



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학전문석사학위연구보고서

**스마트폰 CMOS 이미지 센서  
기술 로드맵 수립 연구: 미국 시장의  
온라인 고객 평가 자료를 기반으로**

**A Study on Technology Roadmap for  
Smartphone CMOS Image Sensor: Based on US  
market on-line consumer review data**

2021년 2월

서울대학교 공학전문대학원  
응용공학과 응용공학전공  
최성철

스마트폰 CMOS 이미지 센서 기술  
로드맵 수립 연구: 미국 시장의 온  
라인 고객 평가 자료를 기반으로

A Study on Technology Roadmap for  
Smartphone CMOS Image Sensor: Based on  
US market on-line consumer review data

지도 교수 구윤모

이 프로젝트 리포트를 공학전문석사 학위  
연구보고서로 제출함  
2020년 11월

서울대학교 공학전문대학원  
응용공학과 응용공학전공  
최성철

최성철의 공학전문석사 학위 연구보고서를 인준함  
2021년 1월

위원장 곽우영 

위원 구윤모 

위원 박창우 

# 초 록

디지털 카메라, 스마트폰 카메라, 자율 주행 자동차 또는 산업용 로봇의 눈이 되어 주는 CMOS 이미지 센서(CIS) 시장이 4차 산업혁명 시대를 맞이하여 급속히 커지고 있다. CMOS image sensor를 제조하는 우리나라 반도체 기업들도 현재까지의 패스트 팔로어(fast follower)전략을 벗어나 퍼스트 무버(first mover)로 업계 선도자 위치로의 전환을 꾀하는 중이다. 뿐만 아니라 글로벌 시장 환경이 변화되고, 시장 불확실성이 커지면서 정확한 마켓 센싱(Market sensing)이나 기술 수요 예측의 필요성이 점점 높아지고 있다. 이에 본 연구는 최종 소비자 빅데이터 분석을 이용한 품질 기능전개(QFD, Quality Function Deployment) 모형을 통해, 고객 지향적인 기술 요소 우선순위를 도출하고, 이를 활용하여 소비자 니즈(Customer needs)가 충분히 반영된 기술 로드맵을 수립하는 것이 목적이다.

품질기능전개(QFD) 모형은 신제품 개발 시 고객 요구 사항(CA, Customer Attribute)을 제품 설계 특성(EC, Engineering Characteristic)으로 변환하여 각 특성기술의 우선순위를 파악할 수 있게 한다. 본 연구는 미국 아마존(Amazon.com)의 스마트폰 카테고리의 소비자 리뷰 데이터에 대해, 자연어 처리(NLP, Natural Language Processing) 분야의 기계학습 알고리즘 모델인 word2vec 모델을 이용한 분석을 적용하여, 스마트폰 카메라에 대한 소비자 니즈를 정량화하였고, 이를 품질기능전개(QFD)의 핵심 도구인 품질의 집(HOQ, House Of Quality)의 CA 상대적 중요도에 적용하였다. 이후 미국 특허를 활용한 텍스트 마이닝(Text mining) 통계 분석을 통해 CA와 EC 간의 관계 매트릭스를 작성하였다. HOQ level 2, EC - EC 단계에서는 전문가 설문을 통해 전문가들의 의견을 반영하여 작성하였고,

최종적으로 CMOS 이미지 센서의 최신 기술과 연계시키고자 하였다.

결론적으로, 본 연구는 CMOS 이미지 센서의 최신 기술 요소와 미국 시장 소비자 니즈를 연계하여, 기술 요소 우선순위를 밝히고, 이를 통해 CMOS 이미지 센서 기술 로드맵 수립에 참고 자료로 활용하고자 한다.

**주요어 :** CMOS 이미지 센서(CIS), 기계학습, Word2Vec, 품질기능전개

**학번 :** 2019-26904

# 목 차

<b>I.</b>	<b>서론</b>	<b>1</b>
1.1	연구배경 및 목적	1
1.2	연구 보고서의 구성	3
<b>II.</b>	<b>관련 연구</b>	<b>4</b>
2.1	이론적 배경	4
2.1.1	텍스트 마이닝	4
2.1.2	자연어 처리와 워드임베딩	5
2.1.3	코사인유사도	8
2.1.4	품질기능전개(QFD)	9
2.2	선행 연구	11
2.2.1	온라인 리뷰에 빅데이터 분석을 적용한 연구	11
2.2.2	QFD기법 적용 연구 사례	12
<b>III.</b>	<b>연구 방법</b>	<b>14</b>
3.1	실험 데이터	14
3.1.1	Amazon社의 고객 리뷰 데이터	14
3.1.2	DXOmark社의 스마트폰 카메라 화질 평가 항목	15
3.1.3	미국 특허 데이터	15
3.1.4	삼성반도체 이미지 센서 최신제품 적용 기술	16
3.2	연구 모델	16
3.2.1	머신러닝을 통한 소비자 니즈 파악	16

3.2.2	QFD를 통한 기술속성 우선순위 도출 . . . . .	19
<b>IV.</b>	<b>연구결과 . . . . .</b>	<b>22</b>
4.1	1단계 실험결과 . . . . .	22
4.1.1	데이터 전처리 . . . . .	22
4.1.2	데이터 활용 - CA 상대적 중요도 도출 . . . . .	23
4.2	2단계 실험결과 . . . . .	26
4.2.1	HOQ level 1 . . . . .	26
4.2.2	HOQ level 2 . . . . .	29
<b>V.</b>	<b>결론 . . . . .</b>	<b>35</b>
5.1	연구 요약 및 시사점 . . . . .	35
5.2	연구 성과 및 기대 효과 . . . . .	37
	<b>참고 문헌 . . . . .</b>	<b>40</b>
	<b>Abstract . . . . .</b>	<b>43</b>

# 그림 목차

그림 1.	CBOW와 Skip-gram의 학습과정 . . . . .	7
그림 2.	House Of Quality . . . . .	10
그림 3.	연구 절차 . . . . .	17
그림 4.	전체 연구 모형 . . . . .	21
그림 5.	전처리 과정 . . . . .	23
그림 6.	word2vec으로 부터 CA상대적 중요도 도출 . . . . .	25
그림 7.	미국 특허 단어빈도 테이블(word frequency table) . . . . .	28
그림 8.	HOQ level 1 . . . . .	30
그림 9.	HOQ level 2 . . . . .	32
그림 10.	CIS 특성 기술 우선순위와 그 변화 . . . . .	33

# 표 목 차

표 1.	온라인 리뷰에 빅데이터 분석을 적용한 연구 . . . . .	12
표 2.	QFD 기법 적용 연구 사례 . . . . .	13
표 3.	아마존 고객 리뷰 데이터 . . . . .	18
표 4.	word2vec 결과 . . . . .	24
표 5.	EC항목과 연계한 CPC code table . . . . .	26
표 6.	CIS 특성 기술 우선순위와 그 변화 . . . . .	34

# 제 1 장

## 서론

### 1.1 연구배경 및 목적

1991년 최초로 코닥社에서 디지털 SLR 카메라를 출시하며 디지털 카메라 시대가 시작되었고, 2000년대에 필름 카메라에서 디지털 카메라로 대세가 전환되었다. 디지털 카메라는 CMOS image sensor(이하 CIS)라는 광전 변환 소자를 이용하여 필름을 대신한다. 디지털 카메라로 피사체를 찍게되면, 렌즈를 통해 들어오는 빛을 CIS가 디지털 영상 신호로 바꿔 주고, 영상신호를 처리하는 ISP(Image Signal Processing)을 거쳐 최종적으로 우리가 확인하는 이미지가 된다. CIS는 디지털 카메라 뿐만 아니라 스마트폰 카메라, 자율주행 자동차의 센서, 산업용 로봇 등의 눈이 되어 주고 있으며, 4차 산업혁명 시대를 맞이하여 CIS 시장은 급속히 커지고 있다. 시장 조사업체 IC인사이드에 따르면 전세계 CIS 시장 매출액은 2010년 부터 2019년까지 연평균 16.9퍼센트의 성장률로 매년 사상 최고치를 경신하였고, 올해 2020년 코로나로 인한 전세계 성장률 둔화에 주춤하지만 내년 부터 2024년까지 7.2퍼센트의 성장률로 사상 최고치 경신을 이어나갈 것으로 전망 하고 있다. 6개의 CIS가 들어가는 스마트폰과 10개 이상의 이미지 센서가 들어갈 것이라는 레벨4이상 자율주행 자동차와 함께 각종 IT기기, 산업용 이미지 센서에 대한 수요까지 고려하면 앞으로 이미지 센서에 대한 수요는 폭발적으로 늘어날 것으로 예상할 수 있다. 이에 발 맞추어 우리나라 반도체 기업들의 이미지 센서에 대한 투자도 크게 늘어

나고 있으며, 함께 해당 분야의 매출과 기술수준의 발달도 크게 증가하고 있다. 이전까지는 스마트폰이나 CIS 업체를 리드하던 미국과 일본 기업들의 제품전략을 빠르게 따라가는 것이 강점이었던 우리나라 기업들도 점차 기술 경쟁력을 확보함에 따라 현재까지의 패스트팔로어(fast follower) 전략을 벗어나 퍼스트무버(first mover)로의 전환을 진행 중인 단계이다. 이에 빠르게 변화하고 있는 시장 상황에서의 고객 반응을 정확히 파악하여, 이를 기술 기획 현장에 반영하고, 시장을 선도할 수 있는 기술 로드맵 전략 수립이 필요하다. 각 기업과 국가들간의 경쟁은 점점 치열해 지고 있으며, 한 제품을 기획하고 판매하기 까지 2년여 또는 그 이상의 시간이 걸리는 반도체 산업의 특성상 자본과 시간에 오차없이 소비자 요구사항에 최대한 부합하는 기술을 적시적기에 대응할 수 있어야 한다. 본연구는 미국 시장 스마트폰 고객 리뷰 빅데이터 분석을 통하여 고객요구사항을 도출하고, 이를 소비자 니즈를 기술특성에 반영하는 QFD(Quality Function Deployment, 품질기능전개)기법에 적용하여, 기술 로드맵의 출발점이라 할 수 있는 각 기술요소에 대한 우선순위를 제안하고자 한다.

## 1.2 연구 보고서의 구성

본 연구보고서는 총5장으로 구성되어 있다. 각 장의 내용은 다음과 같다.

1장은 서론으로 본 연구의 배경과 목적에 대해 소개하였다.

2장에서는 본 연구의 이론적 배경으로써 선행연구와 함께 1절에서는 텍스트 마이닝, 2절에서는 자연어 처리와 워드임베딩, 3절에서는 QFD(Quality Function Deployment, 품질기능전개)에 대해 설명하였다.

3장에서는 연구 모델 및 실험 데이터에 대해 기술하였다. 1절에서는 활용한 데이터에 대해 설명하였고, 2절에서는 본 연구의 개념적 모델을 제시하였다.

4장에서는 연구결과에 대해 QFD의 HOQ(House Of Quality, 품질의 집)을 기준으로 하여 1단계와 2단계의 2절로 나누어 각 단계에 대한 실험 결과를 정리하고, 그에 대한 해석을 기술하였다.

