Using Motion Blur to Recognize Hand Gestures in Low-light Scenes

Daisuke Sugimura, Yusuke Yasukawa and Takayuki Hamamoto
Department of Electrical Engineering, Tokyo University of Science,
6-3-1 Niijuku, Katsushika-ku, 125-8585, Tokyo, Japan

Keywords: Hand Gesture Recognition, Low-light Scenes, Motion Blur, Temporal Integration.

Abstract: We propose a method for recognizing hand gestures in low-light scenes. In such scenes, hand gesture images are significantly deteriorated because of heavy noise; therefore, previous methods may not work well. In this study, we exploit a single color image constructed by temporally integrating a hand gesture sequence. In general, the temporal integration of images improves the signal-to-noise (S/N) ratio; it enables us to capture sufficient appearance information of the hand gesture sequence. The key idea of this study is to exploit a motion blur, which is produced when integrating a hand gesture sequence temporally. The direction and the magnitude of motion blur are discriminative characteristics that can be used for differentiating hand gestures. In order to extract these features of motion blur, we analyze the gradient intensity and the color distributions of a single motion-blurred image. In particular, we encode such image features to self-similarity maps, which capture pairwise statistics of spatially localized features within a single image. The use of self-similarity maps allows us to represent invariant characteristics to the individual variations in the same hand gestures. Using self-similarity maps, we construct a classifier for hand gesture recognition. Our experiments demonstrate the effectiveness of the proposed method.

1 INTRODUCTION

Techniques for hand gesture recognition have been well studied and applied to various applications such as human-computer interaction and interactive games (Pavlovic et al., 1997; Freeman and Weissman, 1996; Wachs et al., 2011; Weaver et al., 1998; Lian et al., 2014). Early studies on hand gesture recognition used hand groves (Iwai et al., 1996) or markers (J.Davis and M, 1994; Cipolla et al., 1993) to measure the hand motions of users. However, these approaches have a limitation that the users' hand motions are unlikely to be natural because they have to wear certain specific devices.

In recent years, the development of motion analysis techniques using camera images has highly contributed to the advances in the field of hand gesture recognition. In fact, many methods have been proposed to improve the performance of hand gesture recognition (Pfister et al., 2014; Yamato et al., 1992; Stamer and Pentland, 1995; Freeman and Roth, 1995; Shen et al., 2012; Ikizler-Cinbis and Sclaroff, 2010; Liu and Shao, 2013; Tang et al., 2014; Davis, 2001; Scocanner et al., 2007; Bregonzio et al., 2009; Niebles et al., 2008). Early studies investigated approaches to recognize spatio-temporal hand gesture motion modeling by using a hidden Markov model (HMM) (Yamato et al., 1992; Stamer and Pentland, 1995). In contrast, recent techniques have exploited local feature descriptors to represent hand gesture motions (Shen et al., 2012; Scocanner et al., 2007; Bregonzio et al., 2009; Niebles et al., 2008). Shen *et al.* proposed a spatio-temporal descriptor constructed by analyzing the motion divergence fields of hand gestures (Shen et al., 2012). Scocanner *et al.* extended a SIFT descriptor to represent spatio-temporal motions (Scocanner et al., 2007). Since these methods have scale- and position-invariant characteristics, they have enabled to achieve high recognition performance.

Liu *et al.* pointed out that the use of local image descriptor representations cannot be used for easily modeling precise hand gesture motions because of the loss of the structural relationships between the extracted local features (Liu and Shao, 2013). To address this, they proposed a hand gesture recognition system by automatically synthesizing spatio-temporal descriptors using genetic programming.

On the other hand, with the developments in the field of active sensors, three-dimensional (3D) hand gesture recognition systems have been studied as well. Tang *et al.* proposed a method for a 3D hand posture estimation by using a single depth im-

Sugimura, D., Yasukawa, Y. and Hamamoto, T.
Using Motion Blur to Recognize Hand Gestures in Low-light Scenes.
DOI: 10.5220/0005673603080316
In *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2016) - Volume 4: VISAPP*, pages 308-316
ISBN: 978-989-758-175-5

Figure 1: Examples of hand gestures in a low-light scene. It is obviously difficult to extract the necessary appearance information of a hand gesture. The objective of this study is to recognize hand gestures in low-light scenes.

Figure 2: Examples of the temporal integration of the low-light hand gesture sequence shown in Figure 1. We can see that sufficient information is captured while the motion blur is generated. Further, we can see that both the direction and the magnitude of the motion blur represent discriminative characteristics to differentiate hand gestures. The key idea of this study is to exploit this motion blur for hand gesture recognition.

age (Tang et al., 2014). Moreover, there are previous studies using KINECT (Ren et al., 2013) or Leap Motion (Marin et al., 2014).

