

A Study on a Ginseng Grade Decision Making Algorithm Using a Pattern Recognition Method

Seokhoon Jeong[†] · Kuk Won Ko^{**} · Je-Yong Kang^{***} · Suwon Jang^{****} · Sangjoon Lee^{*****}

ABSTRACT

This study is a leading research project to develop an automatic grade decision making algorithm of a 6-years-old fresh ginseng. For this work, we developed a Ginseng image acquiring instrument which can take 4-direction's images of a Ginseng at the same time and obtained 245 jingen images using the instrument. The 12 parameters were extracted for each image by a manual way. Lastly, 4 parameters were selected depending on a Ginseng grade classification criteria of KGC Ginseng research institute and a survey result which a distribution of averaging 12 parameters. A pattern recognition classifier was used as a support vector machine, designed to "k-class classifier" using the OpenCV library which is a open-source platform. We had been surveyed the algorithm performance(Correct Matching Ratio, False Acceptance Ratio, False Reject Ratio) when the training data number was controlled 10 to 20. The result of the correct matching ratio is 94% of the 1st ginseng grade, 98% of the 2nd ginseng grade, 90% of the 3rd ginseng grade, overall, showed high recognition performance with all grades when the number of training data are 10.

Keywords : Pattern Recognition, Ginseng Grade Decision Making, Pattern Classifier

패턴인식을 이용한 수삼 등급판정 알고리즘에 관한 연구

정석훈[†] · 고국원^{**} · 강제용^{***} · 장수원^{****} · 이상준^{*****}

요 약

본 연구는 비 정형 농산물 중 6년근 수삼의 자동 등급 분류하기 위한 선행연구로, 이를 위해 4방향에서 이미지 취득이 가능한 수삼 영상 측 정기를 제작 하였으며 총 245 수삼 개체에 대해서 영상을 취득하였다. 취득된 영상의 각 수삼 개체마다 12개의 파라미터를 추출하였으며, KGC 인삼공사의 수삼등급 분류 기준과 각 등급별 평균 파라미터의 분포를 조사하여 최종 4개 파라미터를 선정하였다. 패턴인식 분류기는 Support Vector Machine을 사용하였으며 공용 소프트웨어인 OpenCV Library를 사용하여 k-Class 분류기를 설계하였다. 각 등급별 학습 데이터 수를 10, 15, 20으로 조정하여 등급별 인식률, 본인 거부율, 타인 인식율을 조사하였으며, 학습데이터 수가 10개일 때 1등급 인식률 94%, 2등급 인식 률 98%, 3등급 인식률 90%로 가장 높은 인식 성능을 보였다.

키워드 : 패턴인식, 수삼등급판정, 패턴 분류기

1. 서 론

대한민국은 인삼의 종주국으로 담배인삼공사에서 제조· 판 매하는 홍삼 및 홍삼 제품은 그 품질 및 신용도에 있어서 세 계적으로 인정받고 있다. 홍삼제조용 6년근 수삼 구매 시 등

급판정은 사람의 눈과 그간의 경험에 의존하고 있는 실정으로 보다 객관적인 품질분류가 필요하다[1]. 이러한 등급판정에 있어서, 매년 경작인들과의 논란이 있어 왔으며, 원료 삼 구매 시 등급판정에 필요한 연간 소요비용은 약 수억 원 가량으로 연인원 1,530명의 숙련된 검사원이 필요한 실정으로 보고되고 있다. 홍삼제조용 6년근 수삼의 전체적인 자동 분류는 아니더라도 일정 부분 자동 분류가 된다면 상당부분 검사에 필요한 비용을 절감 효과를 기대할 수 있다. 본 연구의 목적은 6년근 수삼의 등급판정을 위한 선행연구로 수삼 영상취득을 위한 시스템 개발과, 수삼의 특징추출, 패턴 인식알고리즘을 공개 소프트웨어인 OpenCV 영상처리 라이브러리를 사용하여 구현하였으며, 학습과 인식을 과정을 통해 KGC 인삼공사의 분류기 준과 비교하여 등급 인식성능 평가를 수행하였다.