마지막으로 5장은 결론으로서 본 연구에 대한 요약, 시사점과 함께 보완점에 대해 기술하고, 향후 연구 방향에 대한 논의를 다루었다.

## 제 2 장

### 관련 연구

#### 2.1 이론적 배경

##### 2.1.1 텍스트 마이닝

텍스트 마이닝(Text mining)은 텍스트 형태의 데이터들을 수학적 알고리즘에 기초하여 수집, 처리, 분석, 요약하는 통계적 텍스트 처리 기법들을 통칭하는 용어이다[1]. 텍스트 데이터를 수치화하여, 알고리즘을 적용하고 분석, 요약하는 통계기법을 '텍스트 데이터 접근'이라고 한다[2]. 텍스트 마이닝이 적용되는 전체 텍스트 데이터를 말뭉치(corpus)라고 하고, 말뭉치는 대용량의 정형화된 텍스트 집합으로 정의된다[3]. 텍스트 통계 모델에는 반드시 정형화된 텍스트 데이터가 투입되어야 하며, 이를 위해서 정형화 되지 않은 원래의 텍스트를 정형화시키는 과정을 텍스트 사전처리(preprocessing) 또는 정제화(cleansing)라고 한다. 일반적으로 대용량의 텍스트 데이터에 대하여 공란처리, 대소문자 통일, 숫자표현 제거, 문장부호 및 특수문자 제거, 불용단어 제거, 어근 동일화 처리, 엔그램 적용 등의 전처리 과정을 거치며, 연구 목적에 따라 각각의 전처리 과정은 생략될 수도 있다. 전처리된 텍스트 데이터는 연구 분석을 위해 빈도데이터(word count)나 워드클라우드(word cloud), 또는 문서-문서, 단어-단어 간의 유사도 계산, 토픽모델, 기계학습 등 다양한 통계모델에 이용될 수 있다. 텍스트마이닝은 대용량 텍스트에 대한 전통적 텍스트 분석의 한계

를 극복할 수 있게 하고, AI를 통해 시간과 비용을 효율화 할 수 있다. 이에 본연구에서는 수집한 고객 리뷰 데이터(약110만건)에 대한 전처리를 수행 후, 자연어 처리 모델 중 기계학습 기반의 word2vec을 이용하여 분석을 진행하였고, 모두 R프로그램(version 3.6.3)을 이용하였다.

## 2.1.2 자연어 처리와 워드임베딩

머신러닝은 컴퓨터가 데이터를 학습하는 알고리즘과 기술을 통칭한다. 현상을 설명하거나 미래를 예측하는 용도로 활용하거나, 이미지 인식, 영상처리와 함께 자연어 처리에 있어서도 유용하게 쓰인다. 과거에는 기계가 자연어를 학습하게 하는 방법으로 통계적인 접근을 사용했으나, 최근에는 인공 신경망을 이용하는 방법이 자연어 처리에서 더 좋은 성능을 얻고 있다. 머신러닝에는 지도학습, 비지도학습, 강화학습의 세가지 유형이 있다. 지도학습은 입력데이터와 타겟값을 알고 있는 데이터를 학습하여, 이들의 관계를 모델링하는 학습방법이며, 새로운 데이터에 대한 타겟값을 예측하는데에 사용한다. 비지도학습은 지도학습과는 다르게, 타겟값이 없이 학습하는 것을 말하며, 입력된 데이터 속에 내재되어 있는 특성을 찾아내는 용도로 많이 쓰인다. 강화학습은 주어진 환경에서 행동을 취하게 되고, 자신이 한 행동에 대한 보상을 바탕으로 목적을 달성하는 학습이다. 보상을 최대화 하는 방향으로 학습을 진행하며, 구글의 알파고도 강화학습을 이용해 만들어 졌다. 본연구에서는 비지도학습의 한분야인 워드임베딩(Word embedding)[4] 중 대표적인 모델인 Word2Vec을 활용하여 고객 리뷰 데이터속에 내재되어 있는 요인을 찾아 VOC(Voice of Customer)로 이용한다.

Word embedding은 단어를 벡터로 표현하는 것으로, 사람이 쓰는 자연어를 컴퓨터가 이해하고, 효율적으로 처리할 수 있는 숫자의 나열인 벡

터로 변환하는 방법이다. 단어를 표현하는 방법에 따라서 자연어 처리의 성능이 크게 달라지기 때문에 이에 대한 많은 연구가 있었고, 여러가지 모델들이 있다. 임베딩을 만들때는 크게 카운트 기반의 벡터화 방법인 Bag of Words(DTM), LSA(Latent Semantic Analysis)와 예측 기반으로 벡터화하는 NNLM(Neural Network Language Model), RNNLM(Recurrent Neural Network Language Model), Word2Vec(Google), FastText(Facebook), Glove(Stanford) 등의 모델들이 있다. 그 중 Word2Vec은 Google의 Tomas Mikolov와 그의 팀원들이 발표한 신경망 알고리즘(NNLM)을 활용한 머신러닝 기법으로, Word embedding 학습모델 중 하나이다[5]. 비슷한 위치에 등장하는 단어들은 비슷한 의미를 가진다라는 언어학의 분포가설(distributional hypothesis)에 기반하여, 데이터 학습을 통해 의미가 유사한 단어들을 서로 가까운 벡터공간에 위치하도록 할당한다. Word2Vec이 단어들의 벡터를 산출하는 방식에는 CBOW(Continuous Bag of Words)와 Skip-Gram 두가지 방식이 있다. CBOW는 주변에 있는 단어들을 가지고, 중심(목표) 단어를 예측하는 방법이며, Skip-gram 모델은 입력단어를 중심으로 주변 단어들을 예측하는 방법이다. CBOW와 Skip-gram의 인공신경망을 도식화하면 그림6와 같다. CBOW의 경우 입력, 출력 학습 데이터 쌍이 문맥단어4개, 타겟단어(w) 하나인 반면, Skip-gram의 학습 데이터는 타겟단어(w(t))와 함께 타겟 직전 두번째 단어(w(t-2)), 타겟 직전 단어(w(t-1)), 타겟 다음 단어(w(t+1)), 타겟 다음 두번째 단어(w(t+2)) 들 각각과 4개쌍이 된다. Skip-gram이 같은 말뭉치로도 더 많은 학습데이터를 확보할 수 있어 임베딩 품질이 CBOW보다 정확성 등의 측면에서 더 우수한 것으로 알려져 있다(Mikolov et al. 2013).

이번 연구에서는 아마존 고객 리뷰에 Skip-gram 학습 방식을 이용하여, 주어진 키워드와 유의미한 단어들을 추출하여 VOC(voice of cus-

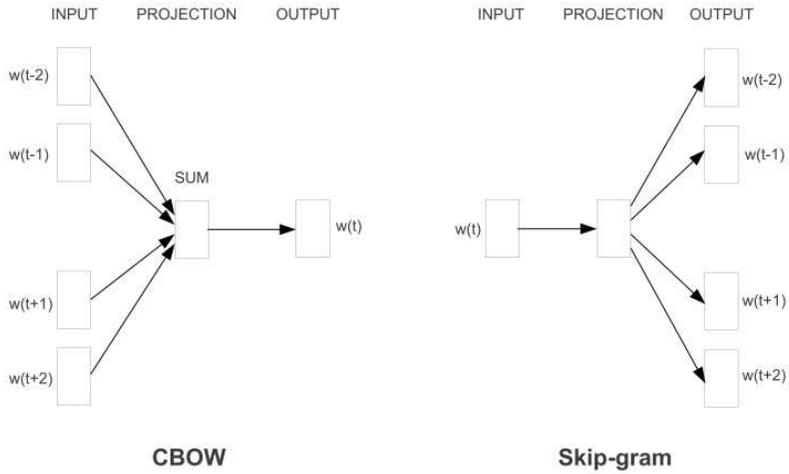


그림 1: CBOW와 Skip-gram의 학습과정

source : Mikolov et al(2013), Efficient estimation of Word representation of Vector space

tomers)로 활용하고자 한다. Skip-gram 모델은 타겟단어가 주어졌을 때 주변단어가 나타날 확률을 식 2.1과 같이 나타내며 이를 최대화하는 방향으로 진행한다.  $T$ 는  $w_1, w_2, w_3, \dots, w_T$ 로 표현되는 각 단어의 위치이며,  $c$ 는 현재 위치 단어인  $w_t$  주변에 등장할 것으로 예측하는 단어의 범위가 된다.  $c$ 가 클수록 더 많은 단어가 현재 위치의 단어와 유사한 단어로 정의되며,  $\frac{1}{T}$ 는 전체 단어에 대하여 위치의 수로 나누는 정규화 항으로,  $\theta$ 는 최적화되는 모든 변수를 의미하게 되며, 확률 값을 최대화할 수 있도록 하는 각각의 단어 벡터 값들이 된다[6].

$$\frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log p(w_{t+j} | w_t) \quad (2.1)$$

이를 참고하여 Word2Vec의 Skip-gram 모델로 부터 얻은 단어들의 벡터를 활용하여 타겟 단어 'camera'가 주어졌을 때 벡터 공간 내에서 주변 단어들과 연관성 정도를 코사인 유사도를 이용해 측정하였다.

### 2.1.3 코사인유사도

본 연구에서는 아마존의 'Cellphones and accessories' 고객 리뷰 데이터를 수집하여, 이를 Word2Vec의 Skip-gram 모델을 통해 분석하였고, 이때 분석된 벡터값을 이용해 'Camera'와 유의미한 관계의 단어들을 추출하였다. 일반적으로 벡터 공간에서 두 벡터간의 유사도를 측정할 때는 두 지점 사이의 직선 거리를 측정하는 유클리디안 디스턴스(Euclidean distance)를 통해 구하거나, 사잇각을 측정해서 구한다. 이 때 사잇각을 이용하는 방법 중 두 벡터 간의 코사인 각도를 측정해서 두 벡터가 얼마나 유사한지를 판단하는 것이 코사인 유사도(Cosine distance)이다. 코사인 유사도는 보통 검색엔진에서 검색어와 문서의 유사도를 구할 때 쓰거나, 텍스트 마이닝에서 단어 집합들간의 유사도를 구할 때 많이 쓰이게 된다. 식 2.2 과 같이 벡터의 내적을 두 벡터의 크기의 곱(norm)으로 나누면서 코사인 유사도를 구하게 된다. 두 단어 벡터의 A와 B사이의 각도는 0도가 되면 코사인 유사도 값은 1이 되고 90도가 되면 0이 된다. 두 벡터가 1에 가까워 질수록 A와 B는 일직선상에 위치하게 되면서 두 단어의 유사도는 높아진다고 할 수 있으며, 코사인 유사도를 통해 측정된 수치로 두 단어가 얼마나 유사한가에 대해 표현할 수 있다.

