FCV2016, the 22nd Korea-Japan Joint Workshop on Frontiers of Computer Vision

Human Wearable Attribute Recognition using Decomposition of Thermal Infrared Images

Brahmastro Kresnaraman*, Yasutomo Kawanishi*, Daisuke Deguchi†, Tomokazu Takahashi‡, Yoshito Mekada§, Ichiro Ide* and Hiroshi Murase*
*Graduate School of Information Science, Nagoya University, Japan
Email: brahmastrok@murase.m.is.nagoya-u.ac.jp, {kawanishi, ide, murase}@is.nagoya-u.ac.jp
†Information Strategy Office, Nagoya University, Japan
Email: ddeguchi@nagoya-u.jp
‡Faculty of Economics and Information, Gifu Shotoku Gakuen University, Japan
Email: ttakahashi@gifu.shotoku.ac.jp
§Graduate School of Computer and Cognitive Science, Chukyo University, Japan
Email: y-mekada@sist.chukyo-u.ac.jp

***Abstract*—This paper addresses an attribute recognition problem in thermal images, specifically on worn objects such as hat and glasses. Although attribute recognition is a growing research field, there are not much work done in thermal infrared spectrum. In this spectrum, since illumination is not a problem, it could be a better option to be used in nighttime or poorly lit areas. The proposed method uses only the attribute information and excludes the unnecessary information for the recognition. To achieve this, we propose attribute recognition based on feature decomposition using Robust Principal Component Analysis (RPCA). An experiment to evaluate the capability of the proposed method was conducted on the dataset created for this research. The results show that the proposed method outperformed the method without decomposition by 14% in average with a maximum of 27% increase in a specific attribute.**

I. Introduction

Security surveillance systems play a vital role in the community nowadays. These systems may be able to prevent crimes in the vicinity and can even be used to help find and identify a criminal. Important keys in identifying or searching criminals are by looking at their characteristics. One of them is what they wear, and we define such information of a person as "wearable attributes". In certain places, a bank for example, the usage of mask, hat, and sunglasses might not be allowed. This research focuses on recognizing these wearable attributes.

A basic surveillance system employs a general consumer market camera, which works in the visible spectrum. Many ongoing image processing researches are conducted in visible spectrum, but illumination is often a big factor on the success of any recognition method used. If an area is not properly illuminated, the captured image might not have sufficient information to offer, which decreases the capability of the method. This makes surveillance in nighttime or in poorly lit areas a challenging task. In these situations, a thermal camera which captures images in thermal infrared spectrum can be a better option.

In thermal infrared spectrum, illumination is not a problem because the camera reproduces an image by capturing infrared radiation whose intensity depends on the temperature. With a thermal infrared camera, surveillance during nighttime or of poorly lit area is made possible. Figure 1 shows a sample image of a person in thermal spectrum. In the image, the person is presented with multiple wearable attributes (hat, glasses, and mask).

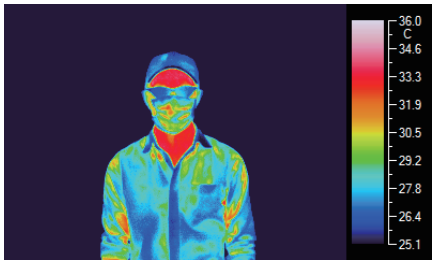

Fig. 1: Image example in thermal infrared spectrum.

The approach taken in this research is to use the thermal characteristics of wearable attributes in the thermal spectrum. As shown in Fig. 1, the size of each attribute is relatively small when compared to the size of the human body. Furthermore, intra-class variation of human body is large. Even in the thermal infrared image, it is still difficult to recognize these attributes. Some works on attribute recognition [7], [8], [13] make use of a smaller region of a person, i.e. face region, which is also divided further into even smaller regions.

In this research, we introduce attribute recognition based on decomposition using Robust Principal Component Analysis (RPCA). The decomposition process produces images with size identical to the original instead of dividing the image into smaller regions. The proposed method decomposes an image to two images; one with the wearable attribute information and the other with the human body without the wearable attribute. More details are provided in Section 3.

II. Related work

Attribute has a very broad meaning as it is defined as a trait or an element of someone/something. For humans, examples

(a) No attribute (b) Mask attribute (c) Backpack attribute

Fig. 2: Example of data with attribute as minority.

of attributes are age, gender, and race. Contrary to these non-wearable attributes, this research handles a specific attribute category which will be referred as wearable attributes such as hat and glasses. To elaborate, this research categorizes wearable attribute as an attribute that is worn by people. Unless specified otherwise, attribute in this research refers to wearable attributes.