In these previous studies, however, the researchers have implicitly assumed a bright lighting condition to precisely capture the hand gestures. In other words, hand gesture images should be taken at a high S/N ratio. In low-light scenes, as illustrated in Figure 1, the S/N ratio of the captured images degrades because heavy noise is likely to be imposed on low-light images. In such case, previous methods are ineffective at recognizing hand gestures.

In order to address such problems caused by low-light scenes, we propose a novel hand gesture recognition method by using a single color image that is constructed by integrating a hand gesture sequence temporally. In general, the temporal integration of images improves the S/N ratio; thus, it enables us to acquire sufficient appearance and motion information of a hand gesture even in low-light scenes.

However, the temporal integration of a hand gesture produces motion blur. In general, motion blur has been widely considered to be one of the annoying noises that degrade performance in various computer vision or image processing techniques. In contrast, in this study, we make effective use of motion blur to recognize hand gestures. Figure 2 shows examples of the integrated gesture image containing motion blur. We consider that both the direction and the magnitude of motion blur represent discriminative characteristics that can be used for differentiating hand gestures.

In order to extract both the direction and the magnitude of the motion blur of a hand gesture, we analyze the gradient intensity distributions of the image as well as color distributions. In fact, an analysis of the gradient intensity distribution of a motion-blurred image plays an important role in the estimation of the moving directions of objects, as reported in (Lin et al., 2011). Furthermore, the color distribution of the integrated gesture image is considered to be helpful information for measuring the magnitude of the motion blur. As shown in Figure 2, motion-blurred regions can be viewed as a mixture of the hand and background components. The magnitude of the motion blur is likely to be large when the skin color intensity (hand motion region) is low, and vice versa. This implies that the analysis of color distributions allows us to estimate the magnitude of the motion blur.

By using both the gradient intensity and the color distributions, we compute a self-similarity map (Walk et al., 2010), which encodes pairwise statistics of spatially localized features by a similarity within a single image. Because a self-similarity map can represent spatial relationships via a similarity, it can assign invariant characteristics to the individual variations in gesture motions, such as differences in lighting conditions and gesture speeds. We exploit the computed self-similarity maps as training data for constructing a hand gesture recognition system.

On the other hand, a single motion-blurred image can also be obtained by taking the image with a long-exposure time. However, it is hard to ensure that the entire hand gesture can be observed within such exposure time. In other words, it is difficult for cameras to find the times when the hand gesture starts and ends. In contrast, the temporal integration allows us to estimate both the start and the end timings of a hand gesture by analyzing the captured images. Thus, we utilize the temporal integration scheme instead of using a single motion-blurred image taken with a long-exposure time. Although the temporal integration makes motion blur discretized as shown in Figure 2, it is enough to obtain both the direction and the magnitude of the motion blur of a hand gesture.

The main contribution of this study is as follows. We exploit a single color image obtained by temporally integrating the hand gesture sequence to recognize hand gestures in low-light scenes. In particular, we make effective use of the motion blur included in the integrated gesture image. To the best of our knowledge, this study is the first to use motion blur in the context of hand gesture recognition. Unlike

the previous studies that implicitly assumed a bright lighting condition, the proposed method enables to perform robust hand gesture recognition under low-light conditions.

2 PROPOSED METHOD

The objective of this study is to construct a hand gesture recognition system for low-light scenes where the previous methods cannot easily process effectively. The proposed method consists of three steps: temporal hand gesture integration, self-similarity map construction, and classifier construction, as illustrated in Figure 3.

2.1 Temporal Hand Gesture Integration

To acquire hand gesture motions with a high S/N ratio in low-light scenes, we temporally integrate the hand gesture sequence $\{I_t\}_{t=1,...,T}$, where I_t denotes the gesture image at t-th frame, into a single image. We then exploit the motion blur included in the integrated image to construct a hand gesture recognition system.

In the proposed method, it is important to ensure that the motion blur has consistent characteristics for each hand gesture. However, the motion blur is likely to vary because of the variations in the gesture speed and timing. In order to compensate for such variations in the motion blur, we estimate the temporal integration time (i.e., the start and the end times of a hand gesture). We then integrate the hand gesture images between the start and the end times.

2.1.1 Integration Time Estimation

In this step, we estimate the times when a gesture motion starts and ends, t_s and t_e, by using an amount of movements of hand gesture.