본 연구에서는 벡터 공간내에서 'Camera'와 주변 단어들의 코사인 유사도 수치를 구해서 유의미한 단어들을 파악한 후, HOQ(House Of Quality)의 CA(Customer Attribute) 항목의 상대적 중요도 분석에 이용하였다.

$$similarity = \cos(\Theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}} \quad (2.2)$$

## 2.1.4 품질기능전개(QFD)

‘고객의 소리를 최종 제품으로’라는 한마디로 대표되는 품질기능전개(Quality Function Deployment, 이하 QFD)는 신제품 개발 또는 품질향상을 목적으로 고객요구를 반영하여 제품 또는 서비스를 빠른 시간 내에 개발 또는 개선하게 하는 방법론 중 하나이다. QFD는 역사적으로 1960년대 일본의 아카오 요지에 의해서 제시된 품질 기법으로 품질기능전개란 단어는 1972년 일본의 ”신제품 개발과 품질 보증”의 주제로 발표된 연구에서 ”품질전개시스템”이 사용된 것을 시초로 삼고 있다[7]. 이후 1972년 미쓰비시 중공업 고베 조선소에 최초로 적용되어, 주로 제조업에서 고객 요구에 부합하는 제품을 설계하기 위해 사용되기 시작되었으나, 1980년대 세계적으로 확산되어 제조업뿐만 아니라 다양한 산업에 적용되었을 뿐만 아니라 제품이나 서비스가 창출되는 비즈니스의 전 분야에 널리 적용되고 있다. 이전의 품질 관리 기법들이 대부분 내부적으로 불량을 줄이고 품질을 개선하는 것에 초점을 두고 있었다면, QFD는 고객니즈를 만족시킬 수 있는 제품을 개발하는 것에 초점을 두고 있다. 고객니즈를 만족할 만한 상품을 위한 품질향상 항목들을 정리하고, QFD의 핵심적인 도구인 품질의 집(House Of Quality, 이하 HOQ)을 통해 기술 특성으로 변환하여 고객의 니즈와 기술을 연계시킨다. <그림2>과 같이, HOQ의 구성은 크게 소비자 요구사항(CA, Customer Attribute), 각 CA에 대한 상대적 중

요도, 제품 설계속성(Engineering Characteristic), CA-EC간의 관계점수로 된다. 먼저 ①고객의 요구사항과 각 요구사항에 대한 ②상대적 중요도를 설문조사, 컨조인트 분석, AHP, 델파이 기법 등의 통계 분석도구를 통해서 수집 및 결정하게 된다. 다음으로는 고객의 니즈를 만족시키는 데 필요한 ④제품 설계특성(EC), 기술/서비스 속성등을 합리적인 방법론을 기반으로 결정해야 한다. 충분히 파악된 후에는, ⑤ CA와 EC간의 연관 관계를 작성하고, 이후 고객의 요구를 충족시키기 위한 EC간의 중요도 및 우선 순위를 평가하게 되며, 최종 HOQ 단계를 거치면 고객의 요구사항이 기술 특성들에 반영되며, 기술 요소의 우선순위를 도출할 수 있게 된다.

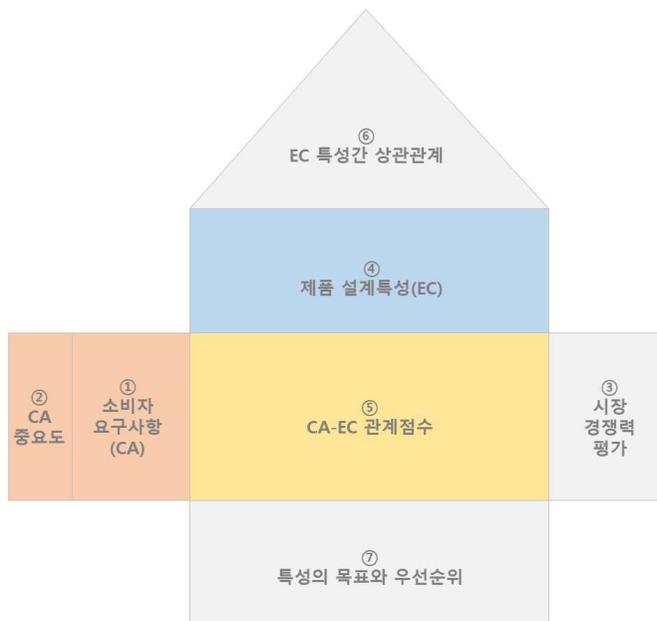


그림 2: House Of Quality

## 2.2 선행 연구

### 2.2.1 온라인 리뷰에 빅데이터 분석을 적용한 연구

그동안 ICT의 발달과 함께 지리적, 시간적 제약 없이 편리하게 이용할 수 있는 온라인 거래는 꾸준히 증가하였으며, 최근들어 COVID19 등으로 인해 급속히 늘어나고 있다. 소비자들은 모바일앱 이나 인터넷으로 필요한 물품이나 서비스를 구매하거나 또는 관련 경험에 대한 리뷰를 남긴다. 이를 통해 이루어 지는 구매자와 잠재 구매자들간에 커뮤니케이션을 온라인 구전(eWOM, electronic Word of Mouth))이라고도 한다[8]. 기업들은 온라인 리뷰를 통해 자사 제품에 대한 피드백을 확인하고, 새로운 제품에 대한 아이디어로도 활용하고 있다. 소비자들은 고객 리뷰의 내용과 그 양에 직접적으로 영향을 받으며[9], 이를 통해 리뷰를 이용하여 소비자들의 선호도에 대한 영향파악과 해당 제품에 대해 가격과 수요 예측에도 활용할 수 있다.[10]. 온라인 리뷰는 제품 구매에 대한 온라인 쇼핑물 리뷰 뿐만 아니라, 관광지에 대한 리뷰에 word2vec기법을 활용하여 관광지에 대한 측정도구 지표가 될 수도 있으며[11], 취업사이트 리뷰에 텍스트 마이닝과 word2vec모델을 사용하여 기업들의 직무만족 측정에도 적용하였다[12]. 또한 SNS(Social Network Service) 데이터를 이용하기도 한다. 트위터 텍스트와 구글맵 리뷰를 분석하여 장소에 대한 선호도를 측정하고[13], 인스타그램의 해시태크를 이용하여 워드클라우드와 단어간 유사어를 추출하여 소비자 인식 분석에도 활용하였다[14]. 본연구에서는 아마존 고객 리뷰 데이터에 대해 word2vec모델을 이용한 분석을 통하여 소비자 니즈를 정량화시켜 CA 상대적 중요도로 활용하고, 각각의 EC항목과 연관된 미국 특허 데이터를 수집하여 텍스트 마이닝 통계분석을 통해 QFD 기법에 적용하여 분석을 진행하였다.

저자	방법론	주제와 연구 내용
심영석, 김홍범 (2018)	Word2vec	Word2vec을 활용한 관광지 개성과 여행객 평점간 구조적 관계 분석, 온라인 리뷰에 Word2vec 모델을 활용하여 관광지 개성을 추출하고 WLS(가중최소제곱법)을 통해 모형 추정에 활용하여 관광지 개성 계수를 추출하여 소비자의 평점과의 연관성을 알아봄
오유진, 정재윤, 조남욱 (2018)	Word2vec, LDA	Word2vec과 LDA기법을 이용한 게임 리뷰 분석, LDA와 Word2vec을 이용하여 주제를 선정, 감정단어와 평점에 대한 회귀식을 이용하여 감정변수가 리뷰내용을 어떻게 반영하는 지에 확인함.
안병대, 최진욱, 서용무 (2020)	Word2vec	기업개성이 직원의 직무만족과 기업 이직률의 관계에 미치는 영향, Word2vec 모델을 활용해 기업 개성 설명 사건을 만들고, 이 사건을 통해 직원들이 남긴 리뷰를 분석해 각 기업의 기업 개성을 측정함
정예림, 김지희, 유헌선 (2020)	Word2vec	Word2vec을 활용한 제품군별 시장규모 추정 방법에 관한 연구, 제품 정보 데이터들에 대해 Word2vec 알고리즘을 적용하여 제품군 도출, 각각의 제품에 대한 매출액 정보를 호출하여 해당 제품군의 매출액 규모 추정, 이를 통해 사전에 정의 되지 않은 세부적인 수준의 품목과 제품군에 대해서 시장규모 추정.
정명석, 이주연 (2018)	LDA	LDA 모델 기반의 인공지능 기술 관련 연구 활동 및 동향 분석, 인공지능 기술과 관련된 논문 데이터에 대해 LDA 적용하여 핵심 키워드를 분류하고, 해당 키워드들에 대한 출현빈도와 시계열 분석을 통해 연구 동향 분석.
김인규 (2016)	Word2vec	텍스트 마이닝과 머신러닝을 통한 모바일 게임앱 리뷰 분석, 게임 리뷰를 수집하여 게임인기의 지속성 유무와 유료앱, 무료앱으로 분류하고, 각 분류별 텍스트들에 대해 Word2vec 적용하여 추출단어들의 차이를 비교 분석.
홍석은 (2018)	LDA, 감성분석	빅데이터 마이닝에 의한 환경 민원의 감성 분석 및 공간 패턴 분석, 환경 민원 데이터에 대한 단어를 사전 구축하여 LDA 이용하여 민원 주제 별 단어 추출 후, 민원별, 토픽별 감성 분석과 공간모델링 적용하여, 토픽별 감성 분포와 환경 민원의 지역적 변수 정량화.
채인영, 김지영, 유기윤 (2017)	감성분석	트위터 데이터와 구글 리뷰를 이용한 장소 선호도 분석 방법, 트위터 데이터를 대상으로 서울의 테마공원 POI를 포함하는 텍스트를 추출하여 언급빈도수와 감성분석을 실시하고 구글맵 테마공원 POI에 대한 리뷰를 수집하여 평점별 감성 분석과 감성이 추출하여 테마공원 선호도 산출.

표 1: 온라인 리뷰에 빅데이터 분석을 적용한 연구

## 2.2.2 QFD기법 적용 연구 사례

QFD는 초기에는 대부분 제조업 분야에 적용되었지만 최근에는 신제품 개발, 서비스 품질 관리 등에도 적용 분야가 점점 확대되어 왔으며, 연구 목적에 따라 QFD 적용분야나 절차가 매우 다양하게 적용되고 있다. 이를 국내외의 연구 사례들을 통해 알아보았다. 먼저 본연구의 주제와도 비슷한 카메라의 적용사례를 알아보면, QFD와 Fuzzy이론을 적용하여 DSLR 신제품 개발에 대한 연구가 진행되었다. 소비자 요구 사항 파악을 위해 퍼지이론을 이용하여 객관적인 수치로 변환하여 QFD에 적용하여 DSLR 신제품 개발에 필요한 핵심요소들의 우선순위를 도출하였다[15]. 뿐만 아니라 Kano 모델을 도입하여 고객 설문조사 사항을 분석후 QFD에 도입하여 DSLR 제품 설계 특성의 우선순위를 도출한 사례[16]도 있다. 또한, QFD를 이용하여 고추장 소스 신제품 콘셉트를 개발한 사례도 있다. 제품 전문가와 해외 현지 소비자를 대상으로 설문조사를 실시하였

다. 특히 소비자와 전문가의 설문지 특성 문항은 Delphi 기법을 통해 산, 학, 연 전문가들의 의견을 통해 추출되어, 각각의 기술 특성의 난이도와 상호관계를 고려하여 기술특성 우선순위를 매겨 분석을 진행하였다[17]. 그리고 AHP기법을 이용하여 철도기술 항목들간의 상대적 중요도를 분석하여, 이를 QFD에 적용하여 철도서비스품질 개선 방안을 마련한 사례도 있었다[18] 각 사례를 통해 QFD의 적용 분야와 적용 절차를 살펴보면, 전통적으로 소비자 설문 조사를 통해 소비자 니즈를 파악하고, 그 과정에서 각종 기법을 도입하여 소비자 요구사항을 수치화 하거나, 상대적 중요도를 평가하였다. 본연구에서는 소비자 설문조사를 대신하여 빅데이터 분석을 통해 소비자 요구사항 상대적 중요도를 추출하고, 이를 QFD기법에 적용시켜 CIS 기술특성 우선순위를 도출하고자 하였다.

