손 동작 인식을 위한 Optical Flow Orientation Histogram

Optical Flow Orientation Histogram for Hand Gesture Recognition

Dhi Aurrahman*, Nurul Arif Setiawan*, Chi-Min Oh* and Chil Woo Lee**

Abstract Hand motion classification problem is considered as basis for sign or gesture recognition. We promote optical flow as main feature extracted from images sequences to simultaneously segment the motion's area by its magnitude and characterize the motion's directions by its orientation. We manage the flow orientation histogram as motion descriptor. A motion is encoded by concatenating the flow orientation histogram from several frames. We utilize simple histogram matching to classify the motion sequences. Attempted experiments show the feasibility of our method for hand motion localization and classification.

핵심어: Optical flow, gesture recognition, histogram matching

1. Introduction

Gesture recognition research is motivated by one aim to improve human computer interaction to the level of human-human interaction. In this level, the demand wants the establishment of control and interaction between human and intelligent machines installed with gestures as humans normally do in social setting. A system capable of gesture recognition can be used in many applications from helping disable people, enrich educational experience to interacting with robots and intelligent space.

In this paper, we consider hand motion recognition as one of gesture recognition problem which can be employed as a basis for sign or gesture recognition in general. Our work is inspired from [1] which uses Motion Gradient Orientation (MGO) with multiclass Relevance Vector Machine (RVM) classifier to recognize hand motion. MGO is proposed in [2] which are calculated from a motion history image (MHI) and motion energy image (MEI). MHI is basically a template which stored all motion information during specific time interval. MHI is computed from moving edges where the recent moving pixels shown to be brighter.

MHI is a global descriptor since it accounts full image size. Thus it limits the usefulness for more general setting in natural application. For example by detecting hand motion and accumulate MHI in the detected motion area. One way to alleviate this limitation is by using local gesture features. The other weakness is the size of feature is proportional to the image dimension. Simple template matching as in [2] may be sufficient for controlled application. In [1], PCA is used to reduce features before training is performed by RVM.

We propose a hand motion recognition system by using optical flow calculated from sequences of images. Optical flow has some interesting properties as it can separates motion area by its magnitude and characterize motion in one region by its motion direction (orientation). Thus it can be used to localize the moving region and recognize motion occurred in that region. The motion characteristics of the localized region is modeled as histogram of flow orientation (here we call flow orientation histogram) over several frames. For current research, the focus is limited on 2-D motion model.

*주저자 : 전남대학교 전자컴퓨터공학과 석사 e-mail: {dio,arif,sapeyes}@image.chonnam.ac.kr

**교신저자 : 전남대학교 전자컴퓨터공학과교수; e-mail: leecw@chonnam.ac.kr

The remainders are organized as follows. Section 2, optical flow for segmentation and motion descriptor is explained. Then we will describe motion recognition strategy by using a simple histogram matching. Experimental results will be discussed in section 4, and we will conclude and outline the future direction in section 5.

1.1. System Architecture

We begin by describing the proposed system architecture as shown in Fig. 1. Starting from calculating optical flow from sequences of images, our system follows general gesture recognition system by extracting feature and collecting it into one sequence to encode motion as one code/descriptor (here we call flow orientation histogram descriptors (FOH)) and then match the observed descriptors with existing models in the database.

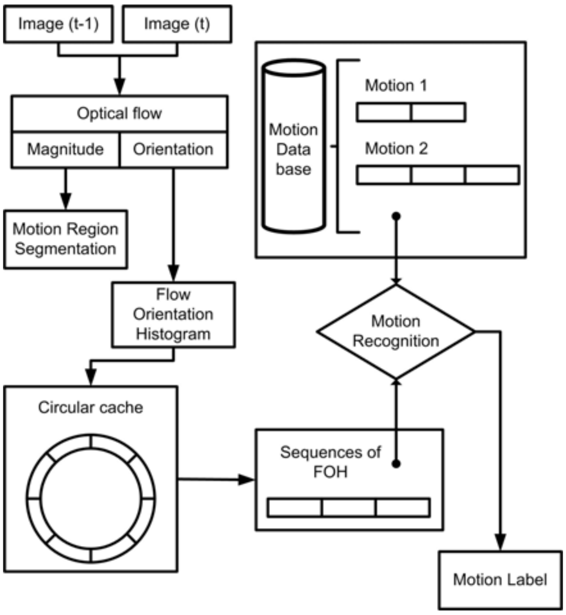


Figure 1. Proposed system architecture.

