

휴대폰 카메라 기반의 악보영상인식 시스템

오성열, 손화정, 김수형
 전남대학교 전산학과
 e-mail:acecap@iip.chonnam.ac.kr

Music Image Recognition System Based on Mobile Phone Camera

Sung-Ryul Oh, Hwa-Jeong Son, Soo-Hyung Kim
 Dept. of Computer Science, Chonnam National University

요 약

우리 삶에서 빼놓을 수 없는 기기인 휴대폰의 다양한 콘텐츠 기술 개발의 필요하다. 이러한 필요성을 충족하기 위하여 휴대폰에서 얻은 저 품질 악보 영상을 인식하는 기능을 구현하고 MIDI파일로 저장된 악보를 연주하는 시스템을 제안한다. 휴대폰 카메라를 통하여 얻은 영상을 인식한 결과 평균 93.4%의 악보 인식률을 얻을 수 있었다.

1. 서론

오늘날 우리의 삶 속에서 빼놓을 수 없는 기기로 여겨지는 이동 전화는 궁극적으로 각 개인의 모든 일상을 관리하는 종합시스템화 될 것으로 사료된다. 이러한 종합시스템의 일부로 휴대폰의 카메라를 이용하여 악보 영상을 획득하고 이를 인식하여 연주하는 기능을 구현하여 기존의 휴대폰과는 차별화된 휴대폰 콘텐츠 기술을 갖출 수 있다.

악보 영상의 인식에 관한 연구는 1980년대 초부터 시작되었으나 이들 대부분은 인쇄된 악보 영상을 평판 스캐너를 통하여 획득하고 인식하였으나, 본 논문에서는 인쇄된 영상을 휴대폰에 장착된 카메라를 통하여 얻은 저품질 악보 영상의 인식에 관한 연구를 수행하고자 한다. 다양한 휴대폰 콘텐츠 기술 개발의 필요성에 충족하는 콘텐츠로서의 악보 인식 시스템 개발을 목적으로 한다. 즉 휴대폰의 카메라를 이용하여 악보 영상을 획득하고 이를 인식하는 시스템을 구현한다.

본 논문에서 사용되는 영상은 S사에서 개발한 B-3200 휴대폰을 이용하여 간단한 동요 영상을 획득하고 촬영된 영상을 가지고 실험하였다.

2. 악보 인식 시스템의 전체 구성

그림 1은 휴대폰을 통한 악보인식의 과정을 보여준다. 휴대폰을 통하여 영상을 촬영하고 악보 영상을 전처리과정을 통하여 오선과 음표 및 기호를 분리한 다음 심볼의 구조적 요소 분석 및 메쉬 특징을 이용하여 인식하였다. 마지막으로 인식한 결과를 토대로 MIDI파일을 생성한 후

미디를 재생하고자 한다.

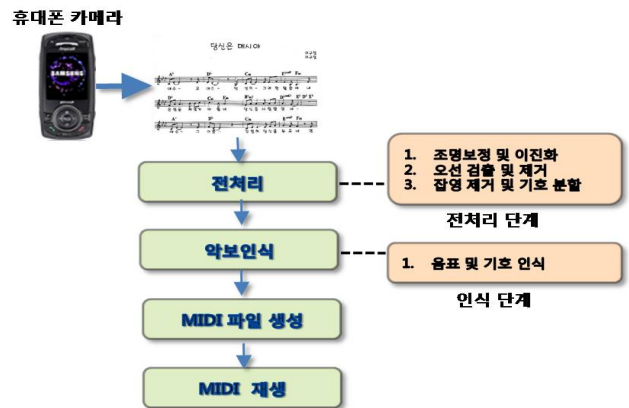


그림 1. 악보 인식 시스템의 구성

3. 전처리

3.1 조명 보정 및 이진화

휴대폰 카메라로 촬영한 경우에는 해상도가 낮을 뿐만 아니라 불규칙적인 조명의 영향을 받아 영상의 품질이 좋지 못하다. 이러한 경우, 단순히 임의의 임계값으로 이진화를 수행하는 일반적인 전역적 이진화 알고리즘을 적용하는 것은 바람직하지 않다. 휴대폰으로 촬영한 영상은 조명에 민감하기 때문에 이러한 영상을 휘도영역에서 32개의 채널로 만들고 그것을 분리하는 방법을 이용하였다. 이 이진화의 결과 영상은 그림 2와 같다.

