ICIC Express Letters
Volume **10**, Number **3**, March **2016**

ICIC International ©2016 ISSN 1881-803X
pp. 745–749

TRACKING OF PERSON FROM MOVING IMAGES TAKEN BY SPARSELY DISTRIBUTED CAMERAS USING MSER AND GEI

Koutarou Takemura[1] and Shigeyoshi Nakajima[2]

[1]ITEC Hankyu Hanshin Co., Ltd.
1-1-31 Ebie, Fukusima-ku, Osaka City 553-0001, Japan
k.takemura426@gmail.com

[2]Graduate School of Engineering
Osaka City University
3-3-138, Sugimoto Cho, Sumiyoshi Ku, Osaka City 558-8585, Japan
nakajima@info.eng.osaka-cu.ac.jp

Received August 2015; accepted October 2015

Abstract. *Needs for surveillance systems increase in these days because of terrorisms, injuries and others. There are many surveillance cameras indoors and outdoors. Most of them are distributed sparsely. We propose a new method to track the same person using sparsely distributed cameras. Tracking of a person using sparsely distributed cameras is more difficult than a case of densely distributed cameras. There was a method using GEI (gait energy image). We combined MSER (maximally stable extremal regions) and GEI. The result of combination was better than a simple GEI case. However, further improvement is needed for practical scene uses.*

Keywords: GEI, MSER, Surveillance camera, Densely distributed cameras, Sparsely distributed cameras, Pedestrian, Identification

1. **Introduction.** There are many surveillance cameras in our day-to-day life indoors and outdoors. Needs for surveillance systems increase in these days because of terrorisms, injuries and others. A person watches some monitors of surveillance cameras in an ordinary surveillance system. However, it costs much human power. So an automatic surveillance system will be very useful.

GEI [1-3] is an accumulated moving image to a single image. It extracts features of a pedestrian's behavior even if his (her) face or other body parts are not separated from a body area in a moving image. MSER [4] is a criterion of a threshold level of a binarization. MSER is expected to improve the GEI ability.

Viola and Jones [5] proposed a fast method to detect face areas from images of persons. Zhou and Hoang [6] proposed a method to track a person and detect face regions.

Some works about identification and tracking of a person and people were developed for marketing instead of security. Zhan et al. [7] analyzed behavior of people in clouded building.

There are some works about person re-identification. Two image shots of sparsely distributed cameras may be shots of the same person or may be shots of different persons. There are two groups of works about person re-identification. Some groups do not concern temporal changes of an appearance of a person. Other groups concern temporal changes.

The former takes their cue from differences of clothes or sizes of persons or shapes of persons. Gray and Tao [8] used many filters to extract features of clothes of persons. They also employed covariance matrix and boosting algorithms.

Chourasiya et al. [1] combined PCA (principal component analysis) and GEI. The combination of PCA and GEI increased recognition rate versus PCA only. Jadav et al. [2] employed GEI. Their method compensated lost gait contour sequences from other extracted gait contours using GEI. Ali et al. [3] extracted features of gaits of humans from analysis of GEI using Radon transform. They used ordinary image data to make GEI data. Our method increases ability of GEI using MSER data of images. GEI is available with a small area of a pedestrian in an image and low resolution of a camera.

We tried to improve GEI using MSER [3].

The rest of this paper is organized as follows. Section 2 outlines the proposed method to measure similarity of moving images. Section 3 presents experimental result and a consideration. Section 4 shows a conclusion.

2. **Similarity of Moving Images.** The proposed method measures similarity of moving images using MSER image as an input of GEI and PCA to measure the similarity of moving image for pedestrian identification. GEI accumulates binary images from a moving image to one gray-level image. However, GEI does not limit a binarization process. We selected MSER as binarization. We expect that MSER is better for GEI than another binarization. PCA is used to make the method robust.

