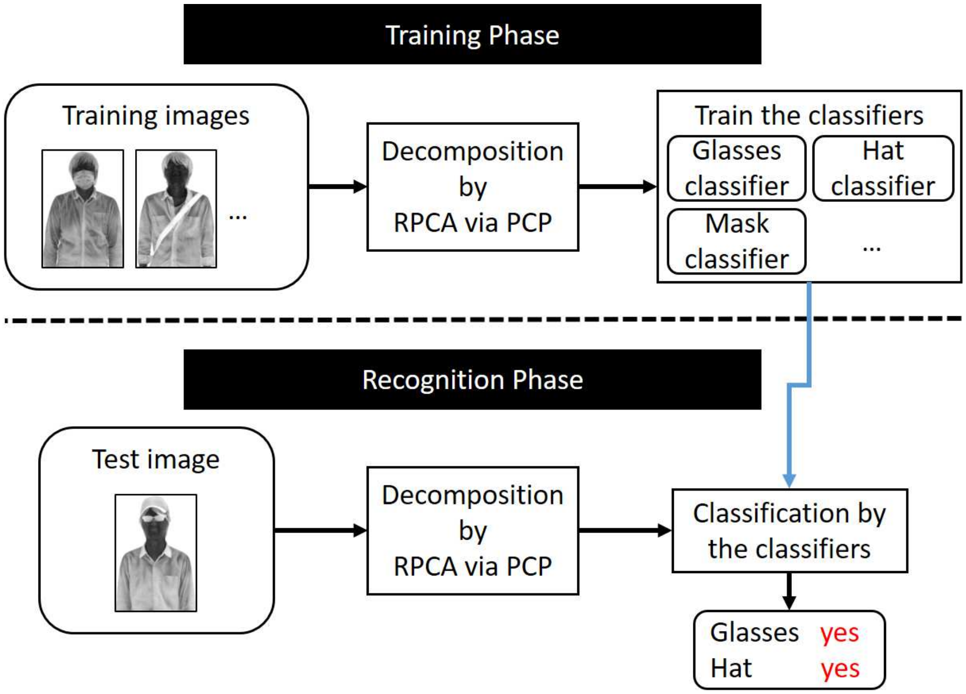

Fig. 3: Process flow of the proposed method.

thermal infrared domain.

In our daily lives, it is common for humans to wear various headgears such as glasses, hats, and other accessories. However, in the subject of face recognition, the presence of these headgears is an obstacle. They often occlude parts of the face, which can hinder the capability of most face recognition methods. This is particularly problematic in areas where the identity of a person is important, for example, in a bank. Fig. 2 shows an image example taken in the thermal infrared spectrum, with and without a headgear. As we can see, the presence of humans in the thermal infrared image is relatively easy to detect. However, identifying them is not an easy task. This difficulty also increases due to the existence of headgears, as we can also see in Fig. 2. Therefore, it is important to recognize these headgears because they can also be considered as a security risk.

In previous years, there has been no research focusing specifically on headgear recognition. The most similar research subject would be attribute recognition which are related to the humans attributes. For example, classifications of human expressions [1], [13], race or ethnicity [14], and gender [5], [18]. Recognition of other attributes, including some headgears and various basic clothing, are being utilized for face verification [10], [11], people search [4], [17], [15], [3] and person re-identification [12], [7]. However, all these researches have been done in the visible spectrum. In the thermal infrared spectrum, there are fewer researches available. Facial expressions recognition [6], [16] and a brief experiment on eyewear detection [17] are some of the examples. To the extent of our knowledge, our previous works [8], [9] are the only research performed on recognizing headgears as attributes in the thermal infrared spectrum.

In our previous work [8], we adopted the decomposition approach before performing the recognition. The decomposition is based on the idea of majority and minority. The implementation of the decomposition is achieved by utilizing Robust Principal Component Analysis (RPCA) [2], which is a modification of the popular Principal Component Analysis (PCA). RPCA aims to separate an image into a low-rank matrix $\boldsymbol{L}$ and a sparse matrix $\boldsymbol{S}$ from data $\boldsymbol{M}$, as seen in 1.

$$\boldsymbol{M} = \boldsymbol{L} + \boldsymbol{S} \tag{1}$$

In this research, we utilize the recognition framework proposed in our previous work [8]. As mentioned previously, the idea of majority and minority is the basis of the decomposition. Assuming a collection of thermal infrared images in which most of them do not have a headgear in the image, but some do, the presence of these headgears in this dataset is the minority. Under this assumption, the decomposition can isolate headgears from the original image and use only the headgear component for the recognition.

