

# 악보 인식을 위한 잡음 제거 및 분류 알고리즘 비교

제승완, 이덕우\*  
계명대학교 컴퓨터공학과

## Noise Removal and Comparison of Classification Algorithms for Music Score Recognition

Seung-Wan Je, Deokwoo Lee\*  
Department of Computer Engineering, Keimyung University

**요약** 본 논문에서는 악보 인식의 전처리 과정 중 잡음 제거와 전처리 이후 분류에 사용하게 될 분류 알고리즘 적용 및 그 결과를 비교한다. 음악 악보에서의 음표 인식을 수행에서 가사, 제목, 작성자 정보는 음표인식 작업에 도움을 주지 않는 잡음으로 분류할 수 있으므로 음표 인식 작업을 하기 전에 이 정보들을 제거하는 작업이 필요하다. 잡음 제거에는 형태학 연산과 윤곽선 탐지 알고리즘을 사용한다. 음표 분류를 수행하기 전에 분류를 위한 특징점 검출을 위해 ORB와 SIFT가 사용되며 특징점 정합을 위해 영상의 스케일링을 다르게 하면서 특징점 검출 및 정합 결과를 비교한다. 비교 방법에서는 크기 변화에 따른 차이를 알아보기 위해 각 원본 영상에 다른 배율을 적용하며, 배율마다 양선형(Bilinear), 바이 큐빅, Lanczos 보간법을 적용해 보간법 알고리즘의 영향도 함께 비교한다. 잡음 제거는 제안한 방법을 통해 맨눈으로도 대부분의 잡음이 제거된 것을 확인할 수 있고, 분류에서는 다양한 방식의 비교를 통해 SIFT 방식이 ORB 방식보다 음표 인식에 더욱 적합한 것을 알 수 있다.

**Abstract** This study compared the classification algorithms used for noise removal required for the preprocessing of music recognition and for classification after preprocessing. Some elements, such as titles, lyrics, and authors, are unnecessary for music recognition. The morphological operations and contour detection algorithms were used to eliminate the noise. Once noise removal was performed, the musical notes were classified with feature extraction. This paper also compared the classification result using the ORB or SIFT algorithm. In the comparison work, various scaling factors were applied to each original image to determine the difference according to the size of the images. The effects of the interpolation method were also analyzed. The proposed noise removal method can confirm that most of the noise has been removed even with the naked eye. In the classification, various comparisons show that the SIFT algorithm is more suitable for note recognition than the ORB algorithm.

**Keywords** : Musical Note, Classification, Feature Extraction, Noise Removal, Recognition

### 1. 서론

최근 영상처리 및 인공지능 분야는 전통적인 영상처리 뿐 아니라 실시간 영상, 의료 영상, 위성 영상 등 광대한

범위로 확장되고 있다[1-4]. 음악 분야에서의 영상처리 기술 적용은 대표적으로 악보 인식이 있는데 이는 1960년대부터 꾸준히 연구되어왔고, 2000년대에 들어 인공지능 기술이 활발히 연구됨에 따라 인공지능을 활용한

이 논문은 2023년도 정부(교육부)의 재원으로 한국연구재단 기초연구사업의 지원을 받아 수행된 연구임.(No. 2022R1I1A3069352)

\*Corresponding Author : Deokwoo Lee(Keimyung Univ.)