2.1. **GEI.** An ordinary gait of a person seems a set of periodical rotations of similar behavior. The behaviors of the same person are similar. However, behaviors of different persons are not similar. There are differences of motions of four extremities and body inclinations. GEI is defined as Equation (1).

$$G(x,y) = \frac{1}{N}\sum_{t=1}^{N} B_t(x,y) \tag{1}$$

A gray scale image becomes a silhouette after binarization. The silhouette is normalized. The normalized image is $B_t(x,y)$. The suffix t indicates a frame number. The 2D vector (x,y) indicates a position of a pixel. The center $B_t(x,y)$ is the gravity center of an upper body because the motion of a lower body is larger than one of an upper body. GEI $G(x,y)$ is defined as Equation (1).

The number N is total number of frames. Figure 1 shows a frame of a gait of a person. Figure 2 shows a normalized binary image $B_t(x,y)$. Figure 3 shows a GEI $G(x,y)$.

FIGURE 1. Gait of person

FIGURE 2. Normalized binary image $B_t(x,y)$

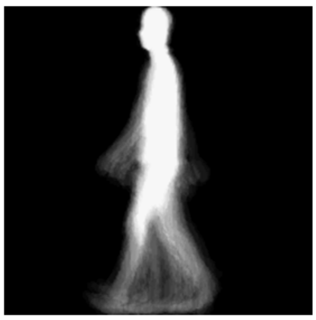

FIGURE 3. GEI $G(x,y)$

2.2. **MSER.** A region of a person or an object ordinarily has a contour in an image. There are many methods to detect contour lines from images such as simple threshold method, and zero-crossing. Matas et al. [4] proposed MSER. Forssen [9] extended MSER to color as MSER as Equation (2).

$$I(p) > I(r) \text{ or } I(p) < I(r) \quad \text{for } \forall p \in Q,\ \forall r \in \partial Q \tag{2}$$

The proposed method uses original MSER image as an input of GEI. MSER is applied to a monochrome image such as 8bit gray scale. MSER outputs an extremal region Q_i as a set of pixels p_j which are connected to other pixels p_k ($p_k \in Q_i$) with 4-neighbor

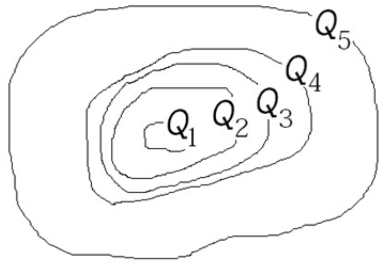

FIGURE 4. Sequence of regions Q_i

or 8-neighbor connection. A region ∂Q is an outer region boundary of a region Q. The boundary ∂Q of Q is the set of pixels being adjacent to at least one pixel of Q but not belonging to Q. ∂Q and Q also satisfy Equation (1).

$I(p)$ is a gray scale value of a pixel p. Let $Q_1, \ldots, Q_{i-1}, Q_i, \ldots$ be a sequence of nested extremal regions, i.e., $Q_i \subset Q_{i+1}$ like Figure 4. A function $q(i)$ is defined as Equation (3).

$$q(i) = \frac{|Q_{i+\Delta}| - |Q_{i-\Delta}|}{|Q_i|} \quad \text{where } |Q| \text{ is a cardinality of } Q. \tag{3}$$

Δ is one of constant gray scale value differences as a parameter of the method. Extremal region Q_{i^*} is maximally stable iff $q(i)$ has a local minimum at $i = i^*$. The numerator of the fraction of Equation (2) indicates the difference of number of pixels from $Q_{i-\Delta}$ to $Q_{i+\Delta}$. The difference is normalized by the denominator which is area of Q_i. So MSER indicates a local minimum of relative change rate of an area of a region.

2.3. **PCA.** An MSER image can become an input of GEI. Then the GEI output image is translated to a component $\boldsymbol{f}$ by PCA (principal component analysis) as Hotelling [10]. The moving image m_j of j-th person at time t is translated to $\boldsymbol{f}_j$. $\boldsymbol{f}_j$ is an n-dimensional vector. $D(\boldsymbol{f}_j, \boldsymbol{f}_k)$ is a distance between components $\boldsymbol{f}_j$ and $\boldsymbol{f}_k$ shown in Equation (4). The number k means another person.