3.2 오선 검출 및 제거

평판 스캐너가 아닌 휴대폰 카메라를 통하여 영상을 획

특하기 때문에 기울어짐이 존재하게 되어 수평방향 히스토그램만을 이용한 오선 검출을 이용하는 것은 적절하지 못하다. 높은 인식 성능을 위해서는 정확한 오선 검출 및 삭제 알고리즘이 요구된다. 기울어진 오선의 영역 검출 알고리즘 및 오선 삭제 알고리즘을 이용하여 오선을 삭제한다[8].

3.3 잡영 제거 및 기호 분할

오선을 삭제한 뒤에도 약간의 잡영이 존재하게 되는 데 이를 추가적으로 제거하여 보다 정확한 음표 및 기호를 검출한다. 이를 위하여 먼저 연결요소 분석을 통하여 검은 화소로 연결되어 있는 요소들을 각각 하나의 객체로 묶고 추출된 객체들 중에서 너비가 오선 간격의 1/4보다 작을 경우 잡영으로 판단 삭제한다. 결과 영상은 그림 3과 같다.

4. 악보 인식

악보 인식 과정은 계층적으로 구성된다. 즉, 머리와 수직선 성분의 개수를 기준으로 대분류에서 10가지로 구분짓는다. 소분류에서는 10가지로 분류된 각 항목별로 머리나 꼬리의 유무 등을 살펴봄으로써 다양한 음표 및 기호들을 인식한다.

4.1 대분류

대분류는 심볼(symbol)의 머리와 수직선의 개수를 기준으로 표 1과 같이 10 가지 경우로 구분된다. 여기서 nH는 머리의 개수이고 nL은 수직선의 개수를 나타낸다.

4.1.1 수직선의 개수

수직선은 오선 간격과 심볼에 대한 수직방향 히스토그램 값을 이용하여 수직선을 찾고 심볼 영역 내에서 직선의 시작과 끝 사이에 연결요소가 몇 개 있는지를 검사하여 수직선의 개수 계산한다.

4.1.2 머리의 개수

머리는 오선 간격과 심볼에 대한 수직방향 히스토그램

값을 이용하여 계산한다. 여기서는 먼저 머리가 수직선 왼쪽에 있는지 오른쪽에 있는지를 구분한 뒤 머리의 개수를 계산한다. 앞서 계산한 심볼의 수직선 위치에서 처음 만나는 수직선의 위치가 오선 간격보다 크다면, 이 심볼의 머리는 왼쪽 아래에 위치하고 그렇지 않은 경우는 오른쪽 위에 위치한다고 판단한다.

표 1. 대분류의 종류 및 판별 기준

분류	이름	판별기준
1	CLEF	위치
2	NORMAL_NOTE	$nL==1 \ \& \ nH==1$
3	BEAM_NOTE	$nL==nH \ \& \ nL \geq 2$
4	FULL2_NOTE	$nL==1 \ \& \ nH==0$
5	SHARP_NATURAL	$nL==2 \ \& \ nH==0$
6	NO_LINE	$nL==0$
7	MULTI_NOTE	$(nL < nH \ \& \ nL == 1) \ \ (nL == 2 \ \& \ nH == 1)$
8	MULTI_BEAM_NOTE	$nL < nH \ \& \ nL \geq 2$
9	MEASURE	위치
10	CODE	위치

4.1.3 심볼의 오선 상의 위치 결정

심볼의 오선 상의 위치는 그림 4에 나타난 것처럼 숫자로 표기한다. 음표의 경우에는 머리의 중심에 해당하는 위치를 찾고, 기호의 경우에는 심볼의 높이의 중심에 해당하는 위치를 찾는다. 위치를 찾는 방법은 다음과 같다.