email: dwoolee@kmu.ac.kr

Received February 8, 2023

Accepted June 2, 2023

Revised March 29, 2023

Published June 30, 2023

악보 인식 기술들도 활발히 연구되고 있다. 예를 들어 Kim[5]은 SOM과 개선된 ART-1 알고리즘을 악보 영상에 적용하여 인식을 진행하였으며, Trieu 등은 U-net을 이용하여 악보 영상의 오선을 제거하는 방법을 제안하였다[6]. 일반적으로 인공지능을 활용한 악보 인식의 과정은 크게 전처리, 분류, 학습 3가지 단계로 나누어진다. 이 중 전처리 과정에는 오선 제거와 잡음 제거, 음표 추출 등이 있는데 이 과정이 필요한 이유는 악보 영상에는 일반적인 영상보다 제목, 가사, 오선 및 여러 가지 기호 등 다양하고 복잡한 패턴 등 고려해야 할 점이 많기 때문이다. 그러므로 악보에서 음표를 분류 및 인식하기 전에 오선 제거와 잡음 제거, 음표 추출 등의 전처리 과정을 수행할 필요가 있다. 전처리 과정을 수행한 후 분류 및 학습에 불필요한 데이터를 사전에 제거하여 악보인식 오류의 확률을 줄일 수 있다. 기존의 잡음 제거 방법에는 글래스 파이어(glass fire) 알고리즘을 이용하여 잡음을 제거하는 방법이 있다[7]. 글래스파이어 알고리즘은 자기 호출을 이용하여 모든 주변 요소가 라벨링 될 때까지 현재 관심 화소의 주변 화소를 차례로 검사하면서 라벨링하는 방법이다. 다음 과정인 분류 과정은 악보를 인식하는데 필요한 기호와 비기호, 박자를 분류하여 음표 인식에 필요한 데이터와 곡 연주에 필요한 데이터를 분류하는 작업을 한다. 기존의 음표 분류 방법에는 음표의 수직 성분과 크기를 이용하여 음표와 비 음표를 분류하며 음표는 음표 머리의 크기, 꼬리 유무, 빔(beam)의 유무를 이용하여 각각 분류하였다[8]. 음표 분류에는 다양한 알고리즘이 사용되는데 데이터에 적합한 알고리즘을 선택하여야 좋은 결과를 얻을 수 있기에 음표 분류에 있어 좋은 성능을 보이는 알고리즘을 찾는 필요가 있다. 다음으로 학습의 단계에서는 분류된 음표 및 기호를 학습해 악보 영상에서 별도의 전처리 없이 음표를 추출하거나, 악보의 기호 및 음표의 음계를 인식하여 입력된 악보를 통해 해당 곡을 자연스럽게 연주할 수 있게 학습한다[7]. 음악 분야에서도 인공지능 기술이 도입되어 작곡, 보컬 음계 인식 등에 적용되고 있다. 본 연구에서는 음계 인식을 위해 필요한 과정인 음표인식을 위한 분류 결과를 SIFT와 ORB 방법 비교를 수행한다. 본 논문에서는 이러한 악보 인식에 기반이 되는 전처리 과정 중 오선 제거 후 추출된 음표 및 기호를 이용하여 악절의 영역을 구하고, 해당 영역을 제외한 다른 영역을 비워주는 방식의 잡음 제거를 수행한 후, 다양한 분류 알고리즘 중 SIFT[9]와 ORB[10] 알고리즘에 다양한 방법을 사용하여 어느 것이 음표 분류에 있어 정확도와 신뢰도가 더 좋은지 결

과를 비교 분석한다.



Fig. 1. Example musical-score with unnecessary noise

## 2. 잡음 제거

일반적인 악보 영상에는 Fig. 1과 같이 오선, 음표, 기호 이외에도 제목, 가사, 저자와 같은 악보 인식 및 음표 추출에 불필요한 정보가 다수 첨부되어 있다. 이러한 잡음들을 미리 제거하지 않으면 추후 음표를 추출하거나 학습할 때 연산 시간이 더욱 길어질뿐더러 분류 및 학습 과정에서 오류가 발생할 수 있기에 사전에 제거해 주어야 한다. 제거해야 하는 잡음들로는 크게 오선, 제목, 가사가 있다. 이 중 오선의 경우에는 수평 히스토그램을 활용하여 제거할 수 있다[11]. 다음으로 잡음 중 가사 및 제목을 제거하기 위해 악보 인식에 필요한 음표 영역과 가사 영역을 분리해야 한다. 먼저 악보 영상은 겹으로 보기에는 이진 영상이지만, 기본적으로 3개의 채널을 가지고 있는 RGB 영상이다. 그렇기에 연산 속도의 가속화 및 연산 정확성을 위해 먼저 그레이 스케일(Gray Scale) 영상으로 변환 후 오프(OTSU) 알고리즘을 이용하여 영상의 모든 픽셀값을 이진화한다. 분리하는 방법은 오선이 존재하며 이진화가 적용된 악보 영상에 형태학 연산의 침식(erosion)과 팽창(dilation) 연산을 적용한다. 이때 커널의 크기는 침식 연산의 경우  $3 \times 3$ 으로 하고, 팽창 연산의 경우  $1 \times 5$ 로 한다. 형태학 연산의 침식 연산을 통해 악보 인식에 필요한 정보들이 모두 연결되게 되고, 연결된 영상의 윤곽선을 탐색하여 Fig. 2와 같이 가사, 제목을 제외한 악절의 영역을 탐지한다. 이때 악절의 영역을 크게 벗어난 음표 및 기호들이 일부 존재하는데 이를 악절의 영역에 포함하는 작업이 필요하다. 먼저 오선이 없는 악보 영상에서 윤곽선 탐지를 통해 음표 및 기

호들의  $x, y, w, h$  정보를 추출한다. 악절의 범위를 벗어난 음표 및 기호들을 악절의 영역에 포함하는 방법은 탐지된 악절 영역의  $y$ 좌표와 탐지된 음표 및 기호들의 영역을 비교하며 악절의 영역을 확장함으로써 음표 및 기호들이 악절의 영역에 포함된다. 확장은 두 정보의  $y$ 와  $h$ 만을 이용하여 확장하게 되는데 그 이유는 음표들은 보통 악절의 영역 안쪽에 존재하기 때문에  $x$  및  $w$ 정보는 고려할 필요가 없다. 악절의 영역은 다음 두 가지 조건에 만족할 때 확장된다.