We first divide an image into B local image patches. In each patch, we compute a motion vector $\mathbf{u}_t^i$ (i denotes the patch index) by using a pyramidal KLT method (Lucas and Kanade, 1981) for $\{I_t\}$. By using the sum of all the motion vectors in each frame, i.e., $\mathbf{v}_t = \sum_i \mathbf{u}_t^i$, we determine t_s and t_e as

$$t_\mathrm{s} = \min_{|\mathbf{v}_t|>\delta |\mathbf{v}_{\max}|} \{1,...,T\} , \quad (1)$$

$$t_\mathrm{e} = \max_{|\mathbf{v}_t|>\delta |\mathbf{v}_{\max}|} \{1,...,T\} , \quad (2)$$

where $\mathbf{v}_{\max}$ denotes the maximum value of a set $V = \{\mathbf{v}_t\}_{t=1,...,T}$ and δ represents a control parameter.

2.1.2 Gesture Image Integration

We integrate a hand gesture sequence between t_s and t_e. In general, motion blur is likely to be produced when the gesture speed is high; thus, gesture frames that have large movements are more important than the others. On the basis of this fact, we incorporate a weight ω_t computed using V for each gesture frame when integrating $\{I_t\}_{t=t_\mathrm{s},...,t_\mathrm{e}}$. We set a high weight value for images with large movements and a low weight for images with short movements. Therefore, we represent the weight as $\omega_t = |\mathbf{v}_t|/|\mathbf{v}_{\max}|$.

By using t_s, t_e, and ω_t, we obtain a single integrated image I_intg by the following weighted integration:

$$I_\mathrm{intg} = \frac{1}{t_\mathrm{e} - t_\mathrm{s}} \sum_{t=t_\mathrm{s}}^{t_\mathrm{e}} \omega_t I_t \ . \quad (3)$$

2.2 Self-similarity Map Construction

The direction and the magnitude of motion blur are helpful characteristics for distinguishing various hand gestures. In order to extract these features from a motion-blurred image I_intg, we analyze the gradient intensity and the color distributions of I_intg.

However, hand gesture motions are likely to vary depending on the user's individuality or the lighting conditions, even though the same hand gestures are performed. Thus, the direct use of the gradient intensity and the color distributions of I_intg leads to the performance degradation of hand gesture recognition.

To address this problem, we construct self-similarity maps (Walk et al., 2010) by using the gradient intensity and the color distributions. A self-similarity map captures the pairwise statistics of spatially localized features within a single image. In particular, a self-similarity map encodes image features via a similarity to represent invariant characteristics to the variations in the same hand gestures. Therefore, the use of a self-similarity map is effective for robust hand gesture recognition. In the proposed method, we construct two self-similarity maps on the basis of the gradient intensity and the color distributions.

2.2.1 Self-similarity Map using Gradient Intensity Distributions

A gradient intensity distribution is useful for estimating the gesture motion direction from a motion-blurred image as reported in (Lin et al., 2011). We exploit the gradient intensity distribution via a self-similarity map representation.

The steps for self-similarity map construction are as follows. We divide I_intg into B local image patches.

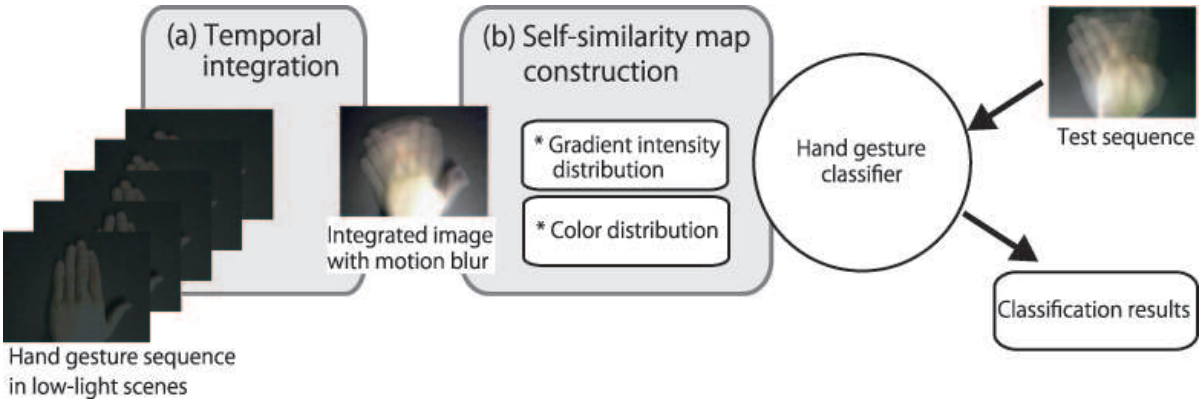


Figure 3: Overview of the proposed method. (a) Temporal integration of a hand gesture sequence. We temporally integrate a hand gesture sequence to acquire hand gesture motion with a high S/N ratio. We obtain a single gesture image containing motion blur. (b) Self-similarity map construction. We analyze the direction and the magnitude of the motion blur on the basis of the gradient intensity and the color distributions. We then construct self-similarity maps by using these image features. We finally build a hand gesture classifier by using the computed self-similarity maps.