저자	방법론	주제와 연구 내용
정영미 (2014)	QFD, AHP, 포커스 그룹 인터뷰	도서관 서비스 품질 향상을 위한 QFD 적용에 관한 사례 연구. 도서관 이용자의 요구 사항을 파악하기 위해 포커스 그룹 인터뷰를 진행하고, AHP를 이용하여 이용자들을 대상으로 각 요구 사항의 중요도와 만족도를 측정. 실무 인터뷰를 기술 특성 도출 하여 QFD 진행
김인규, 박태한, 정연찬 (2019)	QFD, 역공학기법, 설문조사, 인터뷰	역공학과 QFD를 활용한 퍼스널 모빌리티 시스템 요구사항 도출. 경쟁사 제품 분석을 통한 역공학 기법을 적용하여 전동킥보드 시스템 요구사항을 도출함. 설문조사와 인터뷰를 통해 소비자 요구사항 도출하여 QFD 진행
정광태, 김혜선, 전경진, 원병희, 홍재수 (2010)	인터뷰, QFD	휠체어 디자인 개선을 위한 사용자 분석 및 QFD 적용방안. 실사용자들과의 인터뷰를 통해 사용자 요구사항을 도출하고, 휠체어 설계 지침 문헌으로 부터 디자인 요소 결정. 인간공학 전문가 평가를 통해 사용자 니즈 상대적 중요도 결정하여 QFD 진행
송상민, 조인희, 조재립 (2010)	QFD, 설문조사, 다중회귀분석, AHP	QFD를 활용한 스마트폰의 엔터테인먼트 특성 평가. 스마트폰 기능을 분류하여 만족도에 영향 미치는 정도를 설문조사를 통해 분석. 주성분 분석과 크론바하 알파 계수를 이용하여 요인분석과 신뢰성 분석 시행. 다중회귀 분석을 통해 속성별 유의 정도 분석하고, AHP를 통해 특성들 상대적 중요도 측정.
이슬기, 김아영, 홍상필, 이승재, 이민아 (2015)	QFD, Delphi	QFD 방법을 적용한 고추장 소스 콘셉트 개발. 산, 학, 연 전문가들에 대해 Delphi 기법을 통해 설문지의 문항을 구성하고, CA항목은 소비자 요구사항 조사를 통해 얻은 결과를 제시하고, 기술적 특성은 소스의 특성으로 하여 HOQ 작성. 상대적 중요도는 단순가중합 방법을 사용하여 산출하여 진행.
박기남, 조재균, 정석찬, 진종근 (2002)	QFD, Conjoint analysis	품질기능전개와 컨조인트 분석을 이용한 이해관계자 주도의 시장지향적 e-비즈니스 커리큘럼 개발 방법론에 관한 연구. 실무 관계자들과의 설문을 통해 CA항목을 정하고 컨조인트 분석을 통해 상대적 중요도를 산출하고, 전공 교과목을 EC 항목으로 하여 HOQ 작성.

표 2: QFD 기법 적용 연구 사례

## 제 3 장

# 연구 방법

### 3.1 실험 데이터

본 연구에 활용한 데이터는 총 4가지로 미국 아마존(www.amazon.com) 고객 리뷰 데이터, 프랑스의 글로벌 카메라 화질 평가 기관인 DXOmark社(www.dxomark.com)의 스마트폰 카메라 화질 평가 항목, 카메라 관련 미국 특허, 삼성전자(www.samsung.com /semiconductor) CMOS 이미지 센서 반도체 최신기술 스펙 등을 활용하였다.

#### 3.1.1 Amazon社의 고객 리뷰 데이터

먼저, 고객니즈 분석을 위한 미국 아마존 고객 리뷰 데이터는 1996년에서 2016년 까지의 미국 아마존 고객 리뷰 오픈 데이터를 확보하였다[19]. 스마트폰 카메라 핵심부품으로서의 Mobile CIS(Cmos Image Sensor) 기술 로드맵 수립이 목적이므로, 전체 제품 카테고리 중 "Cellphones and Accessories" 카테고리에 해당하는 고객리뷰 데이터(1,128,437건)를 이용하였다. 텍스트 마이닝을 통한 리뷰 데이터의 전처리 단계에서 연도별로 분리하여 DTM(document term matrix)으로 구조화하였고, 연도데이터와 review데이터를 남겨 데이터프레임을 구성하여, 이 중 2016년 데이터(311,798건)를 HOQ(House of Quality)의 CA항목 작성시 활용하였다.

### 3.1.2 DXOmark社의 스마트폰 카메라 화질 평가 항목

DXOmark image labs(www.dxo.com)은 스마트폰 카메라를 메인인 것으로 하여 디지털카메라와 렌즈의 화질과 성능을 평가하며, 스코어를 부여하는 세계적으로 공신력 있는 프랑스의 카메라 화질 리뷰 회사이다. 세계 유수의 언론사에서 스마트폰 카메라 성능 참조자료로 기사에 이용하며, 스마트폰 제조사들도 경쟁사와의 벤치마킹 스코어를 참조하여 개발에 참고한다. 스마트폰 카메라의 경우, 화질 평가는 메인 카메라, 전면 카메라, 줌, 비디오로 나뉘어 진행되며, 각각의 평가 항목과 점수가 별도로 공개된다. 그 중 exposure and contrast, color, Auto-focusing, Texture, Noise, Zoom, Bokeh, Wide angle, Night scene를 EC항목으로 적용했다. 여기서 Bokeh는 일본어이며, blur 또는 not in focus의 의미로 의도적으로 피사체의 배경을 흐리게 만드는 효과를 말한다. 본 연구에서는 DXOmark의 메인 카메라 평가항목을 HOQ level 1에서의 EC(Engineering Characteristic)으로 이용함으로써 아마존 고객 리뷰로 부터 분석된 고객 니즈를 전문가 화질평가와 연계시켰다.

### 3.1.3 미국 특허 데이터

특허는 유망 신기술에 대한 지적 재산을 확보하고자 등록하게 된다. 특히 신기술에 대한 정보로 인해 기술예측이나 기업 경영의 핵심정보로서의 역할을 수행하고 있다. 이 중 미국 특허는 세계 주요 기업들이 국제적으로 보호받기 위하여 가정 먼저 출원하는 곳이다[20]. 따라서 최신 주요 기술들에 대한 정보가 충분할 것으로 판단되어 본 연구는 미국 특허청(United States Patent and Trademark Office, USPTO)의 2016~2018년 출원 특허 중 HOQ 1단계 EC항목들과 연관된 특허 데이터를 CPC(Cooperative

patent classification)코드를 통해 분류하였고(4231건), 이중 초록부분을 텍스트 마이닝을 통한 전처리후 HOQ 작성에 활용하였다.

### 3.1.4 삼성반도체 이미지 센서 최신제품 적용 기술

삼성반도체([www.samsung.com/semiconductor](http://www.samsung.com/semiconductor))는 CIS(CMOS Image Sensor) 세계 시장을 선도하고 있는 업체 중 하나이다. 신제품 라인업에는 업계 최고 경쟁력의 CIS 신기술들이 적용되어 있어, 해당 신기술들은 현시점에서 상용화된 가장 최신 기술이다. 따라서 삼성 CIS제품(HMX, HM1)에 적용된 기술을 활용하면 충분히 CIS 선행 기술 전망의 근거가 될 수 있을 것으로 보인다. 이러한 점에 착안하여 HOQ 2단계 EC항목은 삼성반도체 CIS 신제품에 적용된 최신기술 스펙을 참고하여 작성하였다.

## 3.2 연구 모델

본 연구는 QFD(Quality Function Deployment)를 메인으로 하여, 자연어 처리 분야의 머신러닝 알고리즘 기반 Word2vec 모델을 활용하여 스마트폰 카메라에 대한 소비자 니즈를 파악하는 1단계와 추출된 소비자 니즈를 QFD의 주요 도구인 HOQ(House of Quality)의 CA(고객요구사항)으로 활용하여 CIS(CMOS Image Sensor) 기술로드맵의 기초가 되는 CIS 최신기술들의 우선순위를 알아보는 2단계로 나눌 수 있다. 간략히 그림으로 보자면 아래 <그림3> 와 같다.

### 3.2.1 머신러닝을 통한 소비자 니즈 파악

1단계는 소비자 니즈 파악 단계로, 스마트폰에 대한 미국의 아마존([www.amazon.com](http://www.amazon.com)) 고객 리뷰 데이터[19]를 수집하여, 텍스트 전처리를

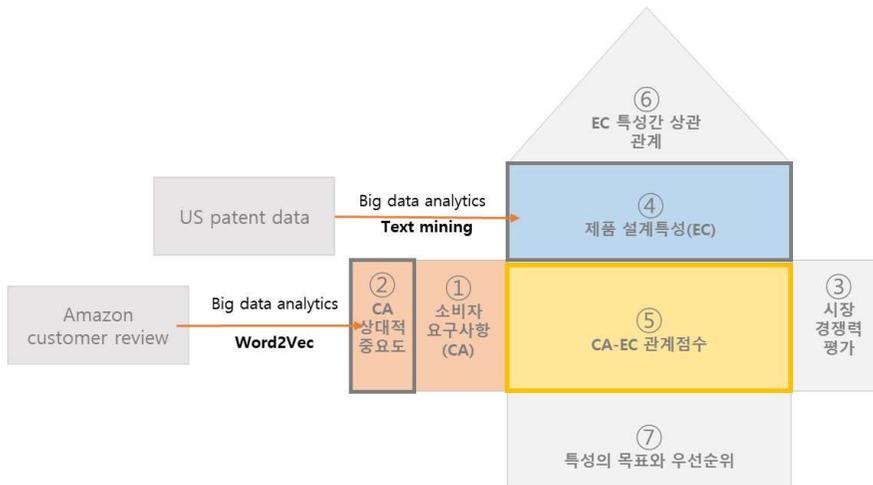


그림 3: 연구 절차

수행하였다. 이후, 'camera'를 타겟단어로 하여 Word2Vec모형을 적용하여 주변단어의 벡터값을 산출하였고, 함께 카메라 주요 스펙별로 카테고리화 하였다. 각각의 카테고리화 하고 함께 분류된 단어들의 코사인 유사도 수치를 2단계 HOQ의 CA(Customer Attribute)으로 활용하였다. 텍스트 마이닝과 머신러닝은 모두 오픈소스 기반 통계 분석 프로그램인 R version 3.6.3을 이용하여 구현하였다.