II. Headgear Recognition Framework

The proposed recognition framework decomposes a thermal infrared image, and separates the headgear components from the non-headgear components of the image. This approach achieves decomposition by utilizing RPCA. Principal Component Pursuit (PCP) is employed as the chosen algorithm to solve the RPCA problem, and is implemented in the

(a) Disguise 2 and motorcycle mask (b) Glasses, motorcycle mask, and helmet (c) Glasses and fake mustache (d) Mask

Fig. 4: Image examples from the dataset.

decomposition part of the framework. Fig. 3 shows the process flow of the proposed framework.

A. Training Phase

The training phase includes the decomposition of the thermal infrared images and the training of the classifier for recognition. Before the decomposition process commences, the available thermal infrared images for training need to be separated into two groups. The first group contains thermal infrared images without any headgear worn. These images are used as base data $\boldsymbol{B} = (\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_I)$. $\boldsymbol{B}$ contains I observations, which will also be used in the recognition phase. The second group contains thermal infrared images with the person wearing at least one headgear. These images are the training data, represented as $\boldsymbol{D} = (\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_J)$, with J observations.

The RPCA problem is described in 1. The decomposition utilizes PCP to solve it, adapting the idea of majority and minority. The problem that need to be solved then changed to the following:

$$\begin{bmatrix}\boldsymbol{B} & \mathbf{d}_j\end{bmatrix} = \boldsymbol{M}^j = \boldsymbol{L}^j + \boldsymbol{S}^j, \tag{2}$$

where

$$\boldsymbol{S}^j = \begin{bmatrix}\boldsymbol{S}^j_{\boldsymbol{B}} & \mathbf{x}_j\end{bmatrix}. \tag{3}$$

This process is performed for every training data and then grouped together as $\boldsymbol{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_J)$, with J observations. After the decomposition process, the training of the classifier is performed. In this framework, virtually any classifier can be used. For example, since there are positive and negative samples for each headgear, a binary classifier can be used. When the training of the classifier is concluded, the training phase of the headgear recognition framework ends.

B. Recognition Phase

For a thermal infrared image $\mathbf{t}$, we aim to recognize which headgear is present in the image. To do so, we go through the same process as in the training phase. The first step is to decompose the image $\mathbf{t}$. The optimization problem is solved by PCP as follows:

TABLE I: Distribution of headgears in the dataset.

Headgears	# of images
No headgear	64
Balaclava (full-face)	25
Balaclava (half-face)	110
Disguise 1	59
Disguise 2	38
Disguise 3	38
Disguise 4	12
Glasses	287
Hat	162
Helmet (half-face)	182
Surgical mask	138
Motorcycle mask	110
Fake mustache	116
Wig	123

$$\begin{bmatrix}\boldsymbol{B} & \mathbf{t}\end{bmatrix} = \boldsymbol{M}^{\mathbf{t}} = \boldsymbol{L}^{\mathbf{t}} + \boldsymbol{S}^{\mathbf{t}}, \tag{4}$$

where

$$\boldsymbol{S}^{\mathbf{t}} = \begin{bmatrix}\boldsymbol{S}^{\mathbf{t}}_{\boldsymbol{B}} & \mathbf{y}\end{bmatrix}. \tag{5}$$

In this case, $\mathbf{y}$ is the sparse component of thermal infrared image $\mathbf{t}$ which contains only the headgear. The sparse component $\mathbf{y}$ then serve as the input to the classifier chosen in the training phase. The output of the classifier is then the final result of the recognition phase.

III. Experiments, Results and Analysis

In this section, we provide explanation regarding the dataset used in this research and the experiment performed. Additionally, we also provide some analysis and discussion on the experimental results.

TABLE II: Results of the proposed method and comparative methods evaluated in F-Score.