In each local image patch, we make an intensity histogram of the gradient image of I_{intg}. We then obtain the similarity $s^{m,n}_{\text{grad}}$ between patch m and patch n by computing a Bhattacharyya distance of the normalized gradient intensity histograms, $\mathbf{g}^m$ and $\mathbf{g}^n$, as

$$s^{m,n}_{\text{grad}} = \sqrt{1 - \sum_{k=1}^{K} \mathbf{g}^m_k \mathbf{g}^n_k}\ , \tag{4}$$

where $\mathbf{g}^m_k$ denotes the frequency in the k-th bin of $\mathbf{g}^m$, and K represents the number of bins.

Finally, we obtain the self-similarity map of the gradient intensity distribution S_{grad} by computing $s^{m,n}_{\text{grad}}$ for all the pairs of image patches:

$$S_{\text{grad}} = [s^{1,2}_{\text{grad}}, ..., s^{1,B}_{\text{grad}}, ..., s^{B-1,B}_{\text{grad}}]\ . \tag{5}$$

Figure 4 shows an example of the computed $s^{m',n}_{\text{grad}}$ (m' denotes the patch of interest).

2.2.2 Self-similarity Map using Color Distributions

Analyzing color distributions of an image will be helpful for measuring the intensity of motion blur. As shown in Figure 2, a mixture of the hand and background components is captured in the motion-blurred regions. This suggests that the intensity of the motion blur is likely to be large when the skin color intensity is low, and vice versa. On the basis of this concept, we use the color distribution via a self-similarity map representation.

We first perform an RGB to HSV color space conversion of I_{intg}. In the HS color space (the V channel is not used because the influences of the changes in lighting conditions need to be excluded), we create a HS color histogram in each local image patch.

(a) (b)

Figure 4: Example of a self similarity map using gradient intensity distributions. (a) Integrated gesture image I_{intg}; (b) Computed self-similarity $s^{m',n}_{\text{grad}}$. The white block in (a) denotes the m'-th patch (the patch of interest).

As in the $s^{m,n}_{\text{grad}}$ computation, we obtain a similarity $s^{m,n}_{\text{col}}$ between patch m and patch n by computing the Bhattacharyya distance of the normalized HS color histograms, $\mathbf{c}^m$ and $\mathbf{c}^n$, as

$$s^{m,n}_{\text{col}} = \sqrt{1 - \sum_{l=1}^{L} \mathbf{c}^m_l \mathbf{c}^n_l}\ , \tag{6}$$

where $\mathbf{c}^m_l$ denotes the frequency in the l-th bin of $\mathbf{c}^m$, and L represents the number of bins.

Similarly, we construct the self-similarity map of the HS color distribution S_{col} as

$$S_{\text{col}} = [s^{1,2}_{\text{col}}, ..., s^{1,B}_{\text{col}}, ..., s^{B-1,B}_{\text{col}}]\ . \tag{7}$$

Figure 5 shows an example of the computed $s^{m',n}_{\text{col}}$ (m' denotes the patch of interest).

2.3 Classifier Construction

We finally construct a classifier for hand gesture recognition. To train a classifier, we use a combination of the two self-similarity maps S_{grad} and S_{col} as

(a) (b)

Figure 5: Example of a self-similarity map using color distributions. (a) Integrated gesture image I_{intg}; (b) Computed self-similarity $s_{\text{col}}^{m',n}$. The white block in (a) denotes the m'-th patch (the patch of interest).

a training sample. By using the trained classifier, we perform hand gesture recognition.

3 EXPERIMENTAL RESULTS

In order to demonstrate the effectiveness of the proposed method, we tested the proposed method using the Cambridge hand gesture dataset (Kim et al., 2007). This dataset includes 5 illumination patterns, 9 motion (gesture) patterns, and 20 people; 900 sequences in total. We used a support vector machine (SVM) for hand gesture classification. To find the optimal parameters of SVM, we performed a five-fold cross validation. We empirically set the other parameters in the proposed system such that δ=0.45, B=400, K=8 and L=3. We used the same parameters for all the experiments. The experiments were run on a Windows PC with Intel Core i7-3770K 3.50GHz for our computation. Our classification cost 30 s in this computing environment.