2단계 HOQ의 CA(소비자 요구 사항)과 그에 따른 상대적 중요도는 AHP(Analytic Hierarchy Process, 계층적 분석과정)이나 설문 기법인 컨조인트 분석을 이용하여 파악할 수도 있으나, 110만건의 아마존 고객 리뷰 빅데이터로부터 도출한 소비자 요구사항에 대한 분석을 이용하여, 점점 커가는 이커머스 시장에서의 소비자 VOC(Voice of Customer)로부터 도출한 고객 니즈를 현장 기술 개발에 반영하고자 시도하였다.

구체적으로 살펴보면, 수집한 아마존 고객 리뷰 데이터는 <표3> 와 같은 형태의 CSV포맷의 영문으로 되어 있으며, 평점, 리뷰어ID, 아마존제품ID, 리뷰어 이름, 리뷰텍스트 등 10개로 구성되어 있다. 이들 중, 리뷰텍

#	rating	reviewTime	reviewerName	reviewText	summary
0	5	08 4, 2014	Claudia Valdivia	Looks even better in person. Be careful...	Can't stop won't
1	5	02 12, 2014	sarah ponce	When you don't want to spend a whole...	I
2	3	02 8, 2014	Kai	so the case came on time, i love the...	Its okay
3	2	02 4, 2014	Sharon Williams	DON'T CARE FOR IT. GAVE IT AS A...	CASE
4	4	02 3, 2014	Bella Rodriguez	I liked it because it was cute, but the...	Cute!
5	2	01 27, 2014	Amazon Customer	The product looked exactly like the...	Not so happy
6	3	01 23, 2014	DaMara Estep	I FINALLY got my case today. It took...	It's cute!
7	5	01 17, 2014	Ashley Nicole Miller	It is a very cute case. None of the...	Cute case
8	1	12 27, 2013	BeeLove21	DO NOT BUY! this item is seriously...	WORST ITEM!
9	4	12 16, 2013	Mrs. Ochoa	I really love this case... you have to...	Pretty Cute!
10	4	10 27, 2013	Paige	its super cute and makes my phone...	case
11	5	10 23, 2013	E. Bryce	Another great product that my daughter...	Bling bling for
12	1	10 8, 2013	Tina	Very cheap broke the first time we put...	cheap plastic
13	4	09 19, 2013	barb	:( I purchased this for my grand-daughters...	Great
14	5	09 10, 2013	KaitlynxO625	Beautiful quality and outstanding product!...	I can't stop using
15	5	08 28, 2013	M. Antillon	It is such a good case for a low price. I have...	I love it

표 3: 아마존 고객 리뷰 데이터

스트와 그에 따른 연도 데이터를 선택 추출하여 고객 니즈 분석에 활용하고자 하였다. 텍스트 마이닝을 통한 전처리 단계에서는 우선 연도별 리뷰 데이터의 데이터 프레임을 생성하였고, 생성된 데이터 프레임을 바탕으로 공란처리, 대소문자 통일, 불용단어 제거, 문장 부호 및 특수 문자 제거 등을 거쳐 텍스트 데이터에 대해 복잡도를 감소시켜 분석에 효율과 정확성을 높였다. 이후, 전처리를 거친 데이터들에 대하여, 첫째로 연도별 고객 리뷰수 변화와 연도별 'camera' 언급 빈도를 확인하여, 스마트폰과 스마트폰 카메라에 대한 관심변화가 빅데이터 분석으로는 어떻게 나타나는지 확인하였다. 둘째로, 수집된 데이터 중, 가장 최근의 데이터를 분석에 활용하기 위해 2016년 고객 리뷰를 추출하여 word2vec의 Skip-gram 모델로 학습시켜, 벡터 공간내에서 타겟단어 'Camera'와 유사한 주변 단어들의 벡터값을 기반으로 타겟 단어와 주변 단어들간의 코사인유사도 수치를 추출하였고, 코사인 유사도 상위 100개 단어를 나열 후, HOQ(House of Quality)의 CA(Customer Attribute)항목들로 정한 Resolution(pixel), Lens,

Auto-focusing, Features, body의 속성에 해당하는 단어들로 각각 분류하였다. 각각의 CA항목들은 스마트폰 카메라나 디지털 카메라의 스펙 부분에 기재되는 주요 특징들을 종합하여 선정하였으며, 각각의 항목에 분류된 단어들 중 코사인 유사도 기준 상위5개 단어들을 추출하여, 해당 단어들의 코사인 유사도 수치들을 평균내어 CA항목들에 대한 상대적 중요도 값으로 활용하였다.

### 3.2.2 QFD를 통한 기술속성 우선순위 도출

2단계에서는 글로벌 스마트폰 카메라 화질 평가기관인 DXOmark社(www.dxomark.com)의 스마트폰 카메라 화질 평가 항목들을 EC(Engineering Characteristic)로 하여 CA(Customer Attribute)항목과 1단계에서 도출된 각 항목에 대한 CA 상대적 중요도 값과 함께 HOQ 1차 레벨을 구성하였다. 2016년에서 2018년까지의 미국 특허 데이터를 수집하였고, 이를 텍스트 마이닝을 진행하여 CA-EC간 관계행렬(Relationship Matrix)를 작성하였다. 2016 ~ 2018년 미국 특허 데이터 중 EC항목인 exposure and contrast, color, Auto-focusing, Texture, Noise, Zoom, Bokeh, Wide angle, Night scene 과 연관있는 특허들을 선진분류체계인 CPC(Cooperative patent classification) code를 이용하여 수집하였다. 이 후, 각각의 항목에 해당하는 특허 데이터들에 대해 텍스트 마이닝을 통한 전처리를 실행하였다. 아마존 고객 리뷰 데이터들에 대한 전처리와 마찬가지로, 공란 처리, 대소문자 통일, 불용단어 제거, 문장 부호 및 특수 문자 제거 등을 진행했고, 전처리된 데이터들에 대해 100위까지의 단어 빈도표(frequency table)을 작성하였다. 한 문서에서 나타나는 단어의 출현빈도가 그 문서의 성격을 대변한다는 BOW(Bag of Words)가정에 근거하여 발현빈도 순위에 따라 ◎(9점), ○(3점), △(1점), 공란(0점)으로 점수에 차등을 두어 CA-EC

관계행렬(relationship matrix)를 작성하였다.

HOQ level 2에서는 삼성반도체([www.samsung.com/semiconductor](http://www.samsung.com/semiconductor)) CMOS 이미지센서 최신제품 기술 스펙을 EC항목으로 하였고, 전문가 판단에 의한 기술 요소들의 중요도, EC-EC 관계매트릭스, 향후 각 기술요소의 중요도 변화를 분석함으로써 빅데이터로부터 도출된 소비자 요구사항을 최종적으로 전문가 의견과 연계하였다. 시간에 따른 소비자 선호의 중요도 변화를 고려한 연구[21]를 참조하여, 추가로 3년후 속성기술의 중요도 변화는 어떨지에 대한 의견을 반영할 수 있게 HOQ 테이블에 추가하였다. 관련 업종 8년 이상의 개발자, 마케팅, 상품기획 담당자 8명에게 설문을 실시 하였으며, EC-EC 연관관계는 1단계와 동일하게 ◎(9점), ○(3점), △(1점), 공란(0점)으로 기술간의 연관 관계의 강도에 따라 점수를 배정하여 화질 평가 항목과 속성기술간의 관계 강도가 가장 높을시 9점을 주었다. 또한 미래의 기술 중요도 변화는 -100 ~ 100% 으로 비율 점수 분포를 두어 현재 대비, 3년후, 5년후의 시간의 변화에 따라 중요도 증가 될 것으로 판단될시 0 ~ 100%, 중요도가 감소한다고 판단될시 -100 ~ 0% 으로 측정하였다. 마지막으로 각 기술요소의 중요도는 10점 척도로 하여, 기술들의 중요도를 파악했으며, 우선순위 산출에 활용하였다. 해당 HOQ는 다음 절인 연구결과 섹션에서 세부적으로 다루겠으며, 최종적으로 1단계와 2단계 전체 연구 모형을 도식화 하면 <그림4>과 같다.

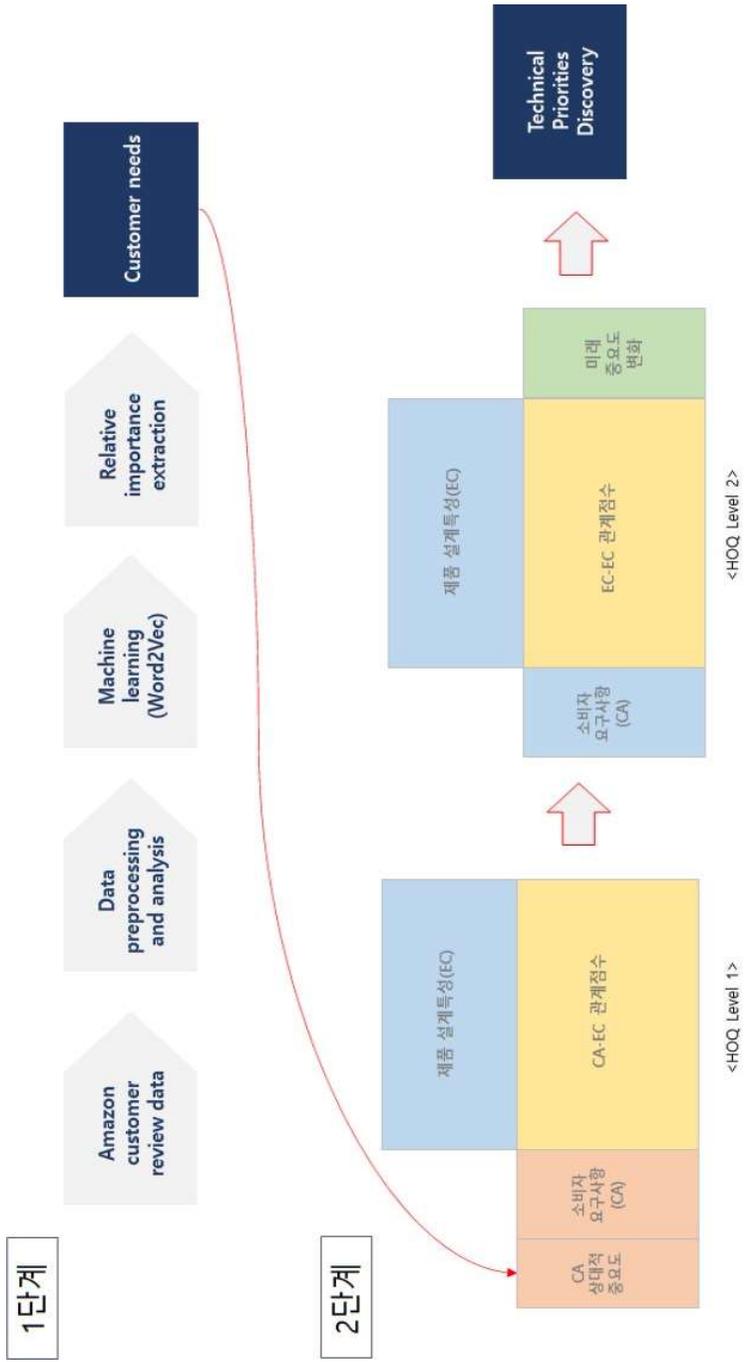


그림 4: 전체 연구 모형

## 제 4 장

# 연구결과

### 4.1 1단계 실험결과

#### 4.1.1 데이터 전처리

수집된 1998년 ~ 2016년의 아마존 고객 리뷰 데이터(1,128,437건)을 본 연구를 위한 기술 통계 분석이나 Word2vec 학습에 사용하기 위해서 전처리 작업을 진행하였다. 데이터의 복잡도를 감소시키기 위해 10개행의 데이터 중 연도와 리뷰데이터 본문만을 남겼고, 이후 연도별 csv파일로 분류하여 corpus(말뭉치)형태로 구성하였다.