A lot of works have been done related to the recognition of non-wearable attributes, as it encompasses the usefulness on various fields such as crime investigations and human computer interaction. There are works available in effort to classify expressions [2] and race/ethnicity [11].

For wearable attributes, the closest works are done by Kumar et al. [7], [8] that utilize various attributes for face verification. Other examples [5], [13] use images taken from surveillance cameras to find people based on certain attributes, e.g. red shirt.

All of the studies mentioned previously were conducted in the visible spectrum. In contrast, there are only few attribute recognition studies available in thermal infrared spectrum. Most of these works are solely focused on recognizing facial expression [6], [12], and a brief eyewear (glasses and sunglasses) detection experiment is done in [13]. To the extent of our knowledge, there are no other works that focus only on wearable attributes in thermal infrared spectrum.

III. Infrared Image Decomposition

The decomposition process adapted by this research is discussed in this section. First, the approach of this research is described in a more detailed fashion. The following subsection explains Robust Principal Component Analysis (RPCA) as the basis of the decomposition process of the proposed method.

A. Attribute Extraction by Decomposition

The decomposition process is where the proposed method takes advantage of the characteristics of the attributes in the thermal images. As mentioned previously, the size of an attribute is relatively small compared to the human body, and as a matter of fact, human bodies themselves have variations. This makes the recognition of the attributes relatively difficult. To tackle this problem, the proposed method decomposes an image and extracts the attributes from the human body region and discards the unnecessary information (including the variation of the human body) for the recognition.

The main idea behind the decomposition is as follows: Assume a collection of thermal infrared human images. When the majority of images have no attributes in them, the presence of an attribute in some images (minority) will be evaluated as "noise". Figure 2 portraits a data example of this idea. The mask attribute in Fig. 2(b) and backpack attribute in Fig. 2(c) are considered as noise in face and torso regions, respectively. Under this assumption, the attribute can be extracted by the decomposition which will be used for the recognition.

B. Robust Principal Component Analysis (RPCA)

As mentioned earlier, the proposed method utilizes Robust Principal Component Analysis (RPCA) to perform the decomposition. RPCA is a modification of the popular Principal Component Analysis (PCA), making it robust to corrupted or noisy observations in the data. This is motivated by the fact that PCA encounters problems on the existence of outliers and/or noisy observations. With RPCA, noisy images can be decomposed, resulting in two images with noise removed from it and the noise itself.

We considered that the decomposition capability of RPCA would benefit our research by considering wearable attributes as noise instead of part of the human characteristics. Thus, RPCA can decompose an image and use only the attribute information for recognition. Further details of RPCA are as follows.

Candes et al. [4] introduced an idealized version of RPCA, aiming to decompose a low-rank matrix L and a sparse matrix S from data M, as shown in the following:

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

where sparse matrix S represents the noisy part of the observation and its magnitude can be arbitrarily large. Similar to PCA, RPCA can be used to obtain a low-rank version of an image. However, the resulting sparse matrix that contains noise can also be utilized. In this research, the proposed method takes advantage of this decomposition capability.

Various techniques are available to achieve this decomposition, such as Principal Component Pursuit (PCP) [4], Stable

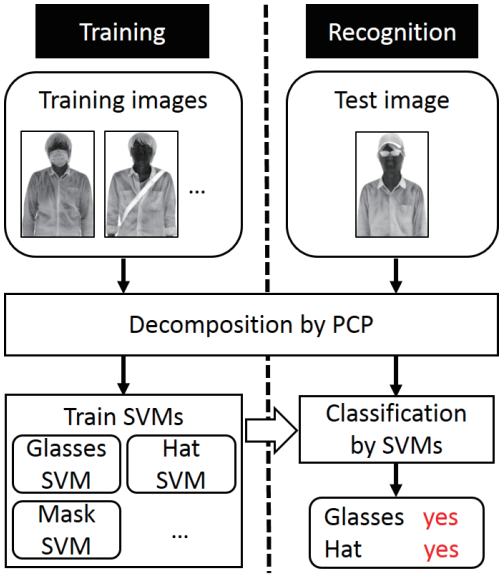

Fig. 3: Process flow.

(a) Original (b) Low-rank (c) Sparse

Fig. 4: Results of applying PCP to thermal images with attributes. Original images are inverted for visualization purpose.

PCP [15] and Local PCP [14]. Bouwmans and Zahzah [3] made a survey on RPCA, comparing some of the techniques aforementioned.