조건 1.  $Sb_y < Sf_y$  and  $(Sf_y + Sf_h) < (Sb_y + Sb_h)$

조건 2.  $(Sb_y < Sb_h) > (Sf_y < Sf_h)$  and  $(Sf_y < Sf_h) > Sb_y$

$Sb$ 와  $Sf$ 는 각각 음표의 영역과 악절의 영역 정보를 가지고 있다. 조건 1에 만족하면 악절의  $y$ 영역이 확장되며 조건 2를 충족하면 악절의  $y+h$ 영역이 확장되고 이것은 Fig. 3을 통해 확인할 수 있다. 마지막으로 확장된 영역을 제외한 나머지 영역의 픽셀값들을 모두 255(하얀색) 값으로 변경해두면 잡음 제거가 완료된다.



Fig. 2. Example of an region of a detected period



Fig. 3. Extended period region



Fig. 4. Input Image Example, From the left, each bit is listed in the order of template and extraction

### 3. 음표 분류

본 논문에서는 무료 음표 영상과 다양한 악보에서 오선 제거 및 잡음 제거 과정을 거친 후 추출된 음표 영상을 사용한다. 비교하는 음표의 종류는 Fig. 4와 같이 서로 모양이 다른 4분음표 2개, 8분음표 2개, 빈 8분음표 2개로 템플릿과 추출된 음표 모두 합쳐 총 12개의 음표를 비교한다. 추출된 음표들은 모두 그레이 스케일로 변환된 후 오츠 알고리즘을 이용하여 이진화한다. 비교 알고리즘을 적용하는 방법에는 네 가지 방법이 있으며, 각 방법에서 크기 변화에 따른 차이를 알아보기 위해 원본 영상 크기의 5배, 7배 9배 크기의 배율을 적용한다. 첫 번째 비교 방식은 템플릿과 추출된 음표 모두 원본 크기에서 5배, 7배, 9배 확대한 비교이며, 두 번째 비교 방식은 템플릿에는 변화를 주지 않고, 추출된 음표에만 원본 크기에 각 배율을 적용하여 비교하는 것이다. 세 번째 비교 방식은 추출된 음표를 템플릿 크기에 맞춰 일차적으로 확대한 후 두 영상을 각 배율로 확대하여 비교하는 것이며, 네 번째 비교 방식은 템플릿을 추출 음표 크기로 먼저 확대해 준 후 각 배율로 확대하여 비교하는 것이다. 확대 방법 또한 두 가지로 나누었는데 한번에 확대하는 방법과 10번에 나누어 점차 확대하는 방법이다. 이 방법을 적용하는 이유는 배율의 적용할 때 사용되는 보간법들이 확대된 영상에 어떠한 영향을 미치는지 알아보기 위해서이다. 사용되는 보간법으로는 양선형[12] 보간법, 바이 큐빅[13] 보간법, Lanczos[14] 보간법을 사용한다. SIFT와 ORB 등의 비교 알고리즘은 두 영상 간의 특징점을 추출하여 선으로 연결해두는데 이때 올바른지 않은 특징점도 같이 연결된다. 그러므로 올바른 특징점을 찾아 줄 필요가 있다. 본 논문에서는 올바른 특징점을 찾기 위해 추출된 특징점의 Keypoint의 좌표가 각 영상에서 차지하는 비율을 계산하고 두 비율의 차가 15% 이내일 경우를 올바른 특징점으로 한다.



Fig. 5. Examples of a music notes with noise removal

## 4. 실험 결과

### 4.1 잡음 제거

본 논문에서는 악보 인식에서 필요하지 않은 요소인 제목, 가사, 저자 등을 형태학 연산과 윤곽선 탐지 알고리즘을 활용하여 효과적으로 제거하였다. 위 Fig. 5(A)는 위에서 제안한 잡음 제거 방법을 Fig. 1 악보 이미지의 잡음을 제거한 후 출력으로 나온 악보이다. Fig. 5를 보면 Fig. 1과 비교해 볼 때, 제목, 가사 등이 성공적으로 제거되었음을 확인할 수 있다.

template/	template/
4_1.jpg:((78, 30))	4_1.jpg:(80, 30)
4_2.jpg:((78, 29))	4_2.jpg:(80, 30)
8b_1.jpg:((87, 72))	8b_1.jpg:(90, 80)
8b_2.jpg:((78, 73))	8b_2.jpg:(80, 80)
8_1.jpg:((78, 51))	8_1.jpg:(80, 60)
8_2.jpg:((78, 32))	8_2.jpg:(80, 40)
compare/	compare/
4_1.jpg:((37, 13))	4_1.jpg:(40, 20)
4_2.jpg:((38, 13))	4_2.jpg:(40, 20)
8b_1.jpg:((34, 37))	8b_1.jpg:(40, 40)
8b_2.jpg:((35, 32))	8b_2.jpg:(40, 40)
8_1.jpg:((38, 19))	8_1.jpg:(40, 20)
8_2.jpg:((33, 19))	8_2.jpg:(40, 20)