$$D(\boldsymbol{f}_j, \boldsymbol{f}_k) = |(\boldsymbol{f}_j - \boldsymbol{f}_k)| \quad \text{where } |\,| \text{ is a Euclid distance of a vector.} \tag{4}$$

f'_j means a component for a movie m'_j of the k-th person taken at another time t'. If $j = k$ movies m_j and m'_k belong to the same person. If $j \neq k$ movies m_j and m'_k belong to different persons. If an Inequality (5) satisfies for all k $(k \neq j)$ we can select the movie of same person j from movies of all persons using $\boldsymbol{f}_j$.

$$D\left(\boldsymbol{f}_j, \boldsymbol{f}'_j\right) < D\left(\boldsymbol{f}_j, \boldsymbol{f}'_k\right) \quad (k \neq j) \tag{5}$$

Figure 5 shows GEI reconstructed from a component of the result of PCA such as $\boldsymbol{f}$.

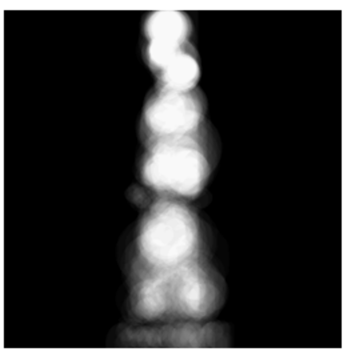

FIGURE 5. GEI of component $\boldsymbol{f}$ after PCA

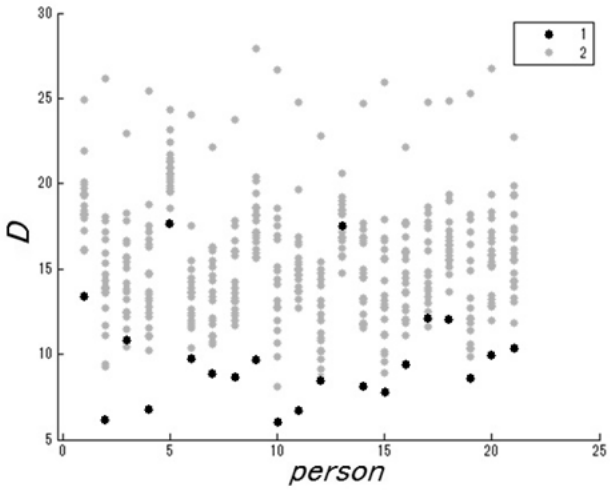

FIGURE 6. D of GEI without MSER

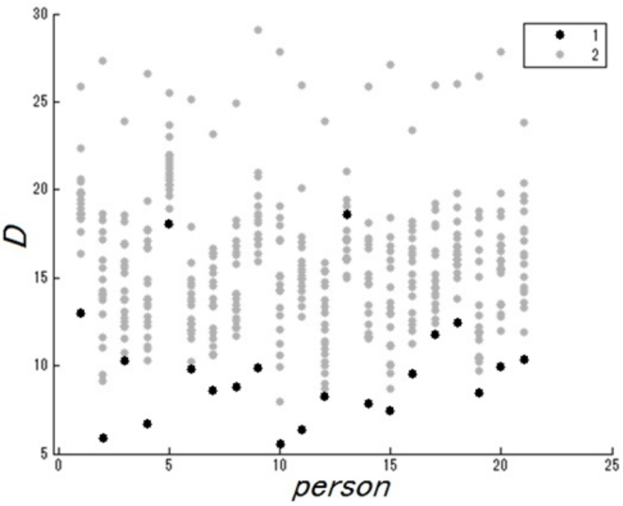

FIGURE 7. D of MSER + GEI

3. **Experimental Result.** We used a video Camera SONY HDR-CX550V. Movies are HD and 30 fps and are downsized to 15 fps and 640×480 pixel size. The back ground was a blue plastic sheet.

We acquired more than two shot movies of 21 persons.

Figure 6 shows the result of $D(\boldsymbol{f}_j, \boldsymbol{f}'_k)$ with simple GEI. There are 18 cases which satisfy Inequality (5). The threshold level of binarisation is selected as to maximize a mean of absolute values of Sobel filtering results on a contour with the level.