The proposed method chooses PCP as the RPCA technique. Given a data matrix M where observations are represented as a column vector, the PCP seeks to solve the following optimization problem:

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1 \text{ s. t. } L + S = M, \tag{2}$$

where $\|.\|_*$ is the nuclear norm, which is the sum of the singular value given a matrix, and $\|.\|_1$ is the l_1-norm where the matrix is treated as a vector. λ is an arbitrary balance parameter. As a baseline, λ is set as:

$$\lambda = \frac{1}{\sqrt{\max(m,n)}}, \tag{3}$$

where m and n represent the numbers of row and column of matrix M, respectively. Usually, λ does not need to be fine-tuned. The only exception is when prior knowledge is available. Under these minimal assumptions, PCP is able to obtain the low-rank and the sparse matrices given a data matrix.

IV. Wearable Attribute Recognition Framework

The process flow of the proposed method is shown in Fig. 3. It is important to note that during the entire process, the thermal infrared images use the "hotblack" color scheme, which means that it is monochrome and the hotter the temperature of an object, the closer the pixel value is to zero. This section explains the decomposition process, followed by the training and recognition by SVM.

A. Decomposition by PCP

For PCP to be able to decompose an image with attributes in it, certain condition needs to be fulfilled. The condition is that a data matrix needs to have a high similarity between the observations. In other words, the attributes can be separated when they are the minority in the data. Therefore, the proposed method controls the data matrix that will be decomposed by the PCP.

The technique to successfully extract attributes from the decomposition process is by creating the data matrix that satisfies the condition mentioned previously. The data matrices are constructed by taking images with no attributes in the image as the base data and one image that will be trained/tested. This arrangement warrants the attributes to be considered as noise by the PCP if it exists in the image because it is counted as the minority. Then, PCP could decompose the image with ease and provides the extracted wearable attributes in the sparse matrix. Figure 4 shows an example of applying PCP with this technique. As shown in Fig. 4(c), the sparse entries show that the PCP performs relatively well in extracting the attributes. However, some details of the hair and clothes are also extracted.

The extraction process in the training phase is as follows. First, from the available dataset, features are calculated and represented as $D = (\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_N)$. In this research, dense variation of Scale Invariant Feature Transform (SIFT [10]) introduced by Liu et al. [9] is selected as the feature. Sampling SIFT densely makes it stable and invariant to key point detection results, which can include the human shape information.

Features of people with no wearable attributes are chosen as base data $B = (\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_P)$ with P observations. This data is used in both training and recognition phases. For training, the data is represented as $W = (\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_Q)$. It is important to note that base data and training data are not

(a) No attribute (b) Mask attribute (c) Glasses, mask and hat attributes (d) Backpack attribute

Fig. 5: Image examples from the dataset.

intersected. The PCP will solve Eq. (4).

$$\begin{bmatrix} B & \mathbf{w}_q \end{bmatrix} = M^q = L^q + S^q, \tag{4}$$

where

$$S^q = \begin{bmatrix} S^q_B & \mathbf{x}_q \end{bmatrix}. \tag{5}$$

In Eq. (5), $\mathbf{x}_q$ is an entry of sparse matrix S^w that corresponds with $\mathbf{w}_q$. After PCP is performed Q times, the training data is arranged as $X = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_Q)$ and further used for training by SVM.

In the recognition phase, a thermal infrared data $\mathbf{t}$ is chosen from D as test data. Equation (6) shows the optimization problem that is needed to be solved in this phase.

$$\begin{bmatrix} B & \mathbf{t} \end{bmatrix} = M^t = L^t + S^t, \tag{6}$$

where

$$S^t = \begin{bmatrix} S^t_B & \mathbf{x}_t \end{bmatrix}. \tag{7}$$

Similar to the training phase, in Eq. (7), $\mathbf{y}$ is an entry of sparse matrix S^t that corresponds with $\mathbf{t}$. $\mathbf{y}$ is then used for recognition by SVM in the next step.

B. Training and Recognition by SVM

Support Vector Machine (SVM) is selected as the recognition method. One SVM is trained for each attribute, creating multiple SVMs as many as the attributes. For each SVM, the training data $X = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_Q)$ is distributed to positives and negatives. As an example, to train SVM of the "glasses" attribute, the positive training data are images with "glasses" in it, regardless of other attributes. The negatives are features of images where "glasses" are not present.

In the recognition phase, test data $\mathbf{y}$ is used as an input to all attribute-specific SVMs. The proposed method provides the results of the recognition by these SVMs as the final output.

V. Experiments

An experiment was conducted to evaluate the capability of the proposed method described in the previous section. The first subsection introduces the dataset used for the experiment, including the attributes used in this research. The setup of the experiment is elaborated in the succeeding subsection. At the end, results of the experiment and its discussion are provided.