Fig. 6. Size of images before (left) and after (right) adding margins

### 4.2 분류

본 논문에서는 Python과 OpenCV를 사용하여 코드를 작성하고 비교하였다. 비교에 앞서 Fig. 4를 통해 템

플릿과 추출된 음표 영상은 모두 다른 크기를 가지고 있다는 것을 알 수 있다. 확대는 원 상태 그대로 10번에 나누어 확대하게 되는데 이 경우, 목표로 하는 크기를 10으로 나누어서 나오는 값이 소수일 경우 소수점 아래가 잘리게 되고, 최종 크기가 줄어들어 영상이 일그러지는 현상이 발생한다. 이를 방지하기 위해 Fig. 6의 우측처럼 각 음표 영상의 크기를 모두 10의 배수가 될 수 있도록 여백을 넣어준다. 이때 목표로 하는 음표 영상이 가운데 올 수 있도록 상단과 하단, 좌, 우 균등하게 여백을 넣어준다. Table 1, 2, 3, 4는 네 가지 방식에 따른 ORB, SIFT를 통해 추출된 특징점들의 추출 결과를 보여주는 표이다. Table 1은 템플릿과 추출된 음표 모두 원본 크기에 5배, 7배, 9배 배율을 적용하는 첫 번째 비교방식의 결과를 나타내며, Table 2는 템플릿에는 변화를 주지 않고, 추출된 음표에만 원본 크기에 각 배율을 적용하는 두 번째 비교방식의 결과를 나타낸다. Table 3은 추출된 음표를 템플릿의 크기에 일차적으로 맞춘 후 각 배율을 적용하는 세 번째 비교방식의 결과를 나타내며, Table 4는 반대로, 템플릿을 추출된 음표의 크기에 일차적으로 맞춘 후 각 배율을 적용하는 네 번째 비교방식의 결과를 나타낸다. 각 Table에서 Total은 해당 확대 방법에서 ORB, SIFT를 이용하여 추출된 특징점들의 전체 수를 의미하며, Good은 추출된 전체 특징점 중 올바른 특징점의 수를 나타낸다. 또 ORB, SIFT의 Total과 Average는 해당 비교방식 및 확대 방법에서의 12 영상의 특징점 합계와 평균을 나타낸다. 먼저, Table 1을 보면 보간법마

Table 1. Compare ORB, SIFT results for the first comparison method

		Five Times				Seven Times				Nine Times				Average	
		Once		Several		Once		Several		Once		Several		Average	
		Total	Good	Total	Good	Total	Good	Total	Good	Total	Good	Total	Good	Total	Good
Bilinear Interpolation	ORB Total	52	20	68	25	163	52	207	72	252	77	330	101		
	SIFT Total	75	46	59	30	109	61	86	45	129	60	108	51		
	ORB Average	8.66	3.33	11.33	4.16	27.16	8.66	34.5	12	42	12.83	55	16.83	29.78	9.64
	SIFT Average	12.5	7.66	9.83	5	18.16	10.16	14.33	7.5	21.5	10	18	8.5	15.72	8.14
Cubic Interpolation	ORB Total	49	20	69	31	188	74	206	72	281	106	328	105		
	SIFT Total	71	36	91	37	117	57	142	54	128	43	167	53		
	ORB Average	2.1	3.33	11.5	5.16	31.33	12.33	34.33	12	46.83	17.66	54.66	17.5	30.13	11.33
	SIFT Average	11.83	6	15.16	6.16	19.5	9.5	23.66	9	21.33	7.16	27.83	8.83	19.89	7.78
Lanczos Interpolation	ORB Total	65	20	58	27	210	73	218	85	323	111	345	131		
	SIFT Total	86	37	92	35	126	49	122	56	139	57	123	51		
	ORB Average	9.33	3.33	9.66	4.5	35	12.16	36.33	14.16	53.83	18.5	57.5	21.83	33.61	12.41
	SIFT Average	14.33	6.16	15.33	5.83	21	8.16	20.33	9.33	23.16	9.5	20.5	8.5	19.11	7.91