Methods	Average F-Score
Without decomposition (baseline)	0.933
Conventional Average	0.944
Proposed method	**0.948**

TABLE III: Detailed results per headgear of the proposed method and comparative methods evaluated in F-Score.

Headgears	Methods		
	Without decomposition (Baseline)	Conventional Average	Proposed method
Balaclava (full-face)	0.83	0.90	**0.93**
Balaclava (half-face)	**0.93**	0.92	**0.93**
Disguise 1	0.90	**0.92**	0.90
Disguise 2	0.93	0.93	**0.95**
Disguise 3	0.93	0.93	**0.94**
Disguise 4	0.94	**0.99**	0.97
Glasses	**0.96**	**0.96**	**0.96**
Hat	**0.92**	0.90	**0.92**
Helmet (half-face)	0.95	0.95	**0.98**
Surgical mask	0.93	**0.95**	0.94
Motorcycle mask	**0.94**	0.92	0.93
Fake mustache	**0.99**	**0.99**	0.98
Wig	**0.99**	**0.99**	**0.99**

A. Dataset

Since there are no publicly available thermal infrared datasets available for our purpose, we created our own dataset. We captured 775 thermal infrared images of humans taken frontally. There are 13 types of headgear available in the dataset, whose distribution is shown in Table I. The subjects in the images may wear an individual headgear or a combination of multiple headgears. Some image examples from the dataset are presented in Fig. 4.

The average size of the human body image is 140×204 pixels. The captured thermal image shows temperatures ranging from 25 to 36 degrees Celcius, taken indoors at room temperature. The thermal infrared image is represented in "hotblack" scheme, which means the closer the pixel value to zero, the higher the temperature.

B. Experimental Setup

The experiment is performed to evaluate the capability of the proposed method. The results are then compared to the results by other methods. The first comparative method is the baseline, where the human images are directly used as the input as the classifier without decomposition. The second comparative method is called "Conventional Average" for convenience. From the base data $\boldsymbol{B} = (\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_I)$, we calculate its average $\bar{\mathbf{b}}$. This average is used to "decompose" both the training and the test data by means of subtraction.

In this paper, the experiment is conducted with cross-validation. As for the classifier chosen for the proposed method, we chose Support Vector Machine (SVM). We reiterate that in principle, any classifier can be used for recognition.

C. Results and Analysis

Table II shows that the proposed method shows an improvement over the baseline method. Table III shows the performance of the proposed method and the other comparative methods for each headgear. It can be seen that the proposed method outperforms other methods in most headgears. Some types of headgears such as the fake mustache and wig show high scores even using the baseline method. This is possibly due to the ideal conditions present at the time of the dataset creation, resulting in very good quality thermal infrared images of each headgear. In real-world surveillance systems, it is less likely that the ideal conditions will be possible. On the other hand, the other types of headgears such as full-face balaclava shows the biggest improvement when using the proposed method, taking full benefit of the decomposition.

The main difference between each of the evaluated algorithms is the decomposition process. Therefore, the information of the processing time/complexity of the algorithms is useful to know. Currently, the decomposition algorithm which we used in the proposed method is relatively time consuming since it performs optimization/minimization for each input. By replacing the optimization process with an approximation algorithm, the processing time can be reduced and the algorithm becomes faster.

IV. Conclusion

This paper addressed the headgear recognition problem in captured human images in the thermal infrared spectrum. The recognition of headgears is necessary for surveillance purposes, for human identification and security. To the extent of our knowledge, there are no other researches specifically preformed to achieve this recognition, even more so in the thermal infrared spectrum.

The headgear recognition approach proposed in this paper adopted a decomposition approach to separate the headgear information from the other information in the thermal infrared image. The proposed method utilized Robust Principal Component Analysis solved by Principle Component Pursuit to perform the decomposition.

To evaluate the proposed framework we conducted experiments on a thermal infrared image dataset. For the purpose of this research, we have created a new thermal infrared dataset. The results of the decomposition show that overall, the proposed method outperforms the other comparative methods. When looking at individual types of headgears, the recognition of the full-face balaclava using the proposed method showed the biggest improvement.