먼저, 심볼이 오선의 위쪽, 오선, 또는 아래쪽 중 어디에 위치해 있는지를 검사한다. 만일 심볼과 가장 가까이 있는 오선이 첫 번째 오선이고 그 차이가 (오선 간격×1/4)보다 크면서, 심볼의 위치가 첫 번째 오선의 시작점보다 위쪽에 있으면, 심볼은 오선의 위쪽에 위치해 있다고 판단한다. 그렇지 않고, 심볼과 가장 가까이 있는 오선이 다섯 번째 오선이면, 그 차이가 (오선 간격×1/4)보다 크면서, 심볼의 위치가 마지막 오선의 시작점보다 아래에 있으면, 심볼은 오선의 아래쪽에 위치해 있다고 판단한다. 위의 두



그림 2. 삼성 애니콜 B3200으로 촬영한 영상

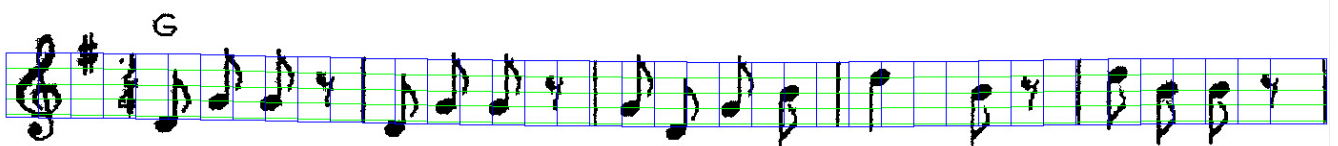


그림 3. 전처리 후 영상

가지 경우가 아닐 경우에는 오선 내에 위치해 있다고 판단한다. 심볼의 오선의 어느 쪽에 위치하는지를 찾은 다음, 그림에서처럼 정확하게 어느 위치에 걸쳐있는지를 숫자로 나타낸다.

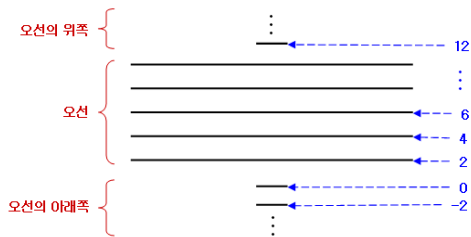


그림 4. 심볼의 오선 상의 위치 표기 방법

4.2 소분류

4.2.1 음자리표 (CLEF)

음자리표 중에서도 높은음자리표(G-clef)는 보표의 맨 좌측에 나타나며 그 폭이 다른 기호들에 비해 큰 특징을 갖는다. 즉, 심볼의 높이가 (오선 전체의 높이+오선 간격×2)보다 크고, 심볼의 중심 위치가 오선의 중앙에 위치하면 높은음자리표로 인식한다. 여기서 오선의 중앙은 5개의 선 중 위와 아래의 직선 2개를 제외한 가운데 3개의 직선 내부를 의미한다.

4.2.2 기본 음표 (NORMAL_NOTE)

기본 음표는 머리와 수직선의 개수가 각각 1개인 경우를 말한다. 기본 음표는 다시 4분 음표, 8분 음표, 16분 음표의 세 가지 종류로 소분류 된다. 이들 음표는 꼬리의 수를 기준으로 구분할 수 있다. 머리를 인식하는 방법은 앞서 설명한 것과 동일하게 수행하며, 음표의 꼬리는 꼬리 부분에 해당하는 영역을 연결요소 분석을 통해 몇 개인지를 검사한다.



그림 5. NORMAL_NOTE의 종류와 소분류

4.2.3 대로 연결된 기본 음표 (BEAM_NOTE)

대로 연결된 기본 음표는 수직선의 수와 머리의 수가 같거나 수직선의 수가 머리의 수보다 더 많을 경우에 해당하는 것으로, 꼬리를 판단하는 방법은 앞서 설명한 것과 동일하며, 점음표를 인식하는 방법은 직선 성분이 없는 기호에 기술되어 있다.



그림 6. BEAM_NOTE의 예

4.2.4 2분 음표, 플랫, 마디선 (FULL2_NOTE)

2분 음표, 플랫(flat), 마디선은 대분류에서 수직선이 1

개이고 머리가 없는 경우에 해당하는 심볼들이다. 여기서 플랫은 심볼의 높이를 기준으로 쉽게 구분이 가능하며, 반대로 마디선은 심볼의 너비를 기준으로 구분이 가능하다. 심볼의 높이가 (오선 간격×2.5) 보다 작다면 플랫으로 인식한다. 그리고 심볼의 너비가 (오선 간격×1/2) 보다 작거나 또는 심볼의 검은 화소 비율이 0.7 이상일 경우에는 마디선으로 인식한다. 여기서, 심볼의 검은 화소의 비율 (PixelRatio) 은 다음과 같이 계산한다. n 은 심볼 내에 있는 검은 화소의 수를 말하고 W 와 H 는 심볼의 너비와 높이를 나타낸다.

$$\text{PixelRatio} = \frac{n}{W \times H}$$

심볼이 플랫과 마디선으로 분류되지 않은 경우에는 2분 음표라고 인식한다. 2분 음표일 경우 머리의 위치를 계산하여 위치를 확인한다.