Table 2. Compare ORB, SIFT results for the second comparison method

		Five Times				Seven Times				Nine Times				Average	
		Once		Several		Once		Several		Once		Several		Average	
		Total	Good	Total	Good	Total	Good	Total	Good	Total	Good	Total	Good	Total	Good
Bilinear Interpolation	ORB Total	0	0	0	0	0	0	0	0	0	0	0	0		
	SIFT Total	37	25	36	26	35	25	34	22	32	25	35	25		
	ORB Average	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	SIFT Average	6.16	4.16	6	4.33	5.83	4.16	5.66	3.66	5.33	4.16	5.83	4.16	5.8	4.311
Cubic Interpolation	ORB Total	0	0	0	0	0	0	0	0	0	0	0	0		
	SIFT Total	32	24	37	28	35	27	37	28	32	25	35	27		
	ORB Average	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	SIFT Average	5.33	4	6.16	4.66	5.83	4.5	6.16	4.66	5.33	4.16	5.83	4.5	5.77	4.41
Lanczos Interpolation	ORB Total	0	0	0	0	0	0	0	0	0	0	0	0		
	SIFT Total	35	24	35	25	38	27	37	29	35	27	36	28		
	ORB Average	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	SIFT Average	5.83	4	5.83	4.16	6.33	4.5	6.16	4.83	5.83	4.5	6	4.66	6	4.44

다 추출된 특징점 전체 수의 평균은 ORB 알고리즘의 경우 양선형 보간법에서 29개, 바이 큐빅 보간법에서 30개, Lanczos 보간법에서 34개이며, SIFT 알고리즘의 경우 양선형 보간법에서 16개, 바이 큐빅 보간법에서 20개, Lanczos 보간법에서 19개이다. 이 중 올바른 특징점의 평균은 ORB의 경우 각 10개, 11개, 12개이며, SIFT의 경우 세 가지 보간법 모두 8개의 특징점을 찾아내었다. ORB의 경우 전체 특징점에 비해 올바른 특징점의 수가 비교적 적은 반면 SIFT 알고리즘의 경우 전체 특징점의 절반 정도의 올바른 특징점을 찾아내었다. 위 결과를 통해 첫 번째 비교방식에서는 ORB가 SIFT보다 매칭 일관성이 떨어짐을 알 수 있으며, Fig. 7을 보면 ORB 보다 SIFT가 전반적으로 특징점 매칭이 잘 된 것을 확인할 수 있다.

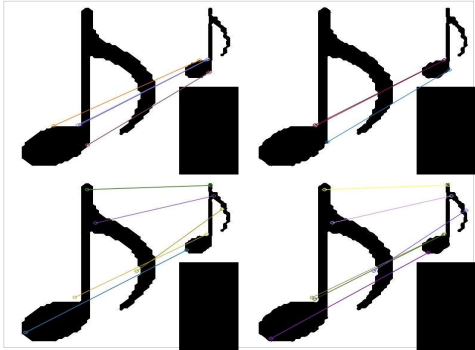


Fig. 7. Image of the matching result of the first comparison method

Fig. 7에서 위쪽은 ORB를 사용한 매칭이고, 아래쪽은 SIFT, 좌측은 한 번에 확대한 경우, 우측은 10번에 나누어 확대한 경우의 결과 이미지이다. 첫 번째 방식에서 가장 좋은 정확도를 보인 방식은 한 번에 확대한 양선형 보간법의 5배 SIFT로 전체 특징점 대비 올바른 특징점의 비율이 약 66%이다. Table 2에서는 Table 1과 마찬가지로 각 배율 및 사용한 보간법, 확대 방식에 따른 결과를 보여준다. Table 2의 결과를 보면 ORB는 특징점이 전혀 추출되지 않은 것으로 보아 작은 영상에서의 특징점 추출에 취약한 것을 알 수 있다. 그러나 SIFT는 모든 영상에서 특징점이 추출되었고, 이에 대한 올바른 특징점 또한 높은 정확도를 보여주고 있다. 이를 통해 SIFT는 ORB와 다르게 작은 영상에서도 특징점이 추출 가능하다는 것을 알 수 있다. 첫 번째 비교방식에서 전체 특징점 대비 올바른 특징점 비율이 약 66%인 것에 비해 두

번째 방식은 약 72% 정도의 매칭 정확도를 보여준다. 이를 통해 ORB보다 SIFT가 매칭 정확도가 더 뛰어난 것을 확인할 수 있다. Fig. 8에서 매칭된 결과를 통해 정확도가 높은 것을 확인할 수 있다.