※ 이 논문(저서)은 2014년 교육부와 한국연구재단의 지역혁신창의인력양성 사업(NRF-2014H1C1A10G6998), KGC 인삼연구원의 지원을 받아 수행된 연구임.

† 비 회 원 : 선문대학교 정보통신공학과 박사과정
** 비 회 원 : 선문대학교 기계ICT융합공학부 부교수
*** 비 회 원 : KGC인삼공사(1급) 팀장
**** 비 회 원 : KGC인삼공사 선임연구원
***** 정 회 원 : 선문대학교 기계ICT융합공학부 조교수

Manuscript Received : April 29, 2016
Accepted : May 17, 2016

* Corresponding Author : Sangjoon Lee(mcp94lee@sunmoon.ac.kr)

2. 본 론

2.1 수삼영상 취득장치

6년근 수삼은 보는 각도에 따라서 형상 및 등급이 다르게 판단될 수 있다. 따라서 수삼 영상 취득 장치 설계 시 가장 고려해야 할 점은 객체의 모든 부분을 검출 가능해야 하며, 취득된 영상의 균일한 품질과 균일한 영상인식결과를 얻기 위해 일정한 조명이 필요하다. 개발된 영상측정 장치는 인삼의 360° 모든 방향에서 객체의 영역이 검출 가능하도록 네 방향 카메라를 탑재하였고, 일정한 조명이 객체를 비칠 수 있도록 두 개의 5W급 LED조명을 탑재했다. 제작된 측정 장치로 취득 가능한 객체의 크기는 (500×500×500)mm의 충분한 크기로 제작되어 6년근 수삼의 세근까지도 측정 가능하도록 설계하였다.

Fig. 1은 6년근 수삼 영상 취득 시스템 개발을 위한 3D도면을 나타내며, Fig. 2는 6년근 수삼 영상취득 프로그램을 나타낸다. 취득 프로그램은 National Instrument 사의 Labwindows (2013 Version)과 OpenCV(2.4.8) 영상처리 라이브러리를 포함하였다.

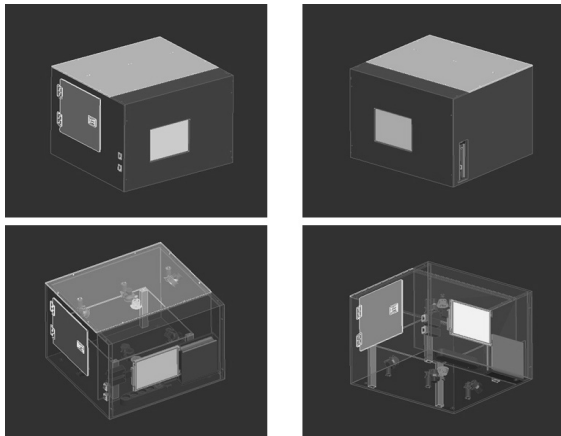


Fig. 1. 6-years Ginseng image acquiring device (3D-sketch)

