댄싱 로봇의 구현을 위한 음악 템포 추출 및 모션 패턴 결정 방법

전명재* 유민수**
*한양대학교 지능형로봇공학과
**한양대학교 공과대학 컴퓨터공학부
e-mail : mjjun@rtcc.hanyang.ac.kr, msryu@hanyang.ac.kr

Music Tempo Tracking and Motion Pattern Selection for Dancing Robots

Myoung-Jae Jun*, Minsoo Ryu**
*Department. of Computer and Robot and Engineering , Hanyang University
**College of Information and Communications, Hanyang University

요 약

Robot 이 음악에 맞춰 어떤 행동을 하기 위해선 먼저 Acoustic 을 이해 할 수 있는 인지 능력이 필요하며 인지한 음악적 내용을 Dance Motion 에 가깝게 Action 을 표현할 수 있어야 한다. 본 논문에서는 신호처리와 기계학습을 사용하여 음악의 Tempo 를 Tracking 하고 이것을 참고하여 행동 Pattern 을 결정하는 Dance Robot System 을 소개한다.

1. 서론

댄스로봇을 정의하면 지능을 가진 로봇이 Acoustic 에 대해 의미 있는 행동으로 응답이 가능한 로봇을 말한다. 본 논문에서는 댄스로봇을 만들기 위한 과정의 하나로 음악 인지방법과 댄스 표현방법에 대해 기술한다. 음악 인지 방법으로는 음악의 아날로그 신호 중 드럼악기의 beat 수를 분석해 음악의 순간 Bit 를 추출하고 신호의 주기, 높 낮이의 정보와 함께 기계학습을 이용하여 음악의 Tempo 를 추적하는 알고리즘을 제시하며 댄스 표현 방법으로는 Tempo data 에 의해 선택된 특정 댄스동작이 음악의 강약에 의해 다시 동작 크기를 조절하는 방법을 제시한다. 논문에서 제시하는 Tempo 추적 알고리즘은 기존에 MID 나 MP3 파일을 로드 하여 신호를 분석해 Tracking 하는 것과는 달리 2 초간의 버퍼링 시간을 가지지만 Acoustic 한 성분이 있는 아날로그 시그널의 Tempo 를 실시간으로 추적 할 수 있다는 것에서 구별된다.

2. Tempo-and Pattern-based Dance Algorithm

Tempo 는 음악의 장르 및 느낌을 구분하는데 중요한 요소가 된다. 그러므로 Robot 이 음악을 인지 하기 위해선 입력되는 Acoustic 에 대해 Tempo 를 추출 할 수 있어야 한다. 방법은 다음과 같다. 입력되는 Audio Signal 을 A/D 변환해 Frequency analysis 를 하게 되는데 이 구간을 Buffering 구간이라 하며 음악이 Robot 으로 입력 된 후 스피커로 출력 되기까지 2 초 동안 이루어 진다. Buffering 구간에선 Robot 이 Tempo tracking 에 필요한 Beat 정보와 High, Low 로 이산화 시킨 진폭(Pitch), 주파수(Hz), Buffering 구간의 에너지(Energy)값을 취득한다. Bit 정보는 FFT. 식(0)를 사용한다.

$$X(k) = \sum_{k=1}^{N} x(n) e^{-j2\pi kn/N} \quad (0)$$

Frequency domain 에서 SD(Snare Dream)의 특정 주파수 대역(80~100Hz)을 모니터링 하면서 2 초 동안 SD 의 타점이 몇 번 발생 했는지를 기록해 Beat 를 Tracking 한다. (예 8Bit: 1 초에 SD 8 번) (그림 2)B Beat 를 포함한 Training data 는 식(1)로 표현하며 각각의 요소는 다음과 같다.

$$Training\ data = \{Bit, Pitch, Hz, Energy\} \quad (1)$$

$$Bit = \{4Bit, 8Bit, 16Bit\} \quad (2)$$

$$Pitch = \sum_{t0}^{t_n - 1} \frac{Pitch}{t_n - 1} \quad (3)$$

$$(Pitch|Low) = \sum_{t0}^{t_n - 1} \frac{Pitch}{t_n - 1} < 50(level) \quad (4)$$

$$(Pitch|High) = \sum_{t0}^{t_n - 1} \frac{Pitch}{t_n - 1} > 50(level) \quad (5)$$

$$(Hz|Low) = \sum_{1Hz}^{10KHz} X(f) > \sum_{10.1KHz}^{20KHz} X(f) \quad (6)$$

$$(Hz|High) = \sum_{1Hz}^{10KHz} X(f) < \sum_{10.1KHz}^{20KHz} X(f) \quad (7)$$

$$(Energy|Low) = \sum_{t0}^{t_n - 1} Pitch < \frac{50(t_n - 1)(level)}{2} \quad (8)$$

$$(Energy|High) = \sum_{t0}^{t_n - 1} \frac{Pitch}{t_n - 1} > \frac{50(t_n - 1)(level)}{2} \quad (9)$$

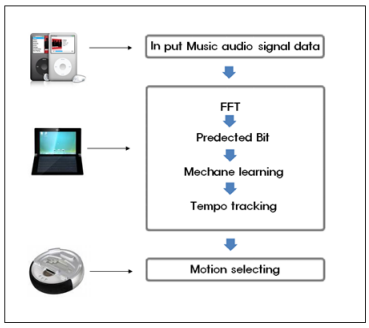
최초로 학습된 Training data 에 의해 결정된 Tree Model 의 출력 Tempo 는 Fast, Regular, Slow 3 가지 형태의 출력을 가지게 된다. 표현은 식(10)과 같이 한다.

$$Tempo = \{Fast, Regular, Slow\} \quad (10)$$

Tempo 를 이루는 각각의 요소는 Robot 의 6 개의 동작 패턴과 매칭 되어 있다.