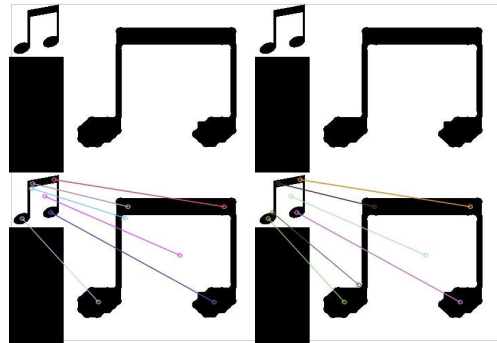


Fig. 8. Image of the matching result of the second comparison method

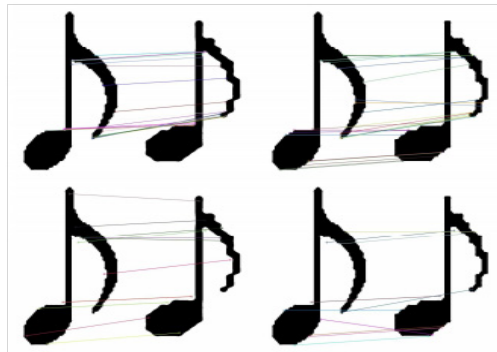


Fig. 9. Image of the matching result of the third comparison method

Table 3을 보면 세 번째 비교방식이 앞선 두 방식보다 더 많은 특징점이 추출되었다는 것을 확인할 수 있지만, 그에 대한 올바른 특징점의 비율은 약 35% 정도로 위 두 방식보다 좋지 못한 결과를 보이는 것을 알 수 있다. Fig. 9는 세 번째 비교방식을 적용한 Lanczos 보간법의 9배 확대 영상이다. Fig. 9를 보면 ORB 매칭에서 특징점이 몇몇 부위에 집중적으로 모여 있는 것을 보아 해당 비교방식은 분류에 적합하지 않다는 것을 알 수 있다. 세 번째 비교방식에서 가장 좋은 정확도는 한 번에 확대한 바이 큐빅의 5배 SIFT로 전체 특징점 대비 올바른 특징점 비율이 약 54%이다.

Table 3. Compare ORB, SIFT results for the third comparison method

		Five Times				Seven Times				Nine Times				Average	
		Once		Several		Once		Several		Once		Several		Average	
		Total	Good	Total	Good	Total	Good	Total	Good	Total	Good	Total	Good	Total	Good
Bilinear Interpolation	ORB Total	228	82	264	83	294	112	376	136	309	112	428	139		
	SIFT Total	81	37	87	33	109	49	111	45	142	44	143	54		
	ORB Average	38	13.66	44	13.83	49	18.66	62.66	22.66	51.5	18.66	71.33	23.16	52.75	18.44
	SIFT Average	13.5	6.16	14.5	5.5	18.16	8.16	18.5	7.5	23.66	7.33	23.83	9	18.69	7.28
Cubic Interpolation	ORB Total	227	78	267	99	307	122	373	126	325	107	492	136		
	SIFT Total	82	44	141	43	118	48	176	51	137	53	230	71		
	ORB Average	37.83	13	44.5	16.5	51.16	20.33	62.16	21	54.16	17.83	82	22.66	55.3	18.55
	SIFT Average	13.66	7.33	23.5	7.16	19.66	8	29.33	8.5	22.83	8.83	38.33	11.83	24.55	8.61
Lanczos Interpolation	ORB Total	232	84	220	83	306	127	350	125	339	124	435	138		
	SIFT Total	88	45	85	35	123	39	151	44	139	50	159	60		
	ORB Average	38.66	14	36.66	13.83	51	21.16	58.33	20.83	56.5	20.66	72.5	23	52.28	18.91
	SIFT Average	14.66	7.5	14.16	5.83	20.5	6.5	25.16	7.33	23.16	8.33	26.5	10	20.69	7.58

Table 4. Compare ORB, SIFT results for the 4th comparison method

		Five Times				Seven Times				Nine Times				Average	
		Once		Several		Once		Several		Once		Several		Average	
		Total	Good	Total	Good	Total	Good	Total	Good	Total	Good	Total	Good	Total	Good
Bilinear Interpolation	ORB Total	33	13	34	15	100	39	126	46	154	50	243	82		
	SIFT Total	53	35	65	28	75	40	72	38	74	41	105	58		
	ORB Average	5.5	2.16	5.66	2.5	16.66	6.5	21	7.66	25.66	8.33	40.5	13.66	19.16	6.8
	SIFT Average	8.83	5.83	10.83	4.66	12.5	6.66	12	6.33	12.33	6.83	17.5	9.66	12.33	6.66
Cubic Interpolation	ORB Total	29	14	42	21	120	50	138	50	186	74	263	92		
	SIFT Total	52	39	80	39	75	42	130	51	78	38	140	56		
	ORB Average	4.83	2.33	7	3.5	20	8.33	23	8.33	31	12.33	43.83	15.33	21.61	8.36
	SIFT Average	8.66	6.5	13.33	6.5	12.5	7	21.66	8.5	13	6.33	23.33	9.33	15.41	7.36
Lanczos Interpolation	ORB Total	38	16	33	17	116	48	125	44	216	83	300	94		
	SIFT Total	64	41	96	39	77	40	146	67	92	43	134	55		
	ORB Average	6.33	2.66	5.5	2.83	19.33	8	20.83	7.33	36	13.83	50	15.66	23	8.39
	SIFT Average	10.66	6.83	16	6.5	12.83	6.66	24.33	11.16	15.33	7.16	22.33	9.16	16.91	7.91