Matching observation into models can be performed in several ways. The commonly known method is Hidden Markov Model (HMM) as used in [3]. However, HMM require large training data and analyze the sign as a whole part, thus making the system complicated and difficult to extend into larger system [1]. Accumulating motion into single template (MGO or MHI) is used in [1-2]. In [1], classification process is performed by means of multiclass classification using RVM. Since the motion information contained in one template, this approach is more appropriate. However, the feature' s size is large. They use PCA to reduce the size of the feature. Here, we will use efficient histogram comparison to recognize appropriate motion. For this application, we use diffusion distance [1] as a measure of histogram comparison. We choose this method (histogram matching) because it more efficient compared to classifier method based since we do not need to generate training data sets. Furthermore the classifier is too dependent on training data and usually we need to create a data set which contains all possible motion variation.

For each motion model, the size of FOH descriptor will be varied depends on the number of frames required to represent the motion. The requirement to compare the descriptors over time, an appropriate way to store the histogram is needed; by this the system allows any FOH descriptors with different length. We use circular cache with fixed size. This cache may accept FOH descriptor at any length and avoid the out of memory problem for storing newly observed flow histogram.

2. Optical Flow for Segmentation and Feature

2.1. Segmentation using Flow Magnitude

We use pyramidal implementation of Lucas-Kanade optical flow. This gives good estimation of optical flow since displacement estimation is propagated through several scales of image pyramid. To improve the image, we use gamma correction and homomorphic filter to brighten the image and maintain constant illumination level. This optical flow can be employed as predicate for compensating motion and detect moving object for mobile robot with omnidirectional camera [5]. To speed up the process the original 320x240 image is downscaled into 160x120 without losing the performance.

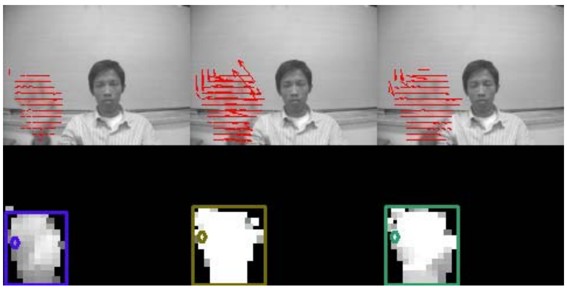


Figure 2. Optical flow image and segmentation of hand area by using flow magnitude.

Optical flow estimates a translation of rectangular patches. We define a dense overlapped patches and estimate flow for each patch. As shown in Fig. 2, patches with large motion magnitude appear brighter than the small motion. We segment motion area by employed a specific threshold value to patches magnitude image and the extracted rectangular region.

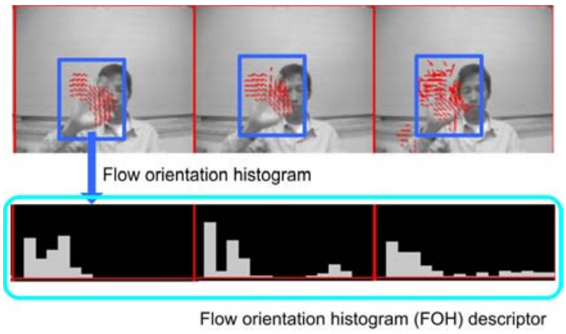
2.2 Flow Orientation Histogram Descriptor

Motion direction is one apparent pattern in hand motion. By using histogram to encode a motion direction in one frame, we obtain a motion cue. The histogram is then accumulated from several frames to create temporal descriptor (FOH). This representation is more compact and local than MHI since it uses histogram to represent the motion and only consider specific region instead of using whole image.