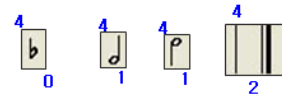


그림 7. FULL2_NOTE의 대분류 및 소분류

4.2.5 샵, 제자리표 (SHARP_NATURAL)

샵(sharp)과 제자리표는 수직선이 2개이고 머리가 없는 경우를 말한다. 두 수직선의 시작 좌표의 차이가 (오선 간격×1/3)보다 작거나 같으면 샵으로 인식하고 그렇지 않을 경우에는 제자리표로 인식한다.

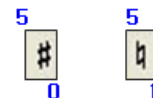


그림 8. SHARP_NATURAL의 대분류 및 소분류

4.2.6 직선 성분이 없는 기호 (NO_LINE)

대분류에서 머리와 수직선이 없는 경우를 직선 성분이 없는 심볼로 분류한다. 직선 성분이 없는 기호에는 그림 9와 같이 온음표와 점음표, 그리고 4분 쉼표, 16분 쉼표 등과 같은 다양한 종류의 쉼표들이 존재한다.

심볼의 높이와 넓이가 (오선 간격×4/5) 이하인 경우를 점음표(dot)로 인식한다. 점음표로 인식되지 않은 다른 심볼들 중 기호의 높이가 (오선 간격×3/4)보다 작거나 같으면서 너비가 오선 간격 이상인 경우를 다음과 같이 온음표와 온쉼표, 2분 쉼표의 세 가지 경우 중 하나로 분류한다.

먼저 심볼의 검은 화소의 비율이 0.7보다 작으면 온음표로 분류한다. 그렇지 않을 경우, 심볼의 상·하의 검은 화소의 개수를 비교하여 위쪽에 있는 검은 화소의 수가 아래쪽에 있는 검은 화소의 수보다 크거나 같으면 2분 쉼표, 그렇지 않을 경우에는 온 쉼표로 구분한다.

점음표, 온음표, 온쉼표, 2분 쉼표로 분류되지 않은 경우에는 심볼의 높이의 비율에 따라 4분 쉼표, 8분 쉼표, 16분 쉼표 중 하나로 분류된다.

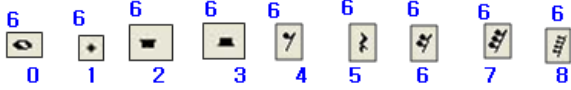


그림 9. NO_LINE의 대분류 및 소분류

4.2.7 단순 겹음표 (MULTI_NOTE)

단순 겹음표는 수직선의 수가 1개이면서 머리의 수가 수직선의 수보다 큰 경우(즉, 2개 이상인 경우) 또는 머리의 수가 1개이면서 수직선의 수가 2개인 경우를 말한다. 그림 10은 이들 두 가지 조건에 해당하는 예를 보여준다. 단순 겹음표를 인식하는 방법은 기본 음표를 인식하는 방법과 동일하다.



그림 10. MULTI_NOTE의 대분류

4.2.8 대로 연결된 겹음표 (MULTI_BEAM_NOTE)

대로 연결된 겹음표는 머리의 수가 수직선의 수보다 많으면서 수직선의 수가 2개 이상인 경우로 그림 11과 같은 형태를 말한다. 인식하는 방법은 기본 음표의 인식 방법과 동일하다.



그림 11. MULTI_BEAM_NOTE의 대분류

4.2.9 박자 (MEASURE)

박자는 일반적으로 높은음자리표 또는 플랫이나 샵과 같은 조표 다음에 위치하는 특징을 갖는다. 따라서 이전 심볼이 높은음자리표인지 또는 조표인지를 확인하고, 동시에 크기 정보를 비교하여 현재 심볼이 박자인지 아닌지를 판단한다. 현재 심볼이 박자일 경우 숫자를 인식한다. 인식 방법은 다음과 같다.