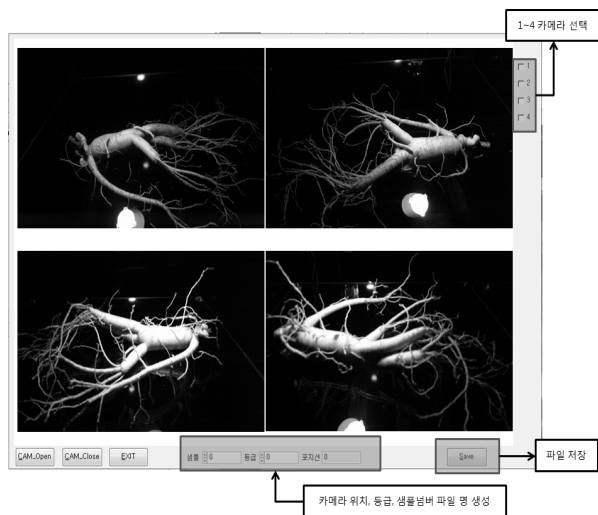


Fig. 2. Image acquiring Software

2.2 수삼의 형상학적 특징

인간의 모양과 외형이 모두 다른 것과 같이 수삼 또한 각각 개체 마다 독특한 외형을 가지고 있다. 하지만 특정 위치마다 부르는 명칭이 틀리다. Fig. 3은 수삼의 대표적인 5 부위를 구분한 그림으로 Fig. 3(a) 부위를 “수근”이라 부르며, 몸통 또는 머리부위에 붙어있는 형태를 취하여 모든 수삼마다 존재하지는 않는다. Fig. 3(b)는 “뇌두”라 부르며 수삼의 머리 부분에 해당되며 모든 수삼마다 존재한다. Fig. 3(c)는 수삼에 몸통에 해당되는 부위로 “동체”라 부르며 수삼의 가장 많은 면적을 차지하며 홍삼제조시 가장 중요한 요소이다. Fig. 3(d)는 수삼의 다리에 해당되는 부위로 “지근”이라 부르며 수삼마다 1~5개 정도 분포한다. Fig. 3(e)는 땅 속의 영양분을 흡수하는 부위로 “세근”이라 부른다.

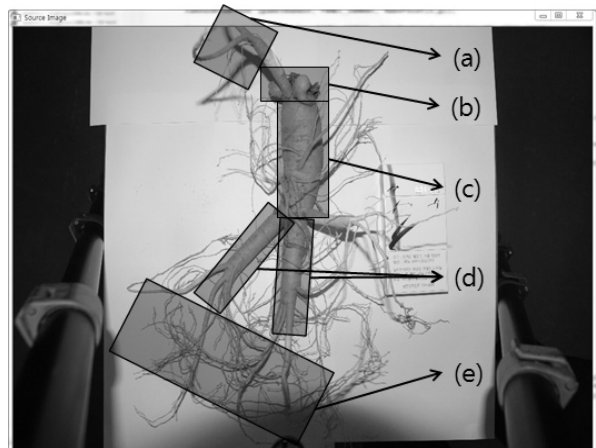


Fig. 3. Parts of the Ginseng category

2.3 수삼의 파라미터 추출

수삼의 파라미터는 수동으로 측정하였으며 Fig. 3과 같이 6년근 수삼 이미지를 로딩하고 뇌두, 동체, 지근에 해당하는 영역을 임의로 구분 하였다. Fig. 1의 측정기에서 수삼의 이미지 취득 시 렌즈의 Focusing 거리 등 영상의 화소수당 길이 연산이 가능 하도록 거리 기준이 되는 물체를 두어 Calibration을 수행하였다. Fig. 4와 같이 Calibration을 위해 기준 Object의 최종단 두 점을 Click하면 Pixel 당 길이다. 아래와 같은 Equation (1)과 Equation (2)와 같이 연산된다.

$$D = \sqrt{(X2 - X1)^2 + (Y2 - Y1)^2} \tag{1}$$

$$Length = \frac{O_L \times P_L}{D} \tag{2}$$

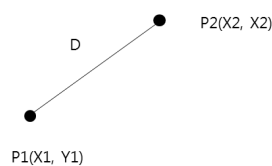


Fig. 4. Distance calculation method between the two pixels

Equation (1)에서 D 는 기준 Object의 화소의 개수를 의미하며, Calibration 대상 Object 길이에서 측정된 D 를 나눈 면 화소수당 거리(mm)가 측정가능하다. Equation (2)에 의해서 측정 하고자 하는 두 점의 화소 수(D)를 입력하면 길이를 연산이 가능하며 O_L Calibration 기준 Object의 길이 (mm), P_L 은 측정하고자 하는 대상물체의 두 점간의 화소 수 D 는 Calibration 기준 Object 화소 수를 의미한다.