The optical flow estimation is still noisy and when the motion is too fast, optical flow cannot estimate the translation correctly. Therefore, in voting the flow orientation histogram, we use magnitude as a threshold. We normalize the histogram using its magnitude. Several histograms from sequences are concatenated to define a specific motion (FOH) descriptor for every motion. This is illustrated in Fig. [3]. Depending on how many frames sufficient to represent the motion, each motion is defined with different FOH's length.

Figure 3. Flow orientation histogram (FOH) descriptor.

3. Motion Recognition Using FOH Comparison



Gesture recognition involves integration of temporal information as one gestures typically spanned over several frame of input images. Previously we have defined FOH descriptors to encode motion pattern of a moving hand region. Now we need a method to compare the observed FOH and models.

3.1. FOH Comparison with Diffusion Distance

We use diffusion distance [4] to measure histogram difference. The performance of diffusion distance has been shown better compared to previously used histogram comparison method such as Bhattacharyya distance and EMD (Earth Mover's Distance) [4]. We choose diffusion distance because it is a fast cross-bin histogram distance type and in our system, FOG mostly differs in the maximum orientation bin. Thus, it will provide fast, robust and effective similarity measure which is one important component that we need.

Principally, diffusion distance is based on heat diffusion equation. Given an isolated temperature field $T(x,y)$ at time $t=0$, the heat will be diffuse and eventually reach equilibrium at time t. Similarly, for two histogram h_1 and h_2, the process diffuses the differences between two histograms until h_1 and h_2 equivalent. The diffusion distance $\hat{K}(h_1,h_2)$ is defined as [4]:

$$\hat{K}(h_1,h_2)=\int_0^{\bar{t}} k(|T(x,t)|)dt \qquad (1)$$

where

$$k(T|x,t|)=2\left(2\int_{-\infty}^{\Delta/2}\phi(x,t)dx-1\right) \qquad (2)$$

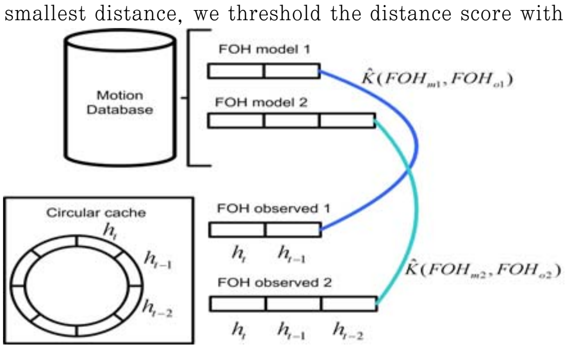
And $\phi(x,t)$ is the Gaussian filter defined as:

$$\phi(x,t)=\frac{1}{(2\pi)^{1/2}t}\exp\{-\frac{x^2}{2t^2}\} \qquad (3)$$

3.2. Classification Procedure

As we have shown in Fig. 1 and discussed in Section 1.1, we implemented circular cache with fixed size to store all histograms in N sequential frames. Because we do not consider a method to infer next possible histogram by using trained data set, in the recognition step, we simply extract FOH descriptors from the cache and then compare the observed FOH with existing model. Depending on the motion model, length of FOH will be different each other. This process is shown in Fig. 4.

We then sort the score for each model. The motion on any time instance is the motion with smallest diffusion distance $\hat{K}(h_1,h_2)$. To make the recognition robust to noise and we do not blindly choose the smallest distance, we threshold the distance score with



minimum similarity score τ .

Figure 4. Recognition of motion at one time instance is performed as FOH comparison through diffusion distance technique.

4. Experiments

4.1. Motion Database

We made a database of hand motion consisting of 8 distinct motions of 4 motion pairs: (1) translating left and right, (2) translating downward and upward, (3) circular clockwise and counter clockwise and (4) waving left and right. The frame length of each model is shown in Table 1.