박자로 인식된 심볼에 대해, 3×4의 12차원 메쉬(mesh) 특징을 이용하여 인식한다. 먼저 심볼 높이의 중간 위치를 기준으로 상위 영역과 하위 영역으로 구분하고, 각 영역별로 메쉬 특징을 추출한다. 일반적으로 상위 영역은 2, 3, 4, 6의 숫자가 나올 수 있고, 하위 영역은 2, 4, 8이 나올 수 있다. 미리 정의된 숫자 특징과 비교하여 최소 거리 분류를 통해 해당 숫자를 인식한다.

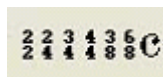


그림 12. MEASURE의 종류

4.2.10 코드 (CODE)

코드는 보표에 상위 영역에 나오게 되는데 이러한 위치 정보를 이용하여 코드가 나오는 영역을 인식하고 영역에 들어온 심볼을 메쉬특징을 이용하여 특징을 추출하고 최

소거리 분류를 통하여 인식한다. 동요수준에서 나오는 대문자 알파벳 7개 (C,D,E,F,G,A,B)와 샵·플랫 및 숫자 (4,6,7)로 이루어진 코드만을 인식한다.



그림 13. CODE의 대분류 및 소분류

5. 실험 결과

5.1 동요 영상의 인식결과

다음 표 2는 B3200으로 촬영한 영상을 인식한 결과를 나타낸다. 코드 인식시에 붙어 있는 코드 및 코드영역에 들어오지 않은 심볼은 미인식 하였다. 9종의 동요 영상에 대하여 평균 인식률을 계산한 결과 93.4%의 인식률을 얻을 수 있었다.

$$\text{인식률} = (\text{심볼의 수} - \text{오인식}) / \text{심볼의 수} * 100 (\%)$$

표 2. 동요 영상에 대한 인식 결과

제목	심볼의수			오인식			인식률
	음표	기호	코드	음표	기호	코드	
애국가	56	25	0	3	0	0	96
도레미파	41	16	18	8	1	0	88
작은별	42	25	25	3	0	0	97
고기잡이	47	33	8	2	1	0	97
고드름	33	20	16	5	0	0	93
개구리유치원	38	20	15	6	1	0	90
개구리	72	31	0	11	1	0	88
가을밤	50	29	28	4	0	0	96
꽃밭에서	95	37	38	3	3	0	96

5.2 인식 결과의 표기 형태

그림 14는 가을밤이라는 제목을 가지는 동요영상의 첫 번째 오선단을 보여준다. 지금까지 서술한 방법으로 그림 14의 영상에 적용한 결과 의해 심볼의 대분류와 소분류를 결정하고, 음표의 경우에는 음의 높이까지를 판단하여 그림 15와 같이 결과를 출력한다.

5.3 오인식의 예

그림 16은 동요수준에 영상을 인식하는 과정에서 발생



그림 14. 삼성 애니콜 B3200 카메라로 획득한 영상

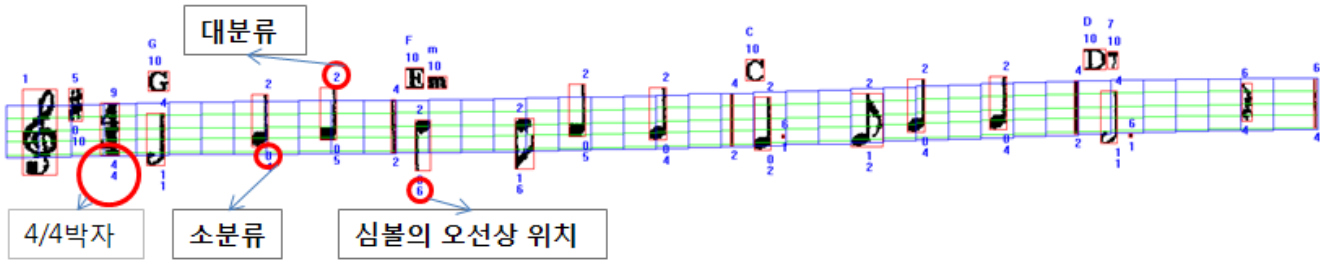


그림 15. 인식 후 결과영상

하는 오인식 결과의 일부분 중에 하나의 영상이다. 심볼이 연결된 경우나, 심볼이 오선 삭제 시에 끊어진 경우에 오인식되는 것을 알 수 있었다.