3.1 Results for Bright-lighting Scenes

We first tested the proposed method in bright-lighting scenes. Although the Cambridge hand gesture dataset has the variations in illuminations, we viewed that the entire illumination level is sufficiently high among this dataset; we thus regarded the Cambridge hand gesture dataset as that taken in bright-lighting scenes.

In this experiment, we compared the proposed method with the following competitive methods: Liu's method (Liu and Shao, 2013) and 3D SIFT descriptor (Scocanner et al., 2007).

The confusion matrices obtained by these methods on the Cambridge hand gesture dataset are shown in Tables 1, 2, and 3. In these confusion matrices, "0, ..., 8" indicate the index of each gesture pattern. In addition, each element in confusion matrix represents the recognition rate (%) of the corresponding gesture. Table 4 shows the comparison results in average recognition rate. We can see that the results obtained by using the proposed method favorably compare with those of the comparison methods.

Table 1: Confusion matrix obtained using the proposed method. Note that each element represents the recognition rate (%) of the corresponding gesture. Further, "0, ..., 8" indicate the index of each gesture pattern.

	0	1	2	3	4	5	6	7	8
0	90	0	0	8	1	1	0	0	0
1	0	85	1	0	10	0	1	3	0
2	1	2	88	1	0	2	2	1	3
3	5	0	0	89	1	0	5	0	0
4	2	21	0	3	72	0	0	2	0
5	2	0	2	0	1	89	0	0	6
6	0	2	0	1	0	0	90	6	1
7	0	4	1	0	2	0	5	88	0
8	0	0	2	0	0	2	2	0	94

Table 2: Confusion matrix reported in (Liu and Shao, 2013).

	0	1	2	3	4	5	6	7	8
0	96	0	0	3	0	0	1	0	0
1	0	97	0	0	2	0	0	1	0
2	3	1	82	0	0	10	0	0	4
3	2	0	0	95	0	0	3	0	0
4	0	12	0	0	85	0	0	3	0
5	0	0	3	4	0	92	0	0	1
6	5	0	0	12	0	0	83	0	0
7	1	19	0	2	5	0	1	72	0
8	1	0	10	0	0	19	0	0	70

Table 3: Confusion matrix obtained using the method (Scocanner et al., 2007).

	0	1	2	3	4	5	6	7	8
0	96	0	0	1	0	0	2	0	1
1	0	90	1	0	6	0	0	3	0
2	0	0	97	0	0	0	0	0	3
3	2	0	0	95	0	0	3	0	0
4	0	7	0	0	92	0	0	1	0
5	0	0	1	0	0	94	0	1	4
6	1	0	0	0	0	0	97	0	2
7	0	0	1	0	1	0	0	98	0
8	0	0	2	0	0	1	0	0	97

Table 4: Comparison results in average recognition rate. The results of Method 1 have been reported in (Liu and Shao, 2013), whereas those of Method 2 have been reported in (Scocanner et al., 2007).

Average recognition rate [%]		
Proposed method	Method 1	Method 2
87.2	85.8	95.1

3.2 Results for Low-light Scenes

In this experiment, we tested the proposed method on a dataset simulating low-light scenes.

(a) A=1 (b) A=0.5 (c) A=0.1 (d) A=0.05 (e) A=0.025

Figure 6: Examples of synthesized low-light hand gesture images.

3.2.1 Synthesis of Low-light Hand Gesture Sequences

We synthesized hand gesture sequences simulating those captured in low-light scenes. We first decreased the brightness of the original gesture sequences from the Cambridge hand gesture dataset. In particular, we multiplied the intensity of original images by the scale value $A(\leq 1)$. We varied A=1 (same brightness level as that of the original sequence), 0.5, 0.1, 0.05 and 0.025. Depending on the value of A, we added Gaussian noise. In order to simulate the amount of noise imposed from real low-light scenes, we analyzed a relation between the brightness level and the amount of noise to be added. In fact, the amount of noise differed for each sequence because of the difference of the relation between the brightness level and the S/N ratio of each image. Examples of the synthesized low-light images are shown in Figure 6.

3.2.2 Results

In order to evaluate the recognition performance, we compared the proposed method with the method using 3D SIFT descriptors (Scocanner et al., 2007) for each brightness level.

Tables 5, 6, 7, 8, 9, 10, 11, and 12 show the confusion matrices obtained by using these methods on the synthesized low-light dataset for A = 0.5, 0.1, 0.05 and 0.025. We observed that our method outperformed the comparison method at each value of A.