연도별로 나뉜 대용량의 텍스트를 포함하는 문서를 정제화하기 위해 개별 텍스트 데이터를 처리하는 R베이스 함수나 stringr 라이브러리 함수 대신 tm 라이브러리 함수를 이용하였다. 일반적으로 전처리 작업은 공란처리, 대소문자통일, 숫자표현제거, 문장 부호 및 특수문자 제거, 불용 단어 제거, 어근 동일화 처리, 엔그램 처리 등의 과정을 거치게 되나, 연구 목적에 맞게 조정하게된다. 예를 들어 본연구의 word2vec 학습의 경우, <그림5>와 같이 타겟단어 'camera'와 연관된 코사인 유사도 기준 상위 30개의 단어를 비교해보았다. 일반적인 7가지의 전처리 과정을 모두 거친 좌측과 word2vec 결과를 검토하며 선별 적용한 우측 도표이다. 우측 표를 보면 1,2,5위가 5mp, 13mp, 8mp이다. 카메라 화소와 연관된 굉장히 의미있는 단어임에도 불구하고, 숫자표현 제거를 하면 확인할 수 없는 데

이터이다. 그외에도 어근동일화 실행시, 의미를 해석할 수 없는 단어들 (afe, blockaq, oi,, etc) 또는 비정형 데이터를 사전처리 함에 따라 같은 단어가 여러 형태로 나타나는 경우(camera - camer, proximity - proxim,,etc) 등 연구자 스스로가 자의적인 해석을 내려야 하는 경우가 많았다. 그래서 최종적으로는 공란처리, 대소문자통일, 특수문자제거, 불용단어 제거 등을 적용하였고, 우측 테이블에도 음영처리 했듯이 카메라와 관계없는 단어들은 추가적으로 분류하여 불용단어 처리를 하였다.

y2016	word	dist
1	camera	0.614823
2	shadow	0.601839
3	megapixel	0.589773
4	apertur	0.566949
5	ncamera	0.564168
6	flash	0.563117
7	vignet	0.547446
8	slr	0.535253
9	unobstruct	0.522406
10	zoom	0.519289
11	dslr	0.516557
12	len	0.515953
13	telephoto	0.509185
14	afe	0.505933
15	blockag	0.505292
16	blur	0.504696
17	resol	0.5036
18	af	0.495884
19	proxim	0.495485
20	obscur	0.495276
21	sensor	0.494693
22	laser	0.492315
23	captur	0.486026
24	camer	0.485624
25	nikon	0.477073
26	microphon	0.470272
27	oi	0.470167
28	hdr	0.469863
29	ambient	0.463074
30	nflash	0.454638



y2016	word	dist(camera)
1	5mp	0.730
2	13mp	0.711
3	flash	0.692
4	cameras	0.670
5	8mp	0.649
6	sensors	0.644
7	lense	0.643
8	autofocus	0.640
9	mic	0.638
10	2mp	0.627
11	rear	0.619
12	microphone	0.618
13	speaker	0.616
14	speakers	0.610
15	proximity	0.606
16	hump	0.594
17	sensor	0.584
18	slr	0.580
19	fisheye	0.578
20	facing	0.573
21	display	0.573
22	lens	0.572
23	cutout	0.568
24	mirrorless	0.564
25	zoom	0.548
26	20mp	0.547
27	unobstructed	0.539
28	front	0.537
29	telephoto	0.536
30	grill	0.535

그림 5: 전처리 과정

#### 4.1.2 데이터 활용 - CA 상대적 중요도 도출

QFD의 CA항목으로는 대표적으로 스마트폰 카메라의 스펙으로 분류되는 항목들인 Resolution(Pixel), Lens, Features(추가기능), Auto focusing, Body(외관)으로 나누었다. Features는 카메라의 기본적인 성능보다는

부가적으로 추가되는 기능들로 정의하여, 근접센서, 손떨림방지, 플래쉬, 가속도센서, 적외선센서 등에 대한 것으로 분류하였다. 전처리 완료 후, word2vec 학습 모델을 통해 <표4>과 같이, 스마트폰 고객 리뷰 중에 카메라와 관련성이 높은, 코사인 유사도 기준 상위100개 단어를 추출하였다.

rank	word	similarity						
1	5mp	0.730	36	stabilization	0.519	71	elevated	0.450
2	13mp	0.711	37	protruding	0.509	72	focuses	0.450
3	flash	0.692	38	laser	0.504	73	optical	0.449
4	cameras	0.670	39	speakerphone	0.504	74	vented	0.448
5	8mp	0.649	40	ois	0.501	75	dolby	0.447
6	sensors	0.644	41	1080p	0.496	76	vision	0.446
7	lense	0.643	42	aperture	0.495	77	pinhole	0.446
8	autofocus	0.640	43	ambient	0.495	78	lip	0.444
9	mic	0.638	44	macro	0.493	79	audio	0.444
10	2mp	0.627	45	mics	0.492	80	480p	0.444
11	rear	0.619	46	vga	0.489	81	botton	0.441
12	microphone	0.618	47	16mp	0.488	82	obstruct	0.441
13	speaker	0.616	48	bezel	0.488	83	viewable	0.440
14	speakers	0.610	49	scenic	0.486	84	firing	0.439
15	proximity	0.606	50	mp	0.485	85	canon	0.438
16	hump	0.594	51	nikon	0.485	86	sehoo	0.437
17	sensor	0.584	52	telescope	0.485	87	blockage	0.437
18	slr	0.580	53	accelerometer	0.484	88	qhd	0.435
19	fisheyc	0.578	54	backlight	0.481	89	len	0.435
20	facing	0.573	55	hdr	0.479	90	scanner	0.434
21	display	0.573	56	ir	0.479	91	ips	0.433
22	lens	0.572	57	stereo	0.478	92	controls	0.433
23	cutout	0.568	58	digital	0.478	93	capturing	0.431
24	mirrorless	0.564	59	recess	0.477	94	presets	0.429
25	zoom	0.548	60	grills	0.477	95	radio	0.427
26	20mp	0.547	61	fl	0.465	96	grille	0.427
27	unobstructed	0.539	62	focus	0.465	97	oval	0.426
28	front	0.537	63	blaster	0.462	98	infrared	0.425
29	telephoto	0.536	64	obscures	0.461	99	spen	0.424
30	grill	0.535	65	helmet	0.459	100	Asymmetrical	0.421
31	crown	0.532	66	microphones	0.459			
32	21mp	0.531	67	resolution	0.458			
33	dslr	0.531	68	blur	0.456			
34	megapixels	0.526	69	camers	0.453			
35	megapixel	0.524	70	captures	0.450			

표 4: word2vec 결과

<그림6>은 추출된 100개의 단어들을 CA항목(Resolution, Lens, Features, AF, Body) 속성에 해당하는 단어들로 각각 분류하고, 분류된 단어들 중 상위 5개 단어씩을 선택 후 평균값과 그에 대한 백분율 값을 산출한 과정을 나타내었다. resolution, lens, body, features, AF의 순으로 코사인 유사도 평균값과 백분율의 word2vec 학습결과가 나왔으며, 고객 리뷰 빅테

이터 분석을 통해 얻은 해당 백분을 값을 HOQ level 1의 소비자 요구사항 상대적 중요도 수치로 사용하였다.

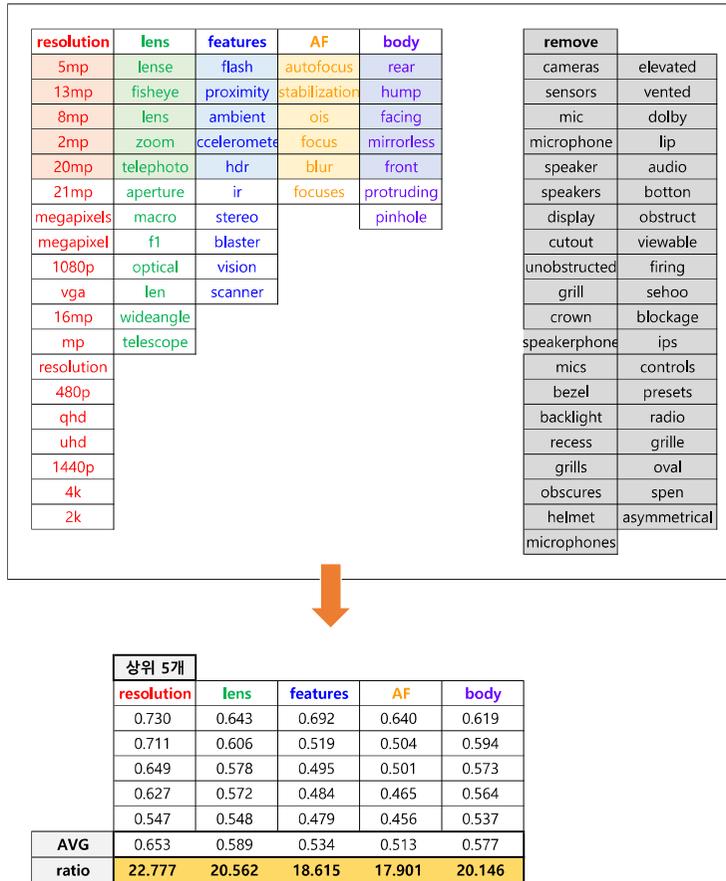


그림 6: word2vec으로 부터 CA상대적 중요도 도출

## 4.2 2단계 실험결과

### 4.2.1 HOQ level 1

HOQ level 1의 EC 항목은 DXOMark의 카메라 화질 평가 항목으로 하였다. 고객의 언어와 화질 전문가들의 평가 항목을 연계시켜, HOQ level 2의 기술 속성과 이어지게 하였다. 3.1절에서 언급하였듯, 기술속성항목은 Exposure and contrast, Color, Autofocus, Texture, Noise, Night scene, Bokeh, Wide angle로 구성하였고, CA(Customer Attribute)와 EC (Engineering Characteristic)과의 관계행렬(Relationship Matrix)은 미국 특허데이터 텍스트 마이닝 분석을 활용하여 작성하였다. 2016년 ~ 2018년의 미국 특허 데이터에 <표5>와 같이 각각의 EC항목과 연관있는 CPC code를 검토하여 데이터 수집에 반영하였다. 예를 들어, Exposure and contrast항목