Figure 7 shows the result of $D(\boldsymbol{f}_j, \boldsymbol{f}'_k)$ with MSER + GEI. There are 19 cases which satisfy Inequality (5).

However, we could not find a constant λ which satisfy Inequality (5) for all j and k ($\neq j$).

$$D\left(\boldsymbol{f}_j, \boldsymbol{f}'_j\right) < \lambda < D\left(\boldsymbol{f}_j, \boldsymbol{f}'_k\right) \quad (j \neq k) \tag{6}$$

Table 1 shows the statistical values of $D(\boldsymbol{f}_j, \boldsymbol{f}'_k)$ $(j \neq k)$ of GEI and MSER + GEI.

TABLE 1. $D(\boldsymbol{f}_j, \boldsymbol{f}'_k)$ $(j \neq k)$

	mean	variance	S.D.	S.E.
MSER + GEI	18.0	9.68	3.11	1.75
GEI	15.6	5.83	2.42	0.51

The values of MSER + GEI were better than GEI with Sobel maximum level.

The data of 13^{th} person in Figure 6 and Figure 7 seem to violate Inequality (5) significantly. We are afraid that there were some errors in video recording of the 13^{th} person.

We could find a threshold value λ_j of Equation (7) which is dependent to person j with 19 cases in Figure 7 as the result of MESR + GEI. If we find λ of Equation (8) which is independent to a person we can decide whether two movies belong to the same person or different people. However, we could not find λ of Equation (8). So the result shows that the proposed method is not enough for a practical surveillance scene.

$$D\left(\boldsymbol{f}_j, \boldsymbol{f}'_j\right) < \lambda_j < D\left(\boldsymbol{f}_j, \boldsymbol{f}'_k\right) \quad (j \neq k) \tag{7}$$

$$D\left(\boldsymbol{f}_j, \boldsymbol{f}'_j\right) < \lambda < D\left(\boldsymbol{f}_j, \boldsymbol{f}'_k\right) \quad (j \neq k) \tag{8}$$

4. **Conclusion.** We proposed a method to calculate distance of movies using MSER and GEI. The proposed method is better than single GEI with a popular binarisation method. However, it is not enough for a practical surveillance scene. We will improve our method in future to combine our method with PCA [1].

REFERENCES

[1] D. R. Chourasiya, V. Richhariya and V. Rajpoot, Human recognition through gait energy image and principal component analysis using TUMIITKP database, *International Journal of Computer Science and Information Technology and Security*, vol.3, no.6. pp.409-416, 2013.

[2] V. R. Jadav, G. Lachiram and N. Swathi, Gait recognition system for human identification using temporal preserving information, *International Journal of Computer Science and Electronics Engineering UK*, vol.4, no.3. pp.1-11, 2014.

[3] H. Ali, J. Dargham, C. Ali and E. G. Moung, Gait recognition using gait energy image, *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol.4, no.3, pp.141-152, 2011.

[4] J. Matas, O. Chum, M. Urban and T. Pajdla, Robust wide baseline stereo from maximally stable extremal regions, *Proc. of British Machine Vision Conference*, Cardiff, UK, pp.384-396, 2002.

[5] P. Viola and M. Jones, Rapid object detection using a boosted cascade of simple features, *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Cambridge, MA, USA, pp.I-511-I-518, 2001.

[6] J. Zhou and J. Hoang, Real time robust human detection and tracking system, *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition – Workshops*, San Diego, USA, p.149, 2005.

[7] B. Zhan, D. N. Monekosso, P. Remagnino, S. A. Velastin and L.-Q. Xu, Crowd analysis: A survey, *Machine Vision and Applications*, vol.19, nos.5-6, pp.345-357, 2008.

[8] D. Gray and H. Tao, Viewpoint invariant pedestrian recognition with an ensemble of localized features, *The 10th European Conference on Computer Vision*, Marseille, France, 2008.

[9] P.-E. Forssen, Maximally stable colour regions for recognition and matching, *Proc. of Computer Vision and Pattern Recognition*, Minneapolis, Minnesota, USA, pp.1-8, 2007.

[10] H. Hotelling, Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology*, vol.24, pp.417-441, 1933.