PKU-SHRC Team Description for RoboCup 2010

Guangnan Ye, Xibin Chen, Kaihua Jiang, Caifu Hong, Xu Wang, Wenqian Zhang,
Guangcheng Zhang, Xihong Wu

Speech and Hearing Research Center,
Key Laboratory of Machine Perception, Peking University, China
{yegn, chenxb, jiangkh, hongcf, wangx, zhangwq, zhanggc, wxh}@cis.pku.edu.cn

Abstract. This paper introduces the hardware and software design of robots
developed by PKU-SHRC Team for RoboCup 2010, Singapore. Each robot
consists of 25 actuated degrees of freedom based on Dynamixel RX28, and
RX64 servos. The whole system runs on a powerful ARM and DSP platform at
a high speed. Software design combines the technology of computer vision,
speech recognition, pattern recognition, and mechanical control. The robots can
detect and track each single object in the playfield by vision. They can localize
themselves accurately by machine learning algorithm. They can communicate
with each other through natural language by speech recognition. They can also
do planning and inference by artificial intelligence technology. What we want
to do is to make the robot more like a human, not only from its appearance and
mechanical design, but also from its innate intelligence.

Keywords: robot, vision, speech, artificial intelligence.

1 Introduction

Nowadays, humanoid robot is a promising and attractive area. It covers almost all
fields of artificial intelligence. Ever since RoboCup first started in 1997, the
competition has performed as a superior experiment platform for researchers.
Although the technology in this area develops very fast, there are still many problems.
To make it more like a human, with more powerful vision, speech, and hearing ability
is probably the dream of every team.

PKU-SHRC Team was started at the end of 2009 in Speech and Hearing Research
Center (SHRC) at Peking University. The lab is also affiliated to the Key Laboratory
of Machine Perception, one of the only three national research centers that integrated
multidisciplinary research to the study of computer vision, speech recognition, and
natural language processing. Since first organized in 1988, our research has been
focused on the basic theory and methodology for machine visual and auditory
perception, voice and text recognition, image recognition and image database,
intelligent system and knowledge engineering, and other fundamental research topics
with potential social and economical benefits. We believe that the future direction for
RoboCup would more and more rely on the fundamental research fields, such as
machine visual perception and auditory information processing.

In many international completions, we have sufficiently proved our research abilities in artificial intelligence. In 2004, we participated in "Speaker Recognition Evaluation" held by NIST, and finally ranked first in the task of speaker recognition in telephone dialogue. In Sep. 2009, we took part in "TRECVID 2009 Surveillance Event Detection" held by NIST. Among all submissions for the evaluation, we performed best and had four tasks ranked in the first place worldwide. Specially, in April 2008, Singapore held an international competition on video and audio retrieval, named A-Star Challenge 2008. The competition was considered to be the first trial for combining the technology of computer vision and speech recognition. Fifty-three teams from all over the world competed for the final. Through six months' effort, our team submitted a system which ranked the first places for the totally automated retrieval task. The above international competitions provide us much precious experience, motivate our technology always being in the leading place. They also set up a solid research background for us to join in the whole new world, RoboCup.

At the end of year 2009, just with one week's preparation, our team participated in RoboCup 2010 China. Because of time limitation for preparation, rather than competed for the best, what we wanted was to learn more from other teams. Through two days competition in kid-size class, our team got the first place in the group stage and ranked fourth place in the final. This excellent performance just gives us more confidence. We decide to participate in RoboCup 2010, Singapore. Although we are a new team, we are fully confident because we just hold the essence of the robot.

2 Hardware Design

2.1 Mechanic Architecture

Our robot (Fig. 1.) uses Dynamixel series robot actuators. Dynamixel RX-64 and Dynamixel RX-28 together constructed the whole structure. The motion mechanism consists of 25 degrees of freedom. Some specific mechanical construction of our Robot can be seen from Table 1.

Table 1. Mechanical construction of the robot.

Quantity	Value	Unit
Overall Height	46	cm
COM Height	33	cm
Weight	2830	G
Leg Length	22	cm
Foot Area	77	cm^2
Head Length	11	cm

2.2 Actuators

The actuators used in our robots are "Dynamixel RX-64" servomotors, produced by Robotis Inc. The Dynamixel series robot actuator is a smart, modular actuator that incorporates a gear reducer, a precision DC motor and a control circuitry with networking functionality, all in a single package. The parameters of the actuators used in our robots are summarized in Table 2.