Table 13 and Figure 7 show the average recognition rate with respect to each value of A. We can see that the recognition rate of the proposed method is high when A is small. In contrast, the recognition performance in the comparison method significantly decreases for low A. From these evaluations, we conclude that the proposed method is effective in low light-scenes.

Table 5: Confusion matrix obtained using the proposed method ($A = 0.5$).

	0	1	2	3	4	5	6	7	8
0	92	0	0	5	0	2	1	0	0
1	0	76	1	1	16	0	3	3	0
2	0	7	90	0	0	0	0	1	2
3	6	0	2	87	5	0	0	0	0
4	0	16	1	2	80	0	0	0	1
5	2	0	3	0	0	92	0	0	3
6	1	1	0	0	1	0	90	5	2
7	0	5	1	0	2	0	4	86	2
8	0	0	5	0	0	3	6	4	82

Table 6: Confusion matrix obtained using the method (Scocanner et al., 2007) ($A = 0.5$).

	0	1	2	3	4	5	6	7	8
0	91	0	1	0	0	1	7	0	0
1	0	78	0	0	8	0	0	13	1
2	0	0	80	0	0	0	0	0	20
3	5	0	0	89	0	0	6	0	0
4	0	11	0	0	84	0	0	5	0
5	0	0	3	0	0	79	0	1	17
6	4	0	0	0	0	0	95	0	1
7	0	1	0	0	0	0	0	97	2
8	0	0	6	0	0	0	0	0	94

3.3 Limitation

In the experimental results, we observed the effectiveness of the proposed method in low-light scenes. However, the proposed method has a limitation.

As described in Section 2.1, the proposed method temporally integrates the input hand gesture sequence into a single color image. However, when the gesture speed is quite high (or the video length is short), the proposed system cannot obtain an image including such gesture motion. Figure 8 shows an example of I_{intg} where the video length is quite short. We can see that I_{intg} was unable to capture the hand gesture. In such a case, the proposed system would fail to recognize hand gestures.

4 CONCLUSIONS

We presented a method for recognizing hand gestures in low-light scenes. In such scenes, hand gesture im-

Table 7: Confusion matrix obtained using the proposed method ($A = 0.1$).

	0	1	2	3	4	5	6	7	8
0	90	1	0	7	0	1	1	0	0
1	1	73	3	2	17	0	0	4	0
2	2	4	87	0	0	2	1	0	4
3	7	1	0	88	3	0	1	0	0
4	0	17	1	1	79	0	0	2	0
5	3	0	3	2	1	89	0	0	2
6	0	2	0	0	0	0	95	1	2
7	0	1	0	0	2	0	3	92	2
8	0	0	7	0	0	1	2	3	87

Table 8: Confusion matrix obtained using the method (Scocanner et al., 2007) ($A = 0.1$).

	0	1	2	3	4	5	6	7	8
0	84	0	1	2	0	1	11	0	1
1	0	70	0	0	8	0	0	21	1
2	0	0	78	0	0	0	0	0	22
3	4	0	0	85	0	0	8	0	3
4	0	14	0	0	75	0	0	10	1
5	0	0	5	0	0	64	0	1	30
6	3	0	2	1	0	0	84	0	10
7	0	3	6	0	1	0	0	82	8
8	0	0	8	0	0	1	0	0	91

Table 9: Confusion matrix obtained using the proposed method ($A = 0.05$).

	0	1	2	3	4	5	6	7	8
0	80	1	2	14	0	1	1	1	0
1	3	70	2	1	20	0	0	3	1
2	1	2	84	1	0	5	1	3	3
3	15	2	1	79	2	0	1	0	0
4	0	24	1	8	63	0	0	4	0
5	4	1	6	3	0	82	1	0	3
6	2	1	1	1	0	0	88	5	2
7	0	4	1	0	2	0	5	86	2
8	0	0	7	0	0	1	4	4	84

Table 10: Confusion matrix obtained using the method (Scocanner et al., 2007) ($A = 0.05$).

	0	1	2	3	4	5	6	7	8
0	74	0	2	3	0	0	15	0	6
1	0	60	4	0	9	0	0	24	3
2	0	0	83	0	0	0	0	0	17
3	13	0	1	63	0	0	14	0	9
4	0	16	2	0	64	0	0	13	5
5	0	0	12	0	0	43	0	1	44
6	4	0	2	0	0	0	72	0	22
7	0	1	6	0	0	0	0	65	28
8	0	0	26	0	0	1	0	0	73

ages are significantly deteriorated because of heavy noise, and the previous methods may not work well. In order to address this problem, we used a single color image that temporally integrates a hand gesture sequence. In general, temporal integration of images improves the S/N ratio; it can capture the necessary

Table 11: Confusion matrix obtained using the proposed method ($A = 0.025$).