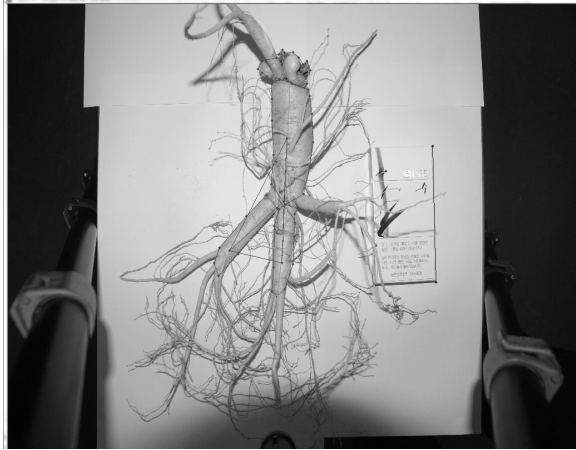


Fig. 5. Image loading and segmentation of the parts

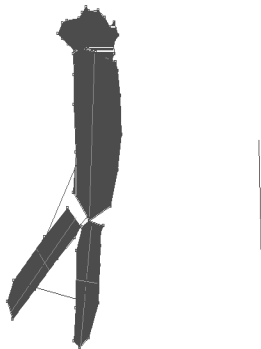


Fig. 6. Segmentation result image

로딩한 수삼 이미지 파일로부터 뇌두, 동체, 지근의 영역의 범위에 적당하도록 위치를 마우스로부터 입력 받아, 각 Segmentation 영역의 넓이를(폐구간의 화소 수) 구한다. Fig. 5는 수삼에 대한 Segmentation 영역 할당하는 그림을 보여주며 Fig. 6은 마우스 클릭 좌표를 기준으로 각 영역의 Polygon을 나타낸 그림이다. 이러한 정보로부터 하나의 수삼 당 12개의 직접적인 파라미터를 측정한다. 측정된 파라미터는 1) 뇌두의 넓이, 2) 동체의 넓이, 3) 대표지근의 1의 넓이, 4) 대표 지근 2의 넓이, 5) 동체의 세로길이, 6) 대표 지근 1의 세로길이, 7) 대표 지근 2의 세로길이, 8) 대표 지근 1의 가로길이, 9) 대표 지근 2의 가로길이, 10) 지근간 각도, 11) 동체 대비 지근 1의 각도 마지막으로 12) 동체 대비 지근 2의 각도를 추출하였으며 Fig. 7은 추출한 파라미터 결과를 보여준다.



Fig. 7. The result of 12 parameters extraction

또한, 추출된 12개 파라미터를 통해 동체의 가로길이, 동체비율 = $\left(\frac{\text{동체세로길이}}{\text{동체가로길이}}\right)$, 지근길이대비 동체 비율을 연산한다.

2.4 OpenCV의 Support Vector Machine Function과 “k-class” 분류기

SVM(support vector machine)은 패턴인식의 학습시스템의 새로운 장르로서 Cortes와 Vapnik에 의해 고안되었다[2]. SVM은 초기에 주목 받지 못하다가 최근 생물정보학, 문자 인식, 필기인식, 얼굴 및 물체인식 등에 다양한 분야에서 사용되고 있으며 뛰어난 성능 때문에 supervised 패턴분류에