Dxomark 항목	CPC code	Description
Exposure	H04N5/2353	Circuitry or methods for compensating for variation in the brightness of the object, e.g. based on electric image signals provided by an electronic image sensor by influencing the exposure time, e.g. shutter
Color	H04N9/735	Circuits for processing colour signals colour balance circuits, e.g. white balance circuits, colour temperature control for picture signal generators
Autofocus	H04N5/23212	Focusing based on image signals provided by the electronic image sensor
	G03B13/36	Autofocus systems
Texture	H04N5/357	Noise processing, e.g. detecting, correcting, reducing or removing noise
Noise	H04N5/357	Noise processing, e.g. detecting, correcting, reducing or removing noise
Zoom	H04N5/23296	Control of means for changing angle of the field of view, e.g. optical zoom objective, electronic zooming or combined use of optical and electronic zooming
	G02B13/009	Miniaturised objectives for electronic devices, e.g. portable telephones, webcams, PDAs, small digital cameras having zoom function.
Bokeh	G06T5/20	Image enhancement or restoration by the use of local operators
	H04N5/23229	Devices for controlling television cameras, e.g. remote control ; Control of cameras comprising an electronic image sensor comprising further processing of the captured image without influencing the image pickup process
Wide angle	H04N5/23296	Control of means for changing angle of the field of view, e.g. optical zoom objective, electronic zooming or combined use of optical and electronic zooming
	H04N5/23238	Control of image capture or reproduction to achieve a very large field of view, e.g. panorama
Night	H04N5/357	Noise processing, e.g. detecting, correcting, reducing or removing noise

표 5: EC항목과 연계한 CPC code table

의 경우, CPC code : H04N5/2353 'Circuitry or methods for compensating for variation in the brightness of the object, e.g. based on electric image signals provided by an electronic image sensor by influencing the exposure time, e.g. shutter' 로 '피사체의 밝기의 변화를 보상해 주는 회로소자 또는 방법'이라는 정의를 가진 CPC code를 특허 데이터 검색조건에 반영하여 Exposure and contrast항목과 관계성이 높은 데이터를 수집하였다.

수집된 특허 데이터들의 abstract 부분을 추출하여 숫자제거, 특수문자 제거, 불용문자 제거, 어근 동일화, 공란처리, 대소문자 통일 등의 텍스트 데이터 전처리를 진행하여 word count를 수행하였다. <그림7> 단어빈도테이블(Word frequency table)과 같이 1 ~ 100위까지의 빈도순위로 정리하였고, BOW(Bag of Words)가정에 근거하여 CPC code 별로 분류된 특허 데이터에서 각 EC항목과 관련된 단어가 많이 언급될 수록 해당 속성과의 관계성이 높다고 가정하고 1위에서 30위는 ◎(9점), 31위에서 60위는 ○(3점), 61위에서 100위는 △(1점), 100위 이하는 (0점)으로 테이블에 차등배점을 적용하였다.

	Exposure	Color	AF	Noise	Zoom	Bokeh	Wide	
1~30	imag captur exposur	imag color process	imag focus control	imag pixel signal	imag camer control	imag camer method	imag camer system	
	3919 879 873	2167 557 537	3420 1081 930	2359 1148 900	3263 1475 1101	197 57 54	3096 1463 993	
	includ control method	metho devic includ	unit includ devic	includ devic unit	devic includ unit	process depth data	includ devic captur	
	847 836 808	472 462 399	905 878 792	772 635 613	974 957 937	49 46 45	877 846 845	
	devic unit apparatus	captu data light	method len camera	appar process method	system method len	devic genera digit	method unit len	
	779 767 694	365 363 306	787 680 679	511 497 489	925 886 842	39 37 34	775 658 638	
	camer light sensor	camer unit sensor	apparati detect captur	plura sensor system	captu appar display	base user interfac	video control display	
	659 572 526	275 259 257	675 628 619	419 390 328	823 770 623	31 29 27	608 606 535	
	proce pixel configur	appar white system	pixel sensor signal	output circuit generat	zoom optic configur	appar exampl includ	appar data optic	
	508 503 484	253 251 239	551 529 486	324 321 306	572 544 537	25 25 25	519 461 460	
	time data system	balan pixel display	system optic configur	conve config light	infor object video	system display electron	confi view process	
	480 475 472	222 208 206	476 451 448	298 287 273	482 477 469	25 23 23	457 448 423	
	set plural detect	base config control	process posit base	photo captur solidst	oper posit process	optic blur light	sensu object plural	
	397 379 363	204 199 194	446 425 366	270 268 258	461 441 409	23 22 21	418 394 348	
	displ base signal	signal genera correct	plural data perform	data charg control	detec base sensor	zoom captur map	posit infor provid	
	355 323 323	182 169 167	359 331 328	230 226 223	377 376 370	21 20 20	338 318 315	
	gener determin perform	deter plural infor	displ light determin	array element	data plural	param pixel	general zoom	
	311 296 261	158 152 147	326 318 305	223 204 203	357 324 319	19 19 19	309 299 298	
	infor oper	perfo receiv adjust	object general infor	vo'tra region convert	view provid determin	photoporior output	base set determin	
	256 256 248	145 137 136	279 270 263	201 198 194	318 315 302	18 18 17	278 266 262	
	31~60	obtai frame object	detec electro accelar	set oper modul	trans camer correct	set user receiv	visual appll frame	oper panora direct
		239 238 227	119 117 116	258 248 238	192 191 190	293 266 265	17 16 16	253 250 239
		receiv len	obtai set valu	receiv region area	elect base	gener light perform	acquir compr receiv	receiv vehid field
		222 205 204	109 107 107	224 211 209	180 174 174	263 235 235	15 15 15	235 234 232
		mode region output	input array	elemen distanc	optic provid portion	signal direct modul	deter point relat	modul proces detect
		203 196 193	107 106 102	208 202 197	170 170 169	225 224 221	14 14 14	230 225 224
		provi adjust shutter	object output sourc	auto focus phase	not's transfer	area focus processor	result transit view find	light area user
		191 190 183	102 102 102	193 190 189	166 159 158	215 208 198	14 14 14	217 202 195
		ptcku optic video	compu provid engin	subject electron	perfo pickup digit	elect acquir move	effect proces adjust	compr electro angl
		182 181 179	99 96 90	186 180 179	158 156 145	191 182 180	13 13 12	193 181 179
		modul calcul period	illum acquir paramet	ditect obtain move	infor detect	track target	area config focus	contel portior relat
		175 172 163	89 88 87	173 173 171	144 143 142	178 178 177	12 12 12	159 159 157
		subje area bright	digit gain	provid calcul	len read determin	digit field	plural pressen background	obtai frame perform
162 160 156		83 82 79	168 167 167	142 136 135	173 172 170	12 12 11	156 155 147	
digit embodi		proce frame oper	adjust drive pickup	compr's region	compr's region	control field human	regio embodi	
155 148 146		78 75 75	163 160 159	131 126 125	169 166 163	11 11 11	143 138 136	
proce focus acquir		scene temp embodi	compr acquir	column line section	obtai vehid relat	provid restor tile	move point acquir	
146 141 140		73 73 72	156 154 151	123 121 121	162 157 154	11 11 11	133 133 132	
motio illumin electron		calcu exposi pre process	differ process output	compr receiv frame	elemen chang	input	input compu group	
138 137 136		71 70 70	149 146 139	118 115 113	151 150 150	11 10 10	131 129 129	

그림 7: 미국 특허 단어 빈도 테이블(word frequency table)

예를 들어 Exposure 의 경우, 1위에서 30위 area에 'pixel'이 랭크되어 있어 Resolution(pixel)에 ◎(9점)으로 배점하였고, 'lens, optic, module, focus' 등이 31위에서 60위 area에 랭크되어 있어 Lens(Lens, Optic), Body(Module), AF(focus) 항목에 ○(3점)으로 배점 하여, Exposure의 CA-EC 관계행렬을 작성하였고, 다른 기술속성항목들도 같은 기준으로 HOQ level 1을 작성하여 <그림8>와 같이 HOQ를 정리하였다. CA 상대적 중요도 수치와 관계행렬의 값들을 연산하여 중요도 수치를 산출한 결과, Autofocus, Bokeh, Exposure and contrast, Noise, Texture, Night scene, Color, Wide angle 의 순으로 확인되었고, 백분율값으로 환산하여 HOQ level 2에서 이용할 EC항목별 상대적 중요도 수치를 구할 수 있었다. 최근 몇년간 DSLR 카메라 AF기술인 위상차 검출 AF(PDAF, Phase Detection Auto Focus)나 Dual pixel AF 등이 스마트폰 카메라에 적용되면서, 보다 빠르고 정확한 Auto Focus 기능이 중요시 되었고, 라이브포커스(Live Focus) 또는 인물사진(Portrait)모드라 불리는 아웃 포커싱(out of focus) 기능이 소비자에게 어필되고 있었음을 볼 때, 미국 특허 빅데이터 분석을 이용하여 기술속성(EC)와 소비자 요구사항의 관계행렬값에 반영한 Auto-focusing, Bokeh 순으로 이어지는 결과는 실제와도 연결 지을 수 있다고 판단할 수 있었다.

## 4.2.2 HOQ level 2

HOQ level 2에서는 각 기술의 중요도, 기술요소간의 관계행렬 작성과 함께, 추가적으로 미래 소비자 선호도 변화에 대한 연구[21]을 참고하여, 미래 중요도 변화 관련 항목을 더하여 전문가 설문을 진행하였다. 그리고, HOQ level 1에서 도출된 각 EC항목의 상대적 중요도 수치를 반영하였다. 전문가 설문은 CIS(CMOS Image Sensor) 신제품 평가 엔지니어 6명을

Row #	Relative Weight	Weight / Importance	D demanded Quality (a.k.a. "Customer Requirements" or "Whats")	Column #								
				1	2	3	4	5	6	7	8	9
			Quality Characteristics (a.k.a. "Technical Requirements" or "Hows")	Exposure and Contrast	Color	AF	Texture	Noise	Zoom	Bokeh	Wide	Night
1	20.6	20.6	Lens	○	▲	○	○	○	○	○	○	○
2	18.9	18.9	AF	○		○	○	○	○	○		○
3	22.8	22.8	Resolution(pixel)	○	○	○	○	○		○	▲	○
4	17.5	17.5	Features		○							
5	20.2	20.2	Body	○		○					○	
<b>Weight / Importance</b>				384.3	278.3	621.3	323.7	323.7	242.1	447.3	268.8	323.7
<b>Relative Weight</b>				12.0	8.7	19.3	10.1	10.1	7.5	13.9	8.4	10.1

그림 8: HOQ level 1

선정하여 진행하였으며, 각 설문 수치들을 합산 후 평균값을 이용하였다. 기술의 중요도는 5점 척도법(five point scale)을 활용하여 전문가들의 판단에 따라 중요도 수치를 측정하였고, 기술속성간 관계행렬은 속성간의 연관도가 높을 경우 ◎(9점)으로 하여, 연관 기술 관련 강도에 따라 ◎(9점), ○(3점), △(1점), 공란(0점)으로 배점하도록 하였다. 미래 중요도 변화 역시 전문가의 직접 판단에 따라 3년후 각각의 해당 기술의 중요도가 현재대비 더 증가하거나 감소할지에 대해 -100 ~ 100% 수치로 설문을 진행하였다.