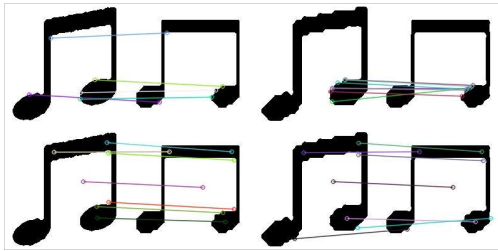


Fig. 10. Image of the matching result of the 4th comparison method

Table 4를 보면 위 다른 방식들과는 다르게 ORB의 특징점들의 전체적인 수가 유독 적은 것을 알 수 있다. 이는 템플릿을 크기가 상대적으로 작은 추출된 음표의 크기에 맞춘 후 배율을 적용하였기 때문에 위의 다른 템플릿보다 상대적 크기가 줄어들어 발생한 현상이다. 이를 통해 두 번째 방식처럼 네 번째 방식에서 ORB가 작은 영상에서의 특징점 추출에 취약함을 다시 확인할 수 있다. Fig. 10은 바이 큐빅 보간법의 5배 확대인 8분 음표의 매칭 결과들로 이 중 우측 아래의 영상은 한 번에 확대한 SIFT의 결과로 추출한 전체 특징점과 올바른 특징점이 모두 일치하는 결과를 보여주었다.

## 5. 결론

본 논문에서는 악보 인식의 세 과정 중 전처리 과정의 잡음 제거의 방법을 제안하며, 음표를 분류하기에 적절한 비교 알고리즘을 찾기 위해 음표 영상에 다양한 비교 방식, 확대 방법 및 확대 크기를 적용하여 비교한다. 본 논문에서 제안한 잡음 제거 방법은 형태학 연산과 윤곽선 탐지 알고리즘을 활용해 악절의 윤곽선과 각 음표 및 기호들의 윤곽선을 탐지한 후 악보 인식에 필요한 영역들을 하나로 묶고, 이외의 영역의 화소값을 모두 255(흰)값으로 변경하는 것으로 잡음을 제거하는 방법을 제안하고, 음표들의 정확한 분류를 위해 적절한 알고리즘을 찾을 필요성을 느껴 대표적인 비교 알고리즘인 ORB와 SIFT를 확대 방법, 확대 크기, 비교 방식을 바꿔가며 음표의 특징점을 비교한다. 이때 탐지된 특징점 중 올바른 특징점을 찾기 위해 두 영상의 Keypoint의 좌표의 비율의 차를 구하여 차가 15% 이내일 경우를 올바른 특징점이라 한다. 확대 방법은 한 번에 확대하는 경우와 10번에 나누어 확대하는 방법을 사용하고, 확대 크기는 원본 크기의 5배, 7배, 9배 크기를 적용한다. 비교 방식에는 템

플릿과 추출된 음표를 원본 상태로 비교하는 방식과 템플릿에는 변화를 주지 않고, 추출된 음표에만 배율을 적용하는 방식, 추출된 음표를 템플릿 크기에 맞춘 후 두 영상을 모두 확대하여 비교하는 방식, 템플릿을 추출된 음표의 크기에 맞춘 후 확대하는 방식의 네 가지 방식을 사용하여 비교한다. 그 결과 첫 번째 방식에서는 ORB가 SIFT에 비해 매칭 일관성이 떨어짐을 알 수 있고, 두 번째와 네 번째 방법에서는 ORB가 크기가 작은 영상의 특징점 추출에 취약함을 알 수 있다. 또, 세 번째 방식에서는 ORB를 사용할 때 특징점이 한쪽에 집중되는 현상이 있는 것으로 보아 해당 비교 방식은 음표를 분류하는데 적절하지 못한 방식임을 알 수 있다. 위 네 가지 방식의 결과로 볼 때 음표를 분류하는 것에 있어 SIFT가 ORB보다 더 적절하다는 것을 알 수 있다. 향후 연구에서는 제안한 전처리 과정과 SIFT 알고리즘을 이용하여 악보의 음표를 추출 및 분류를 토대로 악보의 음계를 인식하는 인공지능 모델의 연구를 진행할 계획이다.