The flow orientation is voted into 12 bin histogram. Histogram is sampled with orientation range $[(-\pi)-(\pi)]$. We accumulated histogram for 4-29 frames depending on the nature of motion. Because the observation is usually affected by noise, we created an ideal FOH model for every motion by assuming histogram distribution in a well controlled environment. Example for two motion, translating right and up is shown if Fig. 5. In the figure, we can notice the differences in FOH for each motion.

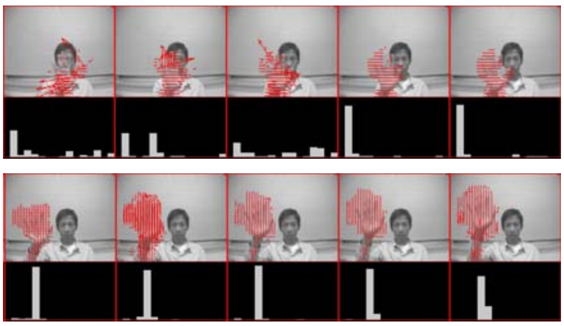


Figure 5. Example FOG for right translation (up) and up translation (bottom).

Table 1. FOH's Length

Motion	Length (frame)
Up	4
Down	4
Left	4
Right	4
Circular CW	12
Circular CCW	12
Waving Left	29
Waving Right	29

4.2. Performance

To evaluate the performance we annotate a video consisting of defined hand motion. We accumulate FOH into circular cache with size $N = 40$ .We use score threshold $\tau = 15$. The performance measure is percentage of correctly recognized performed motion. The recognition rate is shown in Table 2. Because we use different size of models, small length FOH is favored because the span is short and more likely to be selected as current motion. The FOH with short length (Up, Down, Left and Right) have high detection rate compared to the longer

Table 2. Performance of FOH for hand motion recognition

Motion	Accuracy %
Up	85
Down	80
Left	80
Right	85
Circular CW	70
Circular CCW	60
Waving Left	75
Waving Right	65

4.3. Discussion

Since the system is based on optical flow estimation, robust and accurate optical flow estimation algorithm is needed to ensure acceptable performance. Optical flow estimation remains a difficult problem and most algorithms that perform better than Lucas-Kanade usually not fast enough for real-time implementation. Another limitation is optical flow only represent 2D motion. This might be a limitation for current system, however, in some applications, gesture from 2D information is sufficient, as shown in [2]. We are planning to improve optical flow estimation in the future by adding 3-D information e.g. by calculating disparity.

5. Conclusion

We have proposed a novel approach for the motion recognition problem by using optical flow information. The optical flow is used to localize the moving region and recognize motion occurred in that region. The motion characteristic of the localized region is modeled as flow orientation histogram descriptors. We utilize simple histogram matching to classify motion sequences. Experiments result shows the feasibility of our method for hand motion localization and recognition.

Future development of the system includes robust optical flow estimation, covering 3-D motion and creating data sets for evaluation and comparison with other methods.

Acknowledgement

This research is supported by KETI (Korean Electronics Technology Institute) through Future Electronic Home Appliances Technology Development Project

References

[1] Wong, S.F., Cipolla, R., "Real-time adaptive hand motion recognition using a sparse bayesian classifier," Computer Vision in Human-Computer Interaction,ICCV 2005 Workshop on HCI, pp. 170-179, 2005.

[2] Bradski, G.R., Davis, J.W., "Motion segmentation and pose recognition with motion history gradients," Machine Vision and Applications, ed.13, vol.3, pp. 174 184, 2002.

[3] Starner, T., Weaver, J., Pentland, A., "Real-time american sign language recognition using desk and wearable computer based video," IEEE Transactions on Pattern Analysis and Machine Intelligence, ed. 20, vol. 12, pp. 1371 1375, 1998.

[4] Ling, H., Okada, K.," Diffusion distance for histogram comparison," International Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 246 253, 2006.

[5] Arif, S.N., "Development of omnidirectional human detection module for mobile robot". Master's thesis, Chonnam National University, 2007.

[6] Dalal, N.," Finding people in images and videos," PhD thesis, Institut National Polytechnique de Grenoble, 2006.