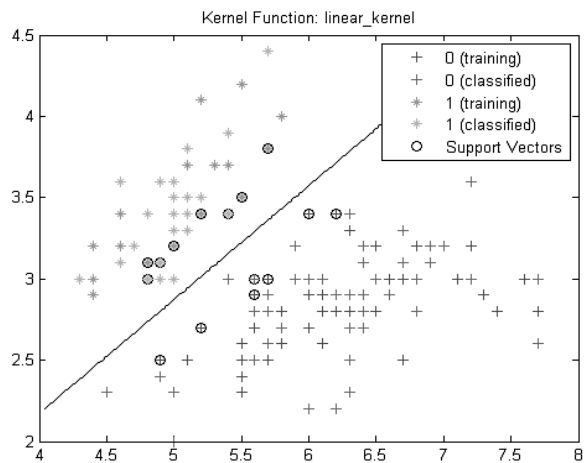


Fig. 8. The basic concept of the SVM and a hyperplane

서 널리 사용된다. SVM 분류기의 원리는 Fig. 8에서 보는 것과 같이 결정경계(decision boundary)와 훈련데이터 사이에 최대 margin을 가지는 초 평면(hyperplane)을 설계하는 것이 목적으로, 패턴의 집합들 중에 결정경계에 가까운 것을 support vector라 부른다.

예를 들어 훈련데이터 집합 $(x_1, y_1), \dots, (x_i, y_i)$, 입력변수 $x_i \in R^N$ 이고 class label $y_i \in \{-1, +1\}$ 일 때, 선형분리가 가능한 문제에서 결정함수의 초 평면은 $w^T x + b = 0$ 과 같다 (w는 초평면 계수의 벡터, b는 bias term을 의미한다.) 2-Class만 분류하는 SVM의 경우에는 $y_i \in \{-1, +1\}$ 로 정의되었지만 k Class 개를 가지는 One Against One SVM의 경우에는 $y_i \in \{-1, \dots, k\}$ 로 정의할 수 있다. 간단한 수학적 조작을 통해 2-Class SVM 분류기 에서 k-Class SVM 분류기로 확장할 수 있다[3]. 이러한 방법은 2개의 Class로부터 하나씩 트레이닝 데이터를 구성하여 개의 분류기를 만든다. One Against One의 SVM 분류방법에서는 voting 전략이 사용되며[4] 결정함수에 의해 많이 투표된 class가 x의 Class라고 예상할 수 있다. 또한 Voting 접근방법은 “Max Wins” 전략으로 불리기도 한다. 본 연구에서 사용한 OpenCV Library는 2-Class 분류기 함수만 존재하므로 위와 같은 방법으로 k-class 분류기를 설계하였다.

3. 실험 방법

3.1 학습을 위한 파라미터 선정

KGC 인삼연구원에서 제공한 수삼 샘플 총 245개체를 사용하여 영상을 취득하였고 각 등급별 개수는 1등급 69개 샘플, 2등급 76개의 샘플, 3등급 100개로 구성된다. 수삼 샘플의 영상 취득 전 기존 전문가에 의한 판정에 선별되었으며 측정 도구를 이용하지는 않았다. 패턴인식 학습 및 테스트를 위해 수삼의 파라미터를 선정해야 하며, 2절에서 언급한 것과 같이 수삼 분석 프로그램에서 12개의 파라미터를 모두 사용이 가능하지만 각 등급의 파라미터 편차가 크기 않으면 인식 오류로 나타나므로 최적의 인식성능을 위해 파라미터들의 평균값 분석을 수행하였다.

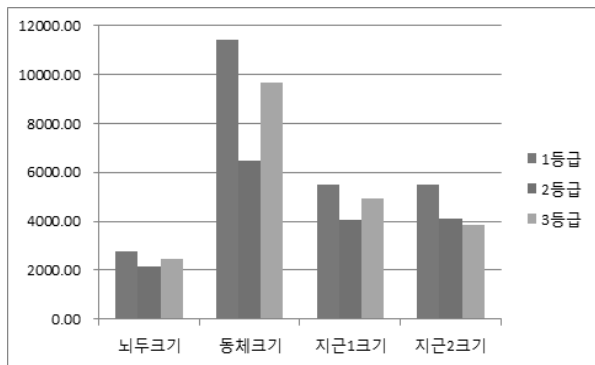


Fig. 9. The average volume value of head, body, leg 1 and leg 2 for each grade (Unit: pixels number)

