

A Study on Motion Recognition of Objects in Soccer Game

Jungsoo Lee*, Jiwon Lee*, Sungwon Moon*, Dowon Nam*

*SW Contents Research Laboratory, ETRI, Daejeon, KOREA

jslee2365@etri.re.kr, ez1005@etri.re.kr, moonstarry@etri.re.kr, dwnam@etri.re.kr

Abstract— Recognition of motion in football videos is necessary for an automatic understanding of the events that occur within football games. In order to recognize the motion, we first define the motions that can occur for each object in the football game, and make our system learn the object that performs the defined motion from the actual football game videos. Then, the system is configured to automatically recognize the motions generated in the actual football game videos within the defined motion. Experiments on motion recognition for 20 players, 1 referee, and 2 assistant referees were performed. The average recognition accuracy per motion for each object was 71.5%, 79.1%, and 93.7%. In this paper, we propose a method for simultaneously recognizing motions of various objects in football videos through learning.

Keywords— Motion recognition, football, learning

I. INTRODUCTION

Motion recognition technology that identifies human behavior based on video plays a very important role in video based automation such as unmanned surveillance system and sports player activity record system. In particular, automatic analysis of motion for an object in a sports video is the basis for automatic recognition of an event occurring in a sports game. In order to analyze the motion automatically, it is necessary to define the motion in advance and it is necessary to classify the position where the defined motion occurs so that the event occurring in the sport (football) game can be extracted. Apart from the location tracking for objects, there are basically three steps for motion recognition for each object. There are three stages: feature extraction, dictionary learning that is expressed in video based on extracted features, and classification into defined motion. In the first step, feature sets such as STIP [1,2] or values such as dense trajectories [3,4] are extracted from the given video. These features are encoded in a useful form such as a vector to recognize the motion. The extracted feature values are used as values for expressing the motion appearing in the sports videos. In the last step, when the motion video is input as a query, the motion is classified by finding the nearest motion among the predefined motions based on the learned motion.

Although motion recognition methods using handcrafted features such as Histogram-of-Oriented-Gradient (HOG) have been used as the motion recognition method, there were disadvantages that it is impossible to simultaneously recognize

multiple motion for multiple objects because of the low processing speed. In addition, if the types of motion videos input to the system are changed, there is a limitation in terms of application range and accuracy since the homogeneous features corresponding to the types of motion videos must be redesigned every time. In order to solve the problems related to accuracy, motion recognition methods using equipment provided with object depth information have been proposed. These methods have cost disadvantages in that they require additional special equipment such as Kinect. Therefore, there is a need for a motion recognition method that shows high accuracy and fast processing speed without special equipment.

In this paper, we propose a faster and more accurate motion recognition method than the conventional method by using feature information extracted by using deep learning technology to solve the above problems.

II. THE PROPOSED METHOD

A. Motion classification

It is necessary to predefine and classify the motions in the football game. For this purpose, we classify and define the motions for each player and referee.

TABLE 1. MOTION CLASSIFICATION FOR OBJECTS

Player(except goalkeeper)/Referee/Assistant	
object	Motion type
Player	Walk
	Run
	Corner kick
	Free kick
	Kick with swinging arms
	Throwing
	Lying down
Referee	Walk without any instructions
	Run without any instructions
	Pointing somewhere lifting one arm
	Yellow card
	Red card
Assistant referee	Walk without lifting a flag
	Run without lifting a flag
	Walk like as crab without lifting a flag

	Lifting the flag over the head
	Lifting the flag to the chest
	Lifting the flag and pointing left or right

B. Motion recognition flow

Figure 1 shows the flow of motion recognition for multiple objects.

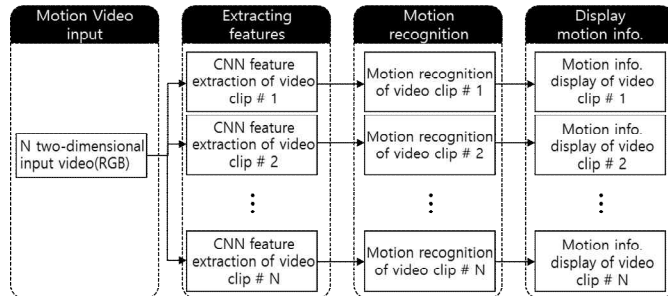


Figure 1. The flow of motion recognition process

First, N motion video clips to be subjected to motion recognition are loaded. Then, N input motion video clips are input to the input end of the feature information extractor for Deep Learning based specific information extraction. Feature information to be used for motion recognition is extracted from each input motion video clip by a feature information extractor. Finally, using the feature information extracted from each motion video clip, motion with the highest correlation with the given motion video clips is determined.

C. The structure of motion recognizer

Figure 2 shows the structure of a motion recognizer based on Deep Learning. Feature extraction and motion recognition are performed from input video clips.