2.1 System Description

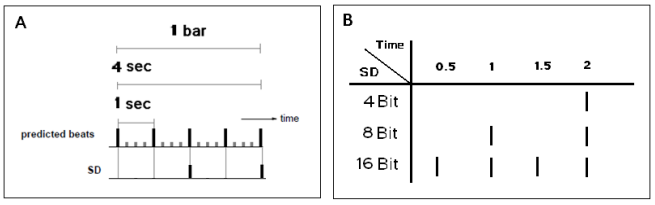
밑의(그림 1)은 전체적인 시스템 구조를 보여준다. 먼저 Audio signal 은 오디오 플레이어를 이용해 PC 에 있는 Line in 입력 단자로 입력을 받는다. 후에 PC 에서 Visual c++를 이용해 구현한 Signal 분석 툴을 사용하여 입력 받은 음악의 Beat 를 추적하고 추적된 Beat 는 기계학습의 Training data 로 사용 되어 기계 학습으로부터 얻어진 Tempo data 와 signal 의 Pitch 값을 이용해 로봇의 Motion 을 Selecting 하게 된다.



(그림 1) System architecture

2.2. Detecting SD

FFT 를 이용해 찾고자 하는 것은 Buffering 구간의 SD point 가 된다. 밑의(그림 2)의 A, Example 은 4Bit 인 음악의 SD point 를 나타낸 것이다. SD 는 4Bit 일 경우 Peak 간격이 2 초가 된다.(그림 2)의 B 는 Bit 별 SD 의 Peak 간격을 나타낸 것이다.SD 의 Peak 가 0.5 초 간격으로 발생 할 때 16Bit 가 됨을 알 수 있다.



(그림 2) Example of Detecting SD point

2.3. Tempo tracking

Frequency Analysis 로 부터 획득한 Data 들은 기계학습을 거쳐 Tempo 를 tracking 하게 된다. 사용되는 기계학습 방법은 Decision Tree 의 ID3 알고리즘을 사용하며(그림 3)은 Decision Tree 의 Training data table 이다.

Bit	Pitch	Hz	Energy	Tempo
16	High	Low	High	Fast
16	Low	High	Low	Fast
16	Low	Low	Low	Fast
4	High	Low	Low	Slow
16	High	High	High	Fast
4	High	High	Low	Slow
4	Low	High	High	Regular
8	High	Low	Low	Fast

(그림 3) Example of Training data table

2.4. Decision Tree(ID3) Algorithm

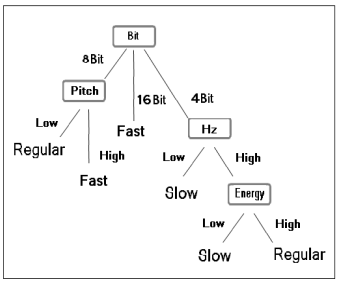
Decision Tree 는 확률적으로 정보이득(information gain)이 가장 많은 것을 root 로드로 하여 Entropy(s)가 0 이 될 때 까지 밑으로 subroot 를 형성하는 구조로 되어 있으며 목적 속성이 C 개의 서로 다른 값을 가질 수 있을 때 예제 인스턴스 집합(s)의 Entropy 는 다음과 같이 정의 된다.

$$Entropy(S) = \sum_{i=1}^{c} -p_i \log_2 P_i$$

이때 예제 인스턴스 집합(s)에서의 속성 A 에 대한 정보이득 Gain(S,A)는 다음과 같이 정의 된다.

$$Gain(S,A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

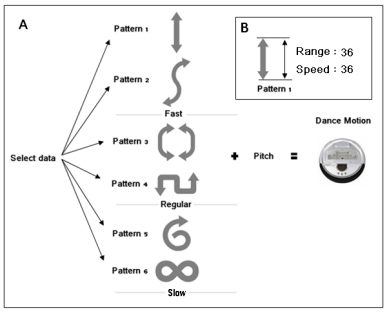
(그림 3)의 Training data 를 ID3 Algorithm 에 적용시킨 결과 생성된 Tree Model 은 (그림 5)와 같다.



(그림 5) Decision Tree of Training data

2.5. Pattern Selection

Select data 는 다음과 같이 정의 하며 Select data = Tempo + Pitch, Select data 에 의해 Robot 은 6 개의 패턴화 되어 있는 댄스 모션 중 하나를 선택 하게 된다. 선택된 패턴은 다시 Pitch 값에 의해 속도와 크기가 결정된다. (그림 6) B



(그림 6) Dance motion selecting process

3. 결론

본 논문에서는 템포 추적을 위한 버퍼링 기법 및 패턴 기반의 로봇 댄스 기법을 제안하였다. 제안하는 기법은 음악의 템포를 추적 할 수 있으며 동작 패턴 선택 기법을 이용하여 효과적인 댄스 로봇의 행위를 표현 할 수 있다.

참고문헌

[1]. Yu Shiu,Hong Jeong and C.-C. Jay Kuo, “Musical structure Analysis Using Similarity Matrix and Dynamic Programming,” Proceedings *of SPIE*, Vol. 6015, 2005

[2]. Judy Robertson, Andrew de Quincey, Tom Stapleford and Geraint Wiggins, “Real-Time Music Generation for a Virtual Environment,” *Proceedings of ECAI-98 Workshop on AI/Alife and Entertainment,* 1998