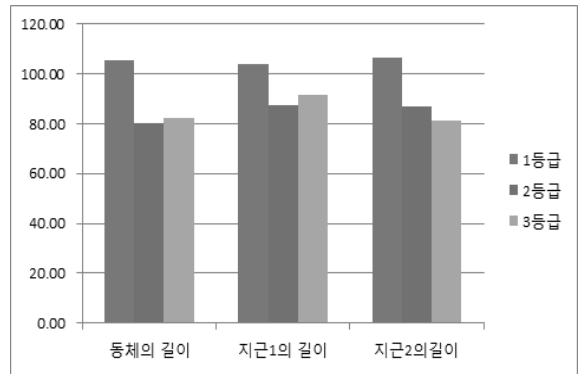


Fig. 10. The average length of body, leg 1 and leg 2 for each grade (Unit: mm)

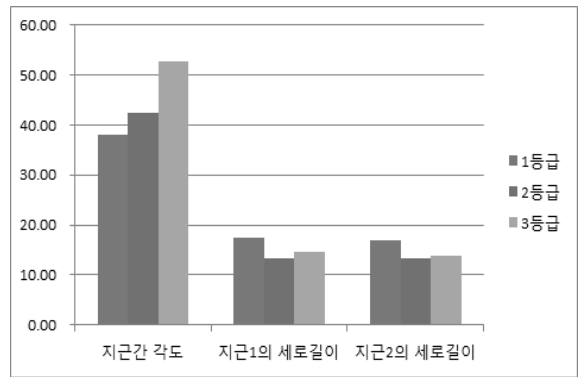


Fig. 11. The average value of the angle between two legs, vertical length of leg 1 and leg 2 for each grade (Unit: leg angle = angle, length = mm)

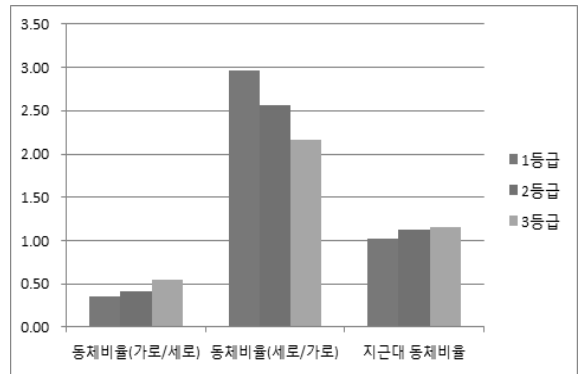


Fig. 12. The average ratio value for each grade (Unit: %)

Fig. 9-12는 수삼 샘플 245개체 파라미터들의 각 등급별 평균값의 분포를 나타낸 그림으로, 등급별 파라미터들의 평균의 차이가 클수록 등급 분류에 큰 영향을 미치며 등급별 차이가 미미하면 패턴인식 결과의 오류로 나타날 확률이 높으므로 패턴인식 학습 시 제외하는 것이 인식률 향상에 도움이 된다. Fig. 9-12에 의해 각등급 별 파라미터 편차가 큰 동체의 크기, 지근 간 각도, 동체의 비율을 패턴인식 학습의 파라미터로 선정하는 것이 옳을 것으로 판단된다. Table 1 은 KGC인삼 공사에서 제공한 1, 2, 3 등급 분류 기준표로

수삼의 무게, 동체의 길이, 비율, 지근의 개수 등의 판정 파라미터로 활용되며, 분류 현장에서 관찰해보면 경력이 오래된 분류 전문가가 정확한 측정 도구 없이 대략적인 판단으로 분류하는 것으로 조사되었다. 이에 의해 학습 및 테스트를 위한 파라미터로 (1) 동체 세로길이, (2) 동체 가로길이, (3) 동체 세로/가로비율, (4) 동체 가로/세로비율 4종류의 수삼 파라미터를 패턴인식 학습과 테스트 데이터로 선정하였다.