TABLE I: Distribution of the seven wearable attributes in the dataset.

Attributes	# of images
No attribute	28
Glasses	168
Mask	168
Hat	112
Helmet	80
Hoodie	40
Shoulder bag	48
Backpack	40

A. Dataset

For the purpose of this research, a new dataset was created. Currently, it contains a total of 408 frontal thermal infrared images from fourteen persons with up to seven different wearable attributes per person. The seven attributes in this dataset are glasses, surgical masks or simply masks, baseball cap (hat), safety helmet, hoodie, shoulder bag, and backpack. The number of images available for each attribute can be seen in Table I. It needs to be noted that one image may contain more than one wearable attributes.

The camera used for data capture was TVS-500EX [1]. The wavelength that can be captured by the camera ranges from 8 to 14 μm. The images were taken indoors at room temperature (around 22–23° Celsius). The size of the images was 320 x 240 pixels. For this dataset, the camera was set to capture infrared signal which temperature ranging from 25–36° Celsius with the "hotblack" scheme. Figure 5 shows some of the captured images. We cropped human regions manually when building the dataset. The average size of the human body image is 140 x 204 pixels.

B. Experimental setup

The aim of the experiment performed in this research is to compare the capability of the proposed method with that of using SVM directly without the decomposition process to recognize the attributes. The experiment is performed in a k-fold cross-validation manner, where k equals to the number of the people in the dataset ($k = 14$). This means the tested

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

Methods	Attributes							Average
	Glasses	Mask	Hat	Helmet	Hoodie	Shoulder bag	Backpack	
Without decomposition (comparative method)	0.75	0.92	0.49	0.49	0.65	0.63	0.79	0.67
Proposed method	0.89	0.95	0.72	0.76	0.71	0.72	0.89	0.81

person is not included in the training data. The results are evaluated by F-score taken averagely from the cross validation.

C. Results

Results of the experiment can be seen in Table II. In average, the proposed method outperformed the comparative method (without decomposition) by 14%. The increase of performance for each attribute are quite apparent except on the mask attribute. The reason for this might be due to the big difference between wearing a mask and not wearing a mask. Even though relatively good results were also yielded for other attributes such as glasses and backpack using SVM directly, the increase after the decomposition by PCP is noteworthy. The most significant improvement can be seen in hat and helmet attributes, increased by 23% and 27%, respectively.

VI. Conclusion

This paper addressed the attribute recognition problem in thermal infrared images. While attribute recognition is continuously being researched, most of the works available were conducted in visible spectrum. The thermal infrared spectrum is fundamentally different from the visible spectrum and it has its own advantages when compared between the two.

The purpose of this research was to recognize wearable attributes on a person in the thermal infrared image. To achieve this, the proposed method utilizes the characteristics of the wearable attributes in the thermal infrared spectrum. The proposed method decomposes an image, extracts the attribute information, and use it for the recognition. The proposed method employed Robust Principal Component Analysis (RPCA) as the basis to perform the decomposition.

The results showed that the proposed method outperformed the non-decomposition method in average by 14%. The most significant performance was obtained in hat and helmet attributes, 23% and 27%, respectively. This proves the decomposition process helps the recognition.

For further research, it is important to increase the size and the variation of the dataset. Adding angle information as one would encounter in a real-world situation, should also be considered. Furthermore, it will also be beneficial to utilize images from both thermal infrared and visible spectra.

Acknowledgment

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

References

[1] Thermal video system advanced thermo TVS-500EX. http://www.infrared.avio.co.jp/en/products/ir-thermo/lineup/tvs-500ex/spec.html. Accessed June 25 2015.

[2] 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 Patter Recognition 2003 Workshops*, volume 5, pages 53–59, 2003.

[3] T. Bouwmans and E. H. Zahzah. Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance. *Computer Vision and Image Understanding*, 122:22–34, 2014.

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

[5] 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.

[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] 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.

[8] 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.

[9] C. Liu, J. Yuen, and A. Torralba. SIFT flow: Dense correspondence across scenes and its applications. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 33(5):978–994, 2011.

[10] D. G. Lowe. Object recognition from local scale-invariant features. In *Proc. of 7th IEEE Int. Conf. on Computer Vision (ICCV)*, volume 2, pages 1150–1157, 1999.

[11] 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.

[12] 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.

[13] 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.

[14] B. Wohlberg, R. Chartrand, and J. Theiler. Local principal component pursuit for nonlinear datasets. In *Proc. of 37th IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pages 3925–3928, 2012.

[15] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *Advances in Neural Information Processing Systems 22*, pages 2080–2088, 2009.