	0	1	2	3	4	5	6	7	8
0	87	0	0	10	0	2	0	0	1
1	3	75	3	1	14	0	0	4	0
2	2	2	83	1	5	3	0	2	2
3	10	1	1	85	2	0	1	0	0
4	0	18	1	3	77	0	0	0	1
5	2	0	3	2	2	88	1	0	2
6	0	0	2	1	1	0	88	8	0
7	0	3	3	0	3	0	9	80	2
8	0	0	4	0	0	4	3	5	84

Table 12: Confusion matrix obtained using the method (Scocanner et al., 2007) ($A = 0.025$).

	0	1	2	3	4	5	6	7	8
0	56	0	14	3	0	0	10	0	17
1	0	55	14	0	7	1	0	8	15
2	0	0	83	0	0	0	0	0	17
3	12	0	10	49	0	2	11	0	16
4	0	22	18	0	42	1	0	8	9
5	0	0	23	0	0	32	0	0	45
6	1	0	16	0	0	0	39	0	44
7	0	1	29	0	0	0	0	23	47
8	0	0	36	0	0	0	0	0	64

Table 13: Average recognition rate. We compared the proposed method with the method (Scocanner et al., 2007).

	Average recognition rate [%]	
A	Proposed method	Comparison method
0.025	83.0	49.2
0.05	79.6	66.3
0.1	86.7	79.2
0.5	86.1	87.4
1.0	87.2	95.1

appearance and motion information for hand gesture recognition in low-light scenes. Temporal integration of a hand gesture produces motion blur, which is considered to be one of the annoying noises that degrade performance in various computer vision or image processing techniques. In contrast, we effectively used motion blur for recognizing hand gestures. We assumed that the direction and the magnitude of the motion blur represent discriminative characteristics that can be used for differentiating various hand gestures. In order to extract these features from a motion-blurred image, we analyzed the gradient intensity distributions as well as the color distributions. We encoded both the gradient intensity and the color distributions by using a self-similarity map to make them invariant to the variations in the users' gestures. By using the self-similarity maps, we constructed a classifier for hand gesture recognition. Through the experiments, we demonstrated that the proposed method was more effective than the previous methods in the case of low-light scenes. We finally discussed the

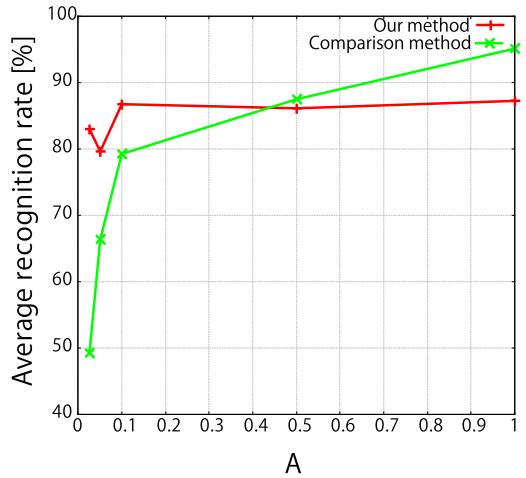
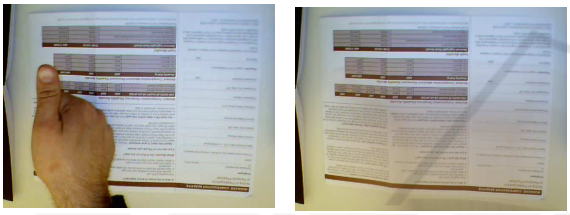


Figure 7: Average recognition rate. The red line denotes the result obtained using the proposed method, and the green line represents that obtained using (Scocanner et al., 2007).



(a) Original gesture (Shen et al., 2012)

(b) Integrated image I_{intg}

Figure 8: Limitation of the proposed method. When the gesture speed is considerably high, our temporal integration cannot capture the gesture motion.

limitation of the proposed method. When the speed of a hand gesture was considerably high, the proposed gesture integration scheme could not capture the hand gesture motion. We believe that the use of a high-frame-rate camera would be effective in solving this problem.

In the experiments shown in Section 3.2, we tested the proposed method using the synthesized hand gesture sequences. In general, however, it is hard to synthesize real-world conditions reliably. In order to ensure the strengths of the proposed method even in real low-light scenes, we will test the proposed method using gesture sequences taken in real low-light conditions. We intend to focus on these points in our future work.