## References

- [1] K. W. Kim, "Medical image processing using artificial intelligence", *KSIDI AI Outlook*, Vol.10, pp.1-15, 2022.
- [2] S. H. Lee, M. J. Lee, "A Study of Establishment and application Algorithm of Artificial Intelligence Training Data on Land use/cover Using Aerial Photograph and Satellite Images", *Korean journal of remote sensing*, Vol.37, No.5, pp.871-884, 2021.  
DOI: <https://doi.org/10.7780/kjrs.2021.37.5.1.4>
- [3] S. Oh, Y. J. Kim, K. G. Kim, "Evaluation of the Effect of using Fractal Feature on Machine learning based Pancreatic Tumor Classification", *Journal of Korea Multimedia Society*, Vol. 24, No. 12, pp. 1614-1623, Dec 2021.  
DOI: <https://doi.org/10.9717/kmms.2021.24.12.1614>
- [4] A. Emmamuel, U. Asim, H. Yu, S. Kim, "3D-CNN Method over Shifted Patch Tokenization for MRI-Based Diagnosis of Alzheimer's Disease Using Segmented Hippocampus", *Journal of Multimedia Information System*, Vol. 9, No. 4, pp. 245-252, Dec 2022.  
DOI: <https://doi.org/10.33851/JMIS.2022.9.4.245>
- [5] K. B. Kim, "Musical Score Recognition with SOM and Enhanced ART-1", *Journal of the Korea Institute of Information and Communication Engineering*, Vol.17, No.5, pp.1064-1069, 2013.  
DOI: <http://dx.doi.org/10.6109/kiice.2013.17.5.1064>
- [6] M. T. Trieu, G. S. Lee, "Staff-line Removal for Music Score Images using U-net", *KIISE Transactions on*



Computing Practices, Vol. 26, No. 1, pp.26-31, 2020, 1.  
DOI: <https://doi.org/10.5626/KTCP.2020.26.1.26>

- [7] K. B. Kim, W. J. Lee, Y. W. Woo, "Automatic Recognition and Performance of Printed Musical Sheets Using Fuzzy ART", Korea institute of Electronic Communication Science, Vol. 6, No. 1, pp.84-89, 2011.  
DOI: <https://doi.org/10.13067/JKIECS.2011.6.1.084>
- [8] J. M. Yoo, G. H. Kim, G. S. Lee, "Music Recognition by Partial Template Matching, Journal of the Korea contents association", Vol. 8, No. 11, pp.85-93, 2008.  
DOI: <https://doi.org/10.5392/JKCA.2008.8.11.085>
- [9] D. G. LOWE, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, 60(2), 91 - 110, 2004.  
DOI: <http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94>
- [10] E. Rublee, V. Rabaud, K. Konolige, G. Bradski, "ORB: an efficient alternative to SIFT or SURF", IEEE International Conference on Computer Vision, 2564 - 2571, 2011.  
DOI: <http://dx.doi.org/10.1109/ICCV.2011.6126544>
- [11] S. W. Je, D. W. Lee, "Morphology based Elimination of Staff Lines for Recognition of Musical Notes", Journal of the Korea Academia-Industrial cooperation Society, Vol. 23, No. 5, pp.334-340, 2022.  
DOI: <https://dx.doi.org/10.5762/KAIS.2022.23.5.334>
- [12] P. R. Smith, "Bilinear interpolation of digital images", Ultramicroscopy, Vol. 6, No. 1, pp. 201-204, 1981.  
DOI: [https://doi.org/10.1016/S0304-3991\(81\)80199-4](https://doi.org/10.1016/S0304-3991(81)80199-4)
- [13] R. Keys, "Cubic convolution interpolation for digital image processing", IEEE, Vol. 29, No. 6, pp. 1153-1160, Dec 1981.  
DOI: <https://doi.org/10.1109/TASSP.1981.1163711>
- [14] B. N. Madhukar, R. Narendra, "Lanczos Resampling for the Digital Processing of Remotely Sensed Images", LNEE, Vol. 258, pp. 403-411, 2013.  
DOI: [http://doi.org/10.1007/978-81-322-1524-0\\_48](http://doi.org/10.1007/978-81-322-1524-0_48)

이 덕 우(Deokwoo Lee)

[정회원]



- 2007년 2월 : 경북대학교 전자전 기컴퓨터학부 졸업 (공학사)
- 2012년 12월 : North Carolina State University, USA, 전기컴 퓨터공학부 석사, 박사
- 2013년 1월 ~ 4월 : Washington Univ. in St. Louis, USA, 박사후 연구원
- 2013년 9월 ~ 2016년 8월 : 삼성전자 무선사업부 책임 연구원
- 2016년 9월 ~ 2018년 2월 : 영산대학교 가상현실콘텐츠 전공 조교수
- 2018년 3월 ~ 현재 : 계명대학교 공과대학 컴퓨터공학전 공 조교수

<관심분야>

신호처리, 영상처리, 컴퓨터비전

제 승 완(Seungwan Je)

[준회원]



- 2017년 3월 ~ 현재 : 계명대학교 컴퓨터공학부 컴퓨터공학전공 학사과정

<관심분야>

영상처리, 영상인식, 인공지능