Fig. 1. The robot

Table 2. Main Specifications of RX-64.

	RX-64	
Weight(g)	116	
Gear Reduction Ratio	1/200	
Input Voltage (V)	at 15V	at 18V
Final Max Holding Torque (kgf.cm)	53	64
Sec/60degree	0.198	0.162

2.3 Sensor

- Camera: 30 FPS CMOS video camera which provides us with a resolution of 640x480 pixels.
- Loudspeaker: Two loudspeakers and English vocal synthesis system.
- Microphone: Two microphones which can localize sound and do speech recognition.
- Accelerometer/Tilt sensor: The accelerometer is to detect a fall down or to stabilize the robot.

3 Software Design

Our robot embedded software modules allowing color and shape detection; object, obstacle or opponent detection; object tracking; self localization based on particle filter; text to speech; sound localization and speech recognition, etc.

3.1 Color and Shape Detection

We use some basic image processing algorithms such as NN based color segmentation [1], line and circle detection by Hough transformation and circle fit method. Based on such methods, we can find the possible area of the goal, ball, beacon, etc. In order to reduce the effect of illumination change, we do the color

segmentation in LAB color space. Also, we automatically do the retraining process when illumination changing happens. [2]

3.2   Object Detection

Detection is an important step in our system. Compared to the method that only uses the color information, we can detect the ball and the opponent robots more accurately through detection module. In [3], Dalal and Triggs have proved that Histograms of Oriented Gradients (HOG) is powerful enough to pedestrian detection. Zhu *et al.* [4] combined the cascade of rejecter approach with the HOG feature and significantly accelerated the speed of the system. In order to get a high performance system, we bring in Zhu's cascaded HOG framework to do the detection. After that, we combined the detection feature (represents the shape information) and color feature together. The accuracy result was totally improved by such combination.

3.3   Object Tracking

The target object appearance always changes significantly. It will affect object detection process. To solve the problem, we use adaptive Online Boosting framework. As it is described by Helmut Grabner in [5], the detection result of each step is regarded as positive data, while the surrounding four blocks are regarded as negative data. During the process, each tracklet is triggered by the result of detection module. Throughout the Online Boosting, we get an expected target for each detected object in the next frame. Consecutively, each expected target is tested by detection module to confirm whether drifting is happening. If the target object is drifting, termination module is triggered and the tracking is ended. Another problem is to refine the tracking results. Because the object in the soccer play always contains the dominant color, orange, yellow, blue, etc. Dominant color similarity can be regarded as a score to identify the target. We use the score to approach a trajectory as long as possible based on tracklets. From the above method, we can finally specify every single object including the ball, the opponent player, the teammates, etc, in every second. We can also judge the velocity of each object so that to trigger the corresponding action. The total framework for detection and tracking can be seen in Fig. 2.
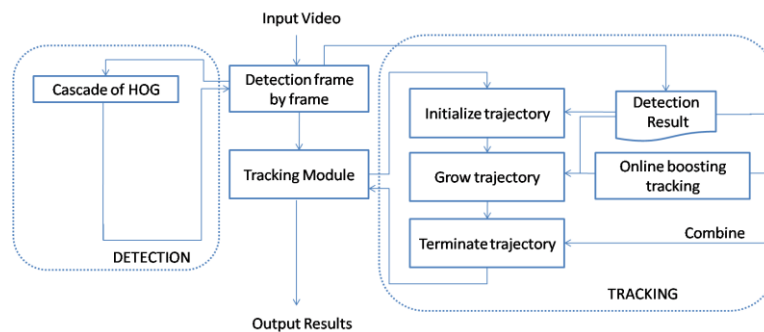
Fig. 2. Detection and Tracking Diagram.