The dataset used for this experiment was captured in ideal conditions and a consistent angle. For real-world surveillance, these conditions are not likely to be achieved. For future work, this research can be expanded to non-frontal human images and various image quality and lighting to simulate real-world conditions. The addition of angle information is a relatively well known problem which provides a big challenge and warrants further research. Lastly, the approach proposed in this paper can be generalized and is not limited to recognition for the purposes of surveillance.

Acknowledgment

Parts of this research were supported by MEXT, Grant-in-Aid for Scientific Research. The authors would like to thank the members of the laboratory for their participation in the creation of the dataset.

References

[1] M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan. Real time face detection and facial expression recognition: Development and applications to human computer interaction. In *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition 2003 Workshops*, volume 5, pages 53–59, 2003.

[2] E. J. Candes, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):11, 2011.

[3] Y. Deng, P. Luo, C. Loy, and X. Tang. Pedestrian attribute recognition at far distance. In *Proc. of 22nd ACM Int. Conf. on Multimedia*, pages 789–792, 2014.

[4] R. Feris, R. Bobbitt, L. Brown, and S. Pankanti. Attribute based people search: Lessons learnt from a practical surveillance system. In *Proc. of 4th ACM Int. Conf. on Multimedia Retrieval*, pages 153–160, 2014.

[5] B. Golomb, D. Lawrence, and T. Sejnowski. Sexnet: A neural network identifies sex from human faces. In *Advances in Neural Information Processing Systems 3*, pages 572–577, 1990.

[6] B. Hernandez, G. Olague, R. Hammoud, L. Trujillo, and E. Romero. Visual learning of texture descriptors for facial expression recognition in thermal imagery. *Computer Vision and Image Understanding*, 106(2–3):258–269, 2007.

[7] S. Khamis, C. Kuo, V. Singh, V. Shet, and L. Davis. Joint learning for attribute-consistent person re-identification. In *Proc. of 13th European Conf. on Computer Vision*, pages 134–146, 2014.

[8] B. Kresnaraman, Y. Kawanishi, T. Takahashi, D. Deguchi, I. Ide, and H. Murase. Human wearable attribute recognition using decomposition of thermal infrared images. In *Proc. of 22nd Korea-Japan Joint Workshop on Frontiers of Computer Vision*, pages 123–127, 2016.

[9] B. Kresnaraman, Y. Kawanishi, T. Takahashi, D. Deguchi, I. Ide, and H. Murase. Human wearable attribute recognition using probability-map-based decomposition of thermal infrared images. *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, E100-A(3):854–864, 2017.

[10] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *Proc. of IEEE Int. Conf. on Computer Vision 2009*, pages 365–372, 2009.

[11] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Describable visual attributes for face verification and image search. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 33(10):1962–1977, 2011.

[12] R. Layne, T. Hospedales, and S. Gong. Person re-identification by attributes. In *Proc. of 23rd British Machine Vision Conf.*, pages 24.1–24.11, 2012.

[13] M. Lyons, J. Budynek, A. Plante, and S. Akamatsu. Classifying facial attributes using a 2-D Gabor wavelet representation and discriminant analysis. In *Proc. of 4th IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 202–207, 2000.

[14] G. Shakhnarovich, P. A. Viola, and B. Moghaddam. A unified learning framework for real time face detection and classification. In *Proc. of 5th IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 14–21, 2002.

[15] J. Thornton, J. Baran-Gale, D. Butler, M. Chan, and H. Zwahlen. Person attribute search for large-area video surveillance. In *Proc. of IEEE Int. Conf. on Technologies for Homeland Security*, pages 55–61, 2011.

[16] L. Trujillo, G. Olague, R. Hammoud, and B. Hernandez. Automatic features localization in thermal images for facial expression recognition. In *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition 2005 Workshops*, pages 14–21, 2005.

[17] D. A. Vaquero, R. S. Varis, D. Tran, L. Brown, A. Hampapur, and M. Turk. Attribute based people search in surveillance environment. In *Proc. of IEEE Computer Society Workshop on Applications of Computer Vision 2009*, pages 1–8, 2009.

[18] S. Wang, Z. Gao, S. He, M. He, and Q.Ji. Gender recognition from visible and thermal infrared facial images. *Multimedia Tools and Applications*, 75(14):8419–8442, 2016.