Table 1. 6-years Ginseng grade classification criteria by a KGC Ginseng institute

	1등	2등	3등	등외
머리	몸통과 균형을 이룬 머리가 있는 것	몸통과 균형을 이룬 머리가 있는 것	몸통과 균형을 이룬 머리가 있는 것	제한없음
	몸통이 부분적으로 비대하거나 굴곡이 심하지 아니한 것	몸통이 부분적으로 비대하거나 굴곡이 심하지 아니한 것	몸통이 부분적으로 비대하거나 굴곡이 심하지 아니한 것	제한없음
몸통 및 다리	몸통의 길이가 6cm 이상이고 5cm 이상의 다리가 2개 이상인 것 다만 몸통지름이 몸통길이의 2/5 이하이어야 한다.	몸통의 길이가 4cm 이상이고 3cm 이상의 다리가 2개 이상인 것이나 몸통의 길이가 8cm 이상이고 직립형인 것. 다만 몸통지름이 몸통길이의 1/2 이하이어야 한다.	몸통의 길이가 3cm 이상이고 몸통과 균형을 이룬 다리가 있는 것이나 몸통의 길이가 5cm 이상이고 직립형인 것. 다만 무게가 50g 이상인 체형 불량삼을 포함한다.	제한없음

패턴인식에 의한 등급분류 방법은 Fig. 13과 같으며 또한 패턴인식을 위한 학습데이터와 테스트를 위한 데이터들은 중복되지 않게 설계하였다. 성능평가를 위한 지표로는 각 등급의 Correct Match Rate(%), 평균 Correct Match Rate, FRR(False Rejection Rate: 본인 거부율), FAR(False Acceptance Rate: 타인 수락율) 지표를 사용하였다. Table 2는 각 4개 파라미터를 이용하였을 때 학습 데이터 수 별 인식률 및 성능을 나타낸다. 인식 결과, Table 2에서 나타난 것과 같이 4개 파라미터 선정, 10개의 파라미터 데이터를 학습 하였을 때, 1등급 인식률 94%, 2등급 인식률 98%, 3등급 인식률 90%를 보였으며, FRR 6.0%, FAR 5.4%로 높은 인식 성능이 나타남을 알 수 있다.

Table 2. The results of Ginseng grade recognition

학습 데이터 수	1등급 인식률 (%)	2등급 인식률 (%)	3등급 인식률 (%)	평균 인식률 (%)	FRR (%)	FAR (%)
10개	94	98	90	94	6	5.4
15개	93.33	97.77	88.9	93.3	6.6	5.9
20개	96.66	100	80	92.2	7.7	6.3