REFERENCES

Bregonzio, M., Gong, S., and Xiang, T. (2009). Recognizing action as clouds of space-time interest points. In *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition*, pages 1948–1955.

Cipolla, R., Okamoto, Y., and Kuno, Y. (1993). Robust structure from motion using motion parallax. In *Proc. IEEE Int. Conf. Computer Vision*, pages 374–382.

Davis, J. W. (2001). Hierarchical motion history images for recognizing human motion. In *Proc. IEEE Workshop. Detection and Recognition of Events in Video*, pages 39–46.

Freeman, W. T. and Roth, M. (1995). Orientation histograms for hand gesture recognition. In *Proc. IEEE Int. Workshop. Automated Face and Gesture Recognition*, pages 296–301.

Freeman, W. T. and Weissman, C. D. (1996). Television control by hand gestures. In *Proc. IEEE Int. Workshop. Automated Face and Gesture Recognition*, pages 179–183.

Ikizler-Cinbis, N. and Sclaroff, S. (2010). Object, scene and actions: Combining multiple features for human action recognition. In *Proc. European Conf. Computer Vision*, pages 494–507.

Iwai, Y., Watanabe, K., Yagi, Y., and Yachida, M. (1996). Gesture recognition using colored gloves. In *Proc. Int. Conf. Pattern Recognition*, pages 662–666.

J.Davis and M, S. (1994). Recognizing hand gestures. In *Proc. European Conf. Computer Vision*, pages 331–340.

Kim, T.-K., Wong, S.-F., and Cipolla, R. (2007). Tensor canonical correlation analysis for action classification. In *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition*, pages 1–8.

Lian, S., Hu, H. W., and Wang, K. (2014). Automatic user state recognition for hand gesture based low-cost television control system. *IEEE Trans. Consumer Electronics*, 60:107–115.

Lin, H. T., Tai, Y. W., and Brown, M. S. (2011). Motion regularization for matting motion blurred objects. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 33:2329–2336.

Liu, L. and Shao, L. (2013). Synthesis of spatio-temporal descriptors for dynamic hand gesture recognition using genetic programming. In *Proc. IEEE Conf. Automatic Face and Gesture Recognition*, pages 1–7.

Lucas, B. and Kanade, T. (1981). An iterative image registration technique with an application in stereo vision. In *Proc. Int. Joint Conf. Artificial Intelligence*, pages 674–679.

Marin, G., Dominio, F., and Zanuttigh, P. (2014). Hand gesture recognition with leap motion and kinect devices. In *Proc. IEEE Int. Conf. Image Processing*, pages 1565–1569.

Niebles, J., Wang, H., and Fei-Fei, L. (2008). Unsupervised learning of human action categories using spatio-temporal words. *Int. Journal of Computer Vision*, 79:299–318.

Pavlovic, V. I., Sharma, R., and Huang, T. S. (1997). Visual interpretation of hand gestures for human-computer interaction: Review. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):677–695.

Pfister, T., Charles, J., and Zisserman, A. (2014). Domain-adaptive discriminative one-shot learning of gestures. In *Proc. European Conf. Computer Vision*, pages 814–829.

Ren, Z., Yuan, J., Meng, J., and Zhang, Z. (2013). Robust part based hand gesture recognition using kinect sensor. *IEEE Trans. Multimedia*, 15:1110–1120.

Scocanner, P., Ali, S., and Shah, M. (2007). A 3-dimensional sift descriptor and its application to action recognition. In *Proc. ACM Int. Conf. Multimedia*, pages 357–360.

Shen, X., Lin, Z., Brandt, J., and Wu, Y. (2012). Dynamic hand gesture recognition: An exemplar-based approach from motion divergence fields. *Image and Vision Computing*, 30:227–235.

Stamer, T. and Pentland, A. (1995). Real-time american sign language recognition from video using hidden markov models. Technical Report TR-375, Media Lab., MIT.

Tang, D., Chang, H. J., Tejani, A., and Kim, T. K. (2014). Latent regression forest: Structured estimation of 3d articulated hand posture. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 3786–3793.

Wachs, J. P., Kolsch, M., Stern, H., and Edan, Y. (2011). Vision-based hand-gesture applications. *Communications of the ACM*, 54:60–71.

Walk, S., Majer, N., Schindler, K., and Schiele, B. (2010). New features and insights for pedestrian detection. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 1030–1037.

Weaver, J., Starner, T., and Pentland, A. (1998). Real-time american sign language recognition using desk and wearable computer based video. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 33:1371–1376.

Yamato, J., Ohya, J., and Ishii, K. (1992). Recognizing human action in time-sequential images using hidden markov model. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 379–385.