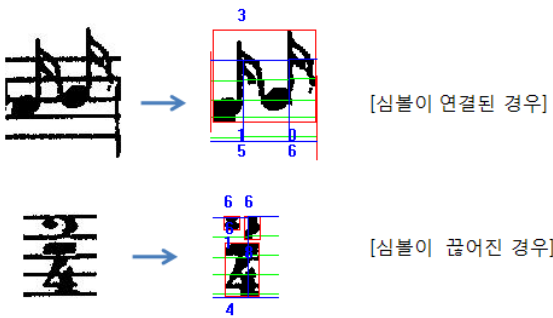


그림 16. 이진화 및 오선 삭제 과정에서 심볼이 연결되거나 끊어지는 경우

6. 결론

본 논문에서는 삼성전자 휴대폰-B3200 카메라에서 촬영한 저품질 악보 영상을 자동으로 인식하고 연주하는데 필요한 기술 방법을 연구하였다. 오선을 검출하기 위해 기울어짐 정도에 따라 서로 다른 개수로 보표를 분할하고, 분할한 각 부분에 대해서 오선을 검출한다. 보표를 분할하는 것은 기울어짐이나 휘어짐이 거의 없도록 하는 것으로, 별도의 기울어짐 보정 과정이 필요하지 않다. 실험을 휴대폰 카메라를 이용해 촬영할 때 발생할 수 있는 기울어짐이나 휘어짐 정도는 충분히 다를 수 있다고 판단된다.

인식과정에서는 심볼이 가지는 구조적 요소를 머리나대의 수를 가지고 인식하였으며, 코드 및 심볼이 연결된 경우 오인식하는 경우가 있었다. 이러한 경우에 인식률을 높일 수 있도록 붙어있는 심볼을 분리하는 알고리즘 연구나 떨어진 기호를 앞 요소와 비교하여 연결하는 과정을 수행하게 된다면 더 좋은 결과가 기대된다.

기존 악보 영상 인식에 관한 연구는 휴대폰 카메라가 아닌 스캐너로부터 얻은 영상에 대한 것으로 그 실효성에

있어 많은 제약 사항이 따르는 단점이 있다. 반면, 제안 방법은 현대 사회에서 누구나 가지고 있는 휴대폰 카메라를 이용한 것으로 휴대폰에서의 악보 인식 및 연주시스템은 새롭고 다양한 휴대폰 콘텐츠 기술 개발의 필요성에 충족할 수 있을 것이라 본다.

참고문헌

- [1] Wellner Pierre D. Wellner, "Adaptive Thresholding for the DigitalDesk," EuroPARC Technical Report EPC-93-110, Rank Xerox Research Centre, Cambridge, UK, 1993.
- [2] 유휘상, 최우영, "수리형태학을 이용한 인쇄악보 인식," 대한전자공학회 학술대회논문집, 제17권, 제1호, pp. 713-717, 1994.
- [3] 박충식, 장경식, 전정우 등, "악보 인식을 위한 전처리 과정," 인공지능, 신경망 및 퍼지관련 학술발표회 논문집, pp. 169-175, 1993.
- [4] 김정규, "음악기호의 계층구조를 이용한 인쇄악보 인식," 한국화상학회지, Vol.3, NO.1, pp.66-75, 1997.
- [5] 박철우, 항인성, 정동석, "인쇄 악보의 자동 인식 및 연주를 위한 연구," 대한전자공학회 춘계학술대회 논문집, 제16권, 제2호, pp.653-656, 1993.
- [6] H. Miyao and M. Okamoto, "Stave Extraction for Printed Music Scores Using DP Matching," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 8, No. 2, pp. 208-215, 2004.
- [7] D. Bainbridge and T. C. Bell. "The Challenge of Optical Music Recognition" *Computers and the Humanities*, Vol. 35. No. 2, pp. 95 - 121, 2001.
- [8] S.H. Kim, H.J. Son, S.R. Oh, G.S. Lee and C.W. Lee, "Staff-Line Detection and Removal Algorithm for a Camera-Based Recognition of Music Score Images" *Proceedings of the 2nd Korean-Japan Joint Workshop on Pattern Recognition*, 2007 (발표예정).