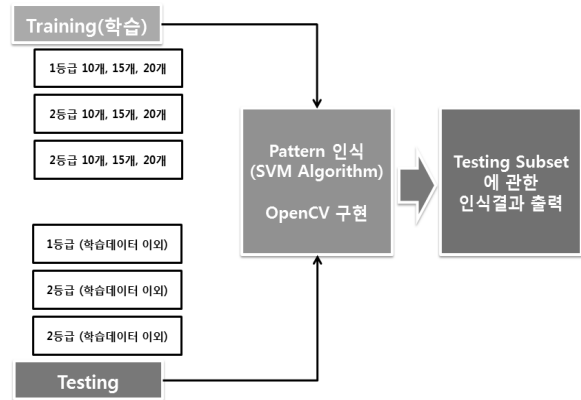


Fig. 13. A Ginseng grade classification method by pattern recognition

4. 결론 및 고찰

수삼 영상 측정기 제작과, 취득된 수삼의 영상에서 파라미터 추출을 하여 패턴인식 알고리즘을 통한 수삼 등급분류 알고리즘 및 프로그램을 공용 소프트웨어를 사용하여 제작하였다. 각 등급별 인식을 분석을 위해서 수삼의 파라미터를 수동으로 추출하여 저장하였고 1, 2, 3 등급 총 245 개체에 대해서 각 12종의 파라미터를 추출하였다. 이 중 인식을 위한 파라미터 선정은 KGC 인삼공사의 수삼 선별기준과 각 등급별 파라미터 평균 데이터의 분포가 큰 것을 기준으로 최종 4종의 파라미터를 선정하였다. 학습 데이터는 10개, 15개, 20개로 조정하여 인식률을 조사하였으며, 학습 데이터수가 10개일 때 평균 94%의 높은 인식성능을 보였다. 다양한 환경과, 분류 장비에 적용하기 위한 추가 실험과 최적의 알고리즘을 설계한다면 좀 더 높은 인식성능을 보장할 수 있을 것이라 판단되며, 자동 수삼 등급 분류기 적용에 있어서 자동으로 파라미터를 추출하는 알고리즘을 설계가 필요하다. 본 연구가 실제 수삼 분류현장에 적용된다면 빠른 분류 속도와 비용 절감 효과를 동시에 기대해도 좋을 것이라 판단된다.

References

[1] J. Y. Kang, M. G. Lee, and Y. T. Kim, "Automatic Decision-Making on the Grade of 6-Year-Old Fresh Ginseng (Panax ginseng C. A. Meyer) by and Image Analyzer - I. Shape and Weight Analyses According to the Grade of Fresh Ginseng)," *Journal of Ginseng Research*, Vol.20, No.1, pp.65-71, 1996.

[2] C. Cortes and V. Vapnik, "Support-vector network," *Machine Learning*, Vol.20, pp.273-297, 1995.

[3] C. J. C. Burges, "A tutorial on Support Vector Machines for pattern recognition," *Data Mining and Knowledge Discovery*, Vol.2, No.2, pp.121-167, 1998.



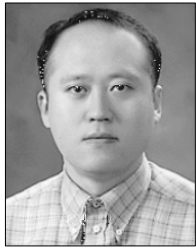
정 석 훈

e-mail : stellong2@gmail.com
2002년 명지대학교 전자공학과(학사)
2004년 명지대학교 전자공학과(석사)
2015년~현 재 선문대학교 정보통신공학과
박사과정
관심분야: 의료신호처리, 패턴인식,
영상처리



장 수 원

e-mail : swjang@kgc.co.kr
2001년 경북대학교 농학과(학사)
2003년 경북대학교 농학과(석사)
2007년 경북대학교 농학과(박사)
2005년 KT&G 연구원 입사
2012년~현 재 KGC인삼공사 선임연구원
관심분야: 인삼(생약재포함)재배/생리, 광합성, 식물호르몬



고 국 원

e-mail : kuks2309@sunmoon.ac.kr
1992년 서울대학교 농업기계학과(학사)
1994년 KAIST 정밀공학과(석사)
2001년 KAIST 기계공학(박사)
2003년~현 재 선문대학교 기계ICT융합
공학부 부교수

관심분야: 3차원 측정시스템, 영상처리, 인공지능



이 상 준

e-mail : mcp94lee@sunmoon.ac.kr
2001년 명지대학교 제어계측공학과(학사)
2005년 명지대학교 전자공학과(석사)
2011년 연세대학교 전기전자공학부(박사)
2014년~현 재 선문대학교 기계ICT융합
공학부 조교수

관심분야: 임베디드시스템, 의료신호처리, 패턴인식, 영상처리



강 제 용

e-mail : jykang@kgc.co.kr
1986년 중앙대학교 생물학과(학사)
1988년 중앙대학교 생물학과(석사)
2005년 경희대학교 한약가공과(박사)
1991년 KT&G 연구원 입사
2010년~현 재 KGC인삼공사(1급) 팀장

관심분야: 인삼재배, 환경제어, 수삼선별