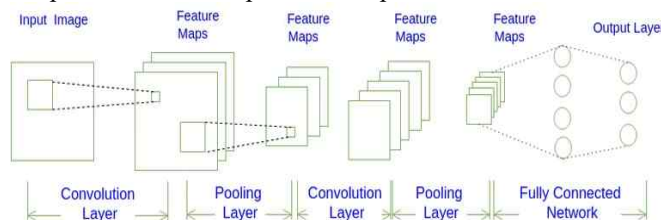
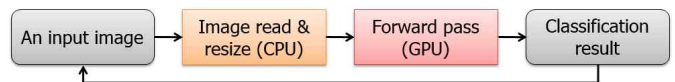


Figure 2. The structure of motion recognizer

Figure 3 shows the process of batch-process by receiving multiple video clips as inputs. This method increases the average efficiency of the CPU and the GPU and improves the processing speed by simultaneously processing multiple input video clips and inputting them in parallel to the input of the deep learning classifier.

- Function flow without batch (Slow)



- Function flow with batch size N (Fast)

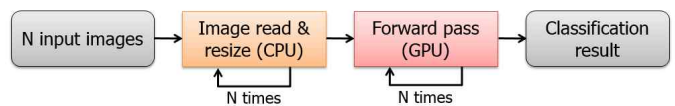


Figure 3. The structure of batch process

Table 2 shows the relationship between processing speed and batch size.

TABLE 2. THE RELATIONSHIP BETWEEN PROCESSING SPEED AND BATCH SIZE

Batch Size	Time per images[ms]	Frame per second[Hz]
1 (No batch)	3.72	368
2	2.68	373
4	1.81	552
8	1.29	775
16	1.11	901
32	0.98	1020
64	0.95	1053
*25	1.03	971

III. EXPERIMENTAL RESULTS

Experiments for motion recognition were performed based on Table 1 for acquired motion data (total 108,000 frames). The given motion data is classified by random selection of 3: 1 ratio of learning and test data based on football videos.

And as shown in Fig. 4, we have developed 16 different models for motion recognition, consisting of 4 cameras and 4 objects (assistant referee1, assistant referee2, main referee, players).

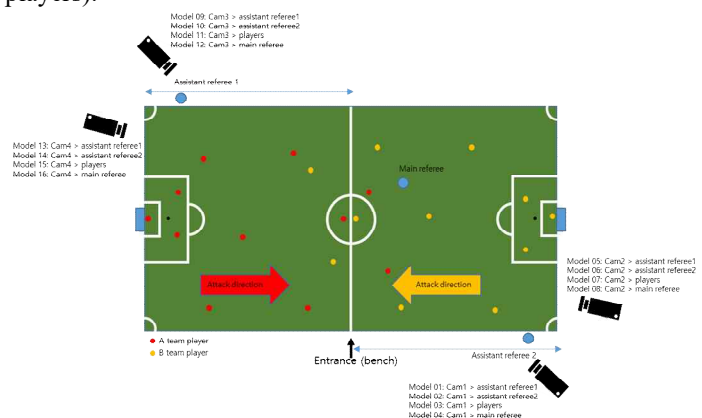


Figure 4. The model for objects' motion learning and recognition

Figure 5 shows the motion recognition of the player, referee, and assistant referee based on the learned motion through the motion recognition system.

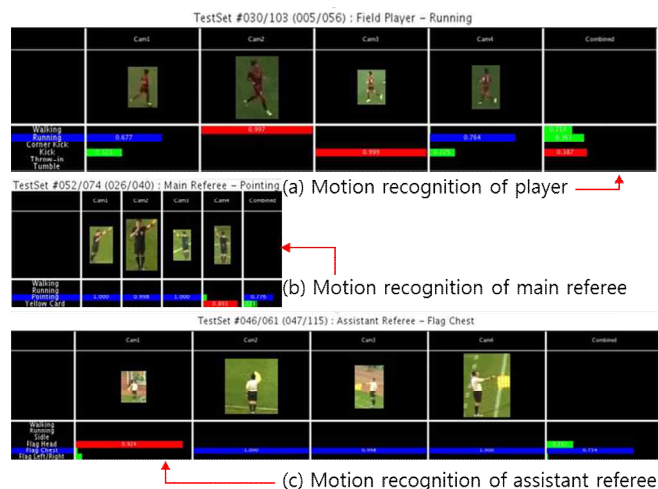


Figure 5. Motion recognition of objects

And from the experiment, we could confirm the average accuracy of each motion recognition of players, referee, and assistant referees as following table 3, 4 and 5 respectively.