3.4 Localization Based on Particle Filter

Self localization is done by Monte Carlo method. Particles are scored by comparing a simulated image from each particle with the current frame captured by camera. The resampling process can trigger a new distribution of the particles. By using particle filtering, each robot would localize itself correspondingly

3.5 Speech Recognition

In RoboCup, to enhance the communication ability of the robots, we can incorporate a speech module into the robot system. This module consists of two parts. One is the Text-To-Speech (TTS) part, which produces the speech signal of the commands and makes the sound through a loudspeaker. The other is the speech recognition part. Robot can detect whether there is a command, and act correspondingly.

The main problem is that the speech recognition part should work effectively online and in strong noisy environment where there may be people talking, whistling or clapping around. We consider using keyword spotting technology instead of large vocabulary automatic speech recognition to solve this problem, because there are enumerated commands that we need to detect and we require the system work fast to meet the online-decoding need.

Our proposed keyword spotting system is under the framework of the traditional HMM-based decoder. The unit of the acoustic model is context-independent (CI) phone for the sake of the limitation capacity of the store memory. Each unit is modeled by a three-state HMM, and the output probability of each state is modeled by Gaussian Mixture Model (GMM). We can use the TTS training corpus to train the acoustic model. To make our system more robust to noise, we use Perceptual Linear Predictive (PLP) as the acoustic feature.

We can design the commands list according to our playing strategy, which would consist of the keywords of the KSS. The keywords are then represented by concatenations of phonetic units. In order to deal with the background noise, we introduce two "garbage models" for people whistling and clapping. To ignore the silence segment, we introduce a silence model. These models, keywords models, background models, silence model are combined to form a searching network. Moreover, considering the people-talking-around situation, we can add all CI-phone models to the searching network.

After building the searching network, we can online decode using frame-synchronous Viterbi beam search algorithm. While decoding, when one keyword turns into a candidate hypothesis, if its confidence is higher than the threshold we preset empirically, the decoder would report a keyword-detection, and then the robot can take actions according to this keyword.

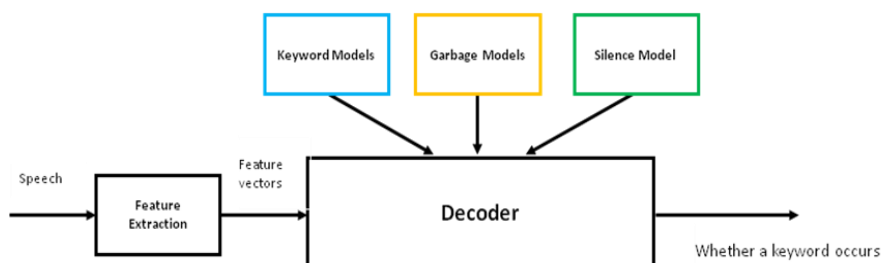The whole framework of the proposed keyword spotting system can be seen in Fig. 3.

Fig. 3. The framework of the proposed keyword spotting system.

4   Motion

We have developed different kinds of action, for example, walking, kicking, getup motion etc. For each action, several strategies have been planned. For instance, based on the velocity of the opponent, we can judge different types of kicking action. Recently, we are trying to do this part in a machine learning framework. For different playfields, different situations, the robot can learn to play better and better.

5   Conclusion

Recently, the state of art technologies of vision and speech recognition have been ported for our Humanoid League robots. In order to make our robot more robust and adaptive to the competition, we are trying to do further work on planning, cooperation and strategy in different situation. We look forward to participating in RoboCup 2010, Singapore. Hope our effort and performance would be a contribution in this fantastic area.

References

1. Smith, T.F., Peter Stone, Kurt Dresner, Peggy Fidelman, Nicholas K. Jong, Nate Kohl, Gregory Kuhlmann,   Mohan Sridharan, and Daniel Stronger. The UT Austin Villa 2004 RoboCup four-legged team: Coming of age. Technical Report UT-AI-TR-04-313(2004)
2. Mohan Sridharan, Peter Stone. Color Learning on a Mobile Robot: Towards Full Autonomy under Changing Illumination. IJCAI(2007)
3. N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. CVPR(2005)
4. Qiang Zhu, Mei-Chen Yeh, Kwang-Ting Cheng, Shai Avidan. Fast Human Detection Using a Cascade of Histograms of Oriented Gradients. CVPR(2006)
5. H. Grabner, T.T. Nguyen, B. Gruber, H. Bischof. On-line boosting-based car detection from arial images. ISPRS(2007)