TABLE 3. AVERAGE ACCURACY RATE FOR PLAYERS

Confusion matrix for motion recognition of players								
		Motion						
		W	R	CK	FK	KB	T	L
Result of classification	W	534	120	0	0	0	27	0
	R	57	360	0	129	21	12	0
	CK	0	0	273	0	90	0	21
	FK	0	0	0	0	0	0	0
	KB	36	15	0	75	243	0	45
	T	0	0	0	0	36	174	3
	L	0	18	0	0	21	9	261
Average accuracy rate for each motion		85.2%	70.2%	100%	0%	59.1%	78.4%	79.1%
Total average accuracy rate		71.5%						

*(W: Walk, R: Run, CK: Corner kick, FK: Free kick, KB: Kick the Ball, T: Throwing, L: Lying down)

TABLE 4. AVERAGE ACCURACY RATE FOR MAIN REFEREE

Confusion matrix for motion recognition of main referee					
		Motion			
		Walk	Run	Pointing	Card
Result of classification	Walk	549	69	0	0
	Run	111	597	51	54
	Pointing	0	0	288	93
	Card	0	0	0	0
Average accuracy rate for each motion		83.2%	89.6%	85.0%	0%
Total average accuracy rate		79.1%			

TABLE 5. AVERAGE ACCURACY RATE FOR ASSISTANT REFEREES

Confusion matrix for motion recognition of assistant referee							
		Motion					
		W	R	CW	LH	LC	LP
Result of classification	W	153	27	0	3	0	0
	R	6	99	0	0	0	0
	CW	0	0	129	0	0	3
	LH	0	0	0	63	0	0
	LC	0	0	0	0	213	0
	LP	0	0	0	21	0	234
Average accuracy rate for each motion		96.2%	78.6%	100%	72.4%	100%	98.7%
Total average accuracy rate		93.7%					

*(W: Walk, R: Run, C: Crab walking, LH: Lifting flag over the head, LC: Lifting the flag to the chest, LP: Lifting the flag and pointing left or right)

IV. CONCLUSION

In this paper, we proposed a method to recognize the motion of the object from the football game video. By learning the motion using the deep learning technique and simultaneously recognizing the various motions, it was possible to determine in real time what motion the object was doing by attaching only the tracker for the object to the front of the motion recognition through the input of the actual football game video. In addition, this paper have made us equip the base technology for event recognition in the game by enabling motion recognition for many objects in real time. And if we improve the accuracy of motion recognition in the future, we expect that it is useful for event extraction and tactical analysis based on this technology.

ACKNOWLEDGMENT

This work was supported by MCST(Ministry of Culture, Sports & Tourism) / KOCCA(Korea Creative Content Agency) (R2016030044 - Development of Context-Based Sport Video Analysis, Summarization, and Retrieval Technologies).

REFERENCES

- [1] Laptev I (2005) On space-time interest points, *Int J Comput Vis* 64(2-3): 107-123
- [2] Laptev I, Lindeberg T(2003) Space-time interest points. In: *International conference on computer vision (ICCV)*
- [3] Wang H, Kläser A, Schmid C, Liu C, *Action recognition by dense trajectories*. In: *Computer vision and pattern recognition (CVPR)*, 2011.
- [4] Wang H, Kläser A, Schmid C, Liu CL, *Dense trajectories and motion boundary descriptors for action recognition*. *Int J Comput Vis* 103(1):60-79, 2013.
- [5] Yuan C, Hu W, Tian G, Yang S, Wang H, *Multi-task sparse learning with beta process prior for action recognition*. In: *Computer vision and pattern recognition (CVPR)*, 2013.
- [6] Everts I, van Gemert J, Gevers T, *Evaluation of color stips for human action recognition*. In: *Computer vision and pattern recognition (CVPR)*, 2013.



Jungsoo Lee received his B.S. and M.S. degrees from Jeonbuk University, Korea in 1995 and 1997, respectively and his Ph.D. degree in Electronic Engineering from Hanyang University, Seoul Korea in 2005. From 2000 to 2005, he was a senior member of MarkAny Research Institute. Currently, he is a senior member of Electronics and Telecommunications Research Institute (ETRI). His research interests are digital watermarking, fingerprinting, image

processing, digital rights management, digital cinema and digital signage.



Jiwon Lee received the B. S. degree in Computer Engineering from Kyungpook National University, Republic of Korea, in 2008, and the Ph. D. degree in Computer Science from Korea Advanced institute of Science and Technology (KAIST), Republic of Korea, in 2013. Since 2013 he has been a senior researcher in the Sportainment Section, SW.Content Reserch Laboratory, Electronics and Telecommunications Research Institute (ETRI), Republic of Korea. His research interests include multimedia security,

image/video watermarking, and image/video processing.



Sungwon Moon received his B.S and M.S degrees of computer science from KAIST, Korea in 2010 and 2012. He is a RESEARCHER of Electronics and Telecommunications Research Institute(ETRI), Korea. His research interests are Digital Watermarking, Video Forensic, and Video Processing.



Do-Won Nam received the B.S. degree in Computer Science from Korea Advanced Institute of Science and Technology (KAIST) in 1996 and M.S. degree in Information Technology from Pohang University of Science of Technology (POSTECH), Korea in 1998. He is working as a principal researcher in the Electronics and Telecommunications Research Institute (ETRI) since 2001. His research interests include data mining, digital rights management, digital cinema system and sports video analysis.