Stage 1의 Importance Score는,

' $\sum(CA(i) \text{의 Relative Weight} \times CA(i) \text{와 각 EC와의 관계행렬값}) \times \text{전문가 점수}$ '

로 산출된다. Ultra High Resolution 항목의 경우,  $(12 \times 3 + 19.3 \times 0.5 + 10.1 \times 9 + 10.1 \times 6 + 7.5 \times 6 + 13.9 \times 0.7 + 8.4 \times 1.7 + 10.1 \times 7) \times 4.1 = 1370.4$  로 Stage1의 Importance Score 값을 얻을 수 있다. 또한 Stage2는 원래의 상대적 중요도 값에 3년후 중요도 변화 % 수치를 반영한 값이다. Ultra high resolution 기술의 3년 후의 변화된 상대적 중요도 산출시, Exposure and Contrast의 상대적 중요도 변화는 '12'  $\times (1-0.06) = 11.2$ 을 시작으로 각각의 Technical Requirements에 중요도 변화를 반영시켜 3년후의 변화한 상대적 중요도 수치를 구한 후 진행한다. <그림9>는 전문가 설문 후, 각각의 수치를 합산 후, 평균값을 구하여 최종적으로 작성된 HOQ 테이블이며, 전문가 점수와 함께 각 기술 속성의 중요도와 백분율 값, 중요도 변화수치, 3년 후의 중요도 수치 변화까지 확인하였다. 하기 Stage 1, 현재시점의 상대적 중요도와 Stage 2, 3년후 상대적 중요도를 통해 최종 우선 순위 결과를 도출 할 수 있었다.

Row #	Max Relationship Value in Row	Relative Weight	Weight / Importance	Column #	1	2	3	4	5	6	7	8	중요도 비율 (%)	
1	9	12.0	384.3	<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <b>Technical Requirements</b>            Exposure and Contrast            Color            AF            Texture            Noise            Zoom            Bokeh            Wide            Night         </div> <div style="width: 45%;"> <b>Parameters</b>            Ultra high resolution(ex.108Mp)            Slim size(ex. 0.7um pixel)            Nona cell(big pixel for low light)            Dual Pixel(ultra fast AF)            Real time HDR(Smart WDR)            Smart ISO(optimal dynamic range with less noise)            FRS(ex. 4K 60fps, FHD 480fps)            3D sensor(TOF)         </div> </div>	3.0	6.0	9.0	1.7	9.0	9.0	9.0	5.5		3년 후(2023년)
2	2	8.7	278.3		2.2	0.5					0.2			4
3	9	19.3	621.3		0.5	4.0	2.0	9.0	0.2				6.0	33
4	9	10.1	323.7		9.0	5.5	2.8	0.2	2.5				1.7	76
5	5	10.1	323.7		6.0	5.2	4.0	1.0	1.0	5.0			1.2	80
6	6	7.5	242.1		6.0	1.2	1.5						2.2	-34
7	1	13.9	447.3		0.7	0.7							9.0	56
8	2	8.4	268.8		1.7	0.7	0.5							-25
9	9	10.1	323.7		7.0	4.5	9.0	0.5	4.0	3.5	0.5			78
				<b>전문가 점수</b>										
Stage 1				Importance Score	1370.4	1377.2	1112.6	790.3	714.6	762.6	282.1	701.1		
				%	19.3	19.4	15.6	11.1	10.0	10.7	4.0	9.9		
Stage 2				Importance Score	2022.3	1944.8	1543.9	1050.9	918.7	1003.1	314.2	1010.2		
				%	20.6	19.8	15.7	10.7	9.4	10.2	3.2	10.3		

그림 9: HOQ level 2

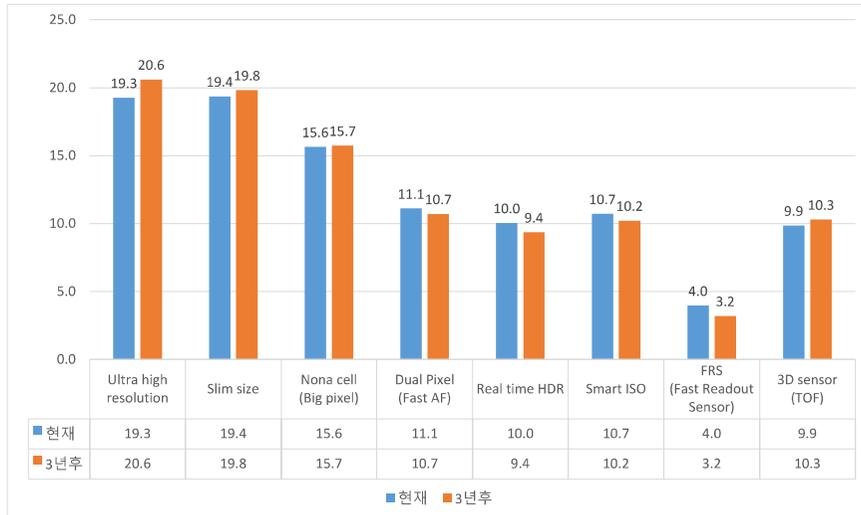


그림 10: CIS 특성 기술 우선순위와 그 변화

Stage 1, 현재시점 기준 EC(Engineering Characteristic) 항목들 간의 우선순위를 <그림10>을 통해 살펴보면, Slim size, Ultra high resolution, Nona cell(Big pixel), Dual pixel(Fast AF), Smart ISO(Optimal dynamic range), Real time HDR, 3D sensor(TOF, Time of flight), ISOCELL, FRS(Fast Readout Sensor) 순으로 기술 특성 우선순위를 확인할 수 있다. Slim size 항목이 Ultra high resolution 항목 대비 미세하게 중요도 수치가 높으나, 3년 후의 중요도 변화 수치를 반영한 결과, Ultra high resolution 항목의 우선 순위가 Slim size를 앞서게 되고, 3D sensor(TOF) 항목이 7위에서 5위로 변함을 <표6>와 같이 확인할 수 있었다.

최근들어 5천만 화소, 1억화소 등의 이미지 센서가 출시되어 소비자들에게 큰 관심을 끌었고, 앞으로도 더욱더 Ultra high resolution(초고화소) 기

Priority	현재	3년후
1	Slim size	Ultra high resolution
2	Ultra high resolution	Slim size
3	Nona cell	Nona cell
4	Dual pixel(fast AF)	Dual pixel(fast AF)
5	Smart ISO	3D sensor(TOF)
6	Real time HDR	Smart ISO
7	3D sensor(TOF)	Real time HDR
8	FRS	FRS

표 6: CIS 특성 기술 우선순위와 그 변화

술에 대한 소비자 니즈는 꾸준히 늘어날 것으로 예상되는 바, 그 중요성은 꾸준히 증가할 전망이다. 그리고 특히, 3D sensor(TOF)의 경우, 3D sensor는 Time Of Flight(TOF) 기술로 대변된다. TOF 센서는 레이저나 LED 빛이 물체에 쏘면 뒤 반사되어 오는 왕복 시간을 계산해 센서와 물체와의 거리를 측정하는 센서이다. 이를 이용하여 물체의 이미지를 3D로 구현할 수 있어 이용 범위가 점점 늘어나고 있다. 최근 몇년 동안 개선되어오면서, 점점 그 활용도가 높아지고 있어, 소비자들도 해당 기술에 대한 관심이 늘어나고 있는 것으로 보인다. 향후 3년 간 기술이 꾸준히 발전하면서, 기술적인 성숙도가 높아지면서 생체인증, 동작인식은 물론 AR(Augmented Reality), VR(Virtual Reality), 3D 스캐닝 등을 포함하여 여러 방면으로 구현하며 점차 그 쓰임이 늘어날 전망이다. 한편 3D sensor 대비, Dual pixel, Real time HDR, Smart ISO, FRS등은 중요도가 조금씩 하락하면서 우선순위가 유지되거나 하락하게 된다. 이는 어느 정도 기술적 한계 또는 경쟁사 기술 대비 동등하거나 그 이상의 기술 수준으로 완성도를 가지면서, 3D sensor에 중요도를 더 집중시킬 것으로 판단된다.

## 제 5 장

### 결론

#### 5.1 연구 요약 및 시사점

CMOS image sensor(CIS)는 렌즈를 통해 들어본 빛 정보를 디지털 신호로 변환해 주는 반도체이다. 필름 카메라에서의 필름 역할을 대신하여 모든 디지털 카메라에서 쓰이고 있으며, 스마트폰 카메라, 자율 주행 자동차, 산업용 로봇 등 각종 IT기기 및 산업분야에서 카메라로서의 이용뿐만 아니라 인공지능의 눈역할을 하게 되면서 앞으로도 꾸준히 그 쓰임이 늘어날 전망이다. 본 연구에서는 미국의 대표적인 온라인몰인 아마존의 스마트폰 고객 리뷰 데이터(110만여건)를 수집하여, 최종 소비자에 대한 빅데이터 분석을 진행하였고, 이를 통해 QFD의 핵심도구인 HOQ의 CA(Customer Attribute) 상대적 중요도로 활용하였다. QFD는 고객의 요구사항을 제품 기술 개발이나 서비스 개발의 각단계에 반영할 수 있게 하는 방법론이다. 기존에 VOC를 수집하기 위해서는 주로 소비자 설문 조사를 기본으로한, Conjoint analysis, AHP기법, Fuzzy이론 등을 통해서 소비자 요구사항에 대한 분석을 진행하여, 많은 인력과 시간, 비용을 필요로 하였다. 본 연구에서는 설문조사에 대한 과정없이, 다량의 소비자 리뷰 데이터에 대해 자연어 처리 모델 중 기계학습 기반인 word2vec을 이용하여 소비자 니즈를 정량화하였고, 이를 상대적 중요도로 활용하였다. EC(Engineering Characteristic)과의 연관관계 작성 단계에서도 미국 특허 데이터(4500여건)를 이용한 텍스트 마이닝 분석을 통해 CA와 EC간의 관

계매트릭스를 작성하였다. HOQ level 2에서는 HOQ level 1에서의 EC항목 중요도 수치를 바탕으로 전문가 설문을 통해 기술 로드맵 수립을 위한 CIS(CMOS image sensor)기술요소들의 우선순위와 시간에 따른 중요도 변화를 최종적으로 도출하였다.

본 연구는 텍스트 마이닝, Word2vec, QFD까지의 총 3가지 방법론을 활용하였다. 먼저, 수집한 아마존 고객 리뷰 데이터 중 Cellphones and accessories 카테고리의 2016년 데이터를 추출 후, 영문 텍스트 데이터 전처리 작업을 수행하였다. 이후, Word2vec 학습을 통해 수집된 데이터를 분석하였다. Word2vec은 타겟단어로 주변단어를 도출하는 skip-gram 방식과 주변단어를 이용하여 타겟단어를 도출하는 CBOW방식이 있다. 본 연구는 skip-gram 방식을 이용하였고 타겟단어를 'camera'로 하여, cellphones 카테고리에 있는 단어들 중 'camera'와 가까운 벡터 공간에 위치하는 단어들 100개를 코사인 유사도 기준으로 추출하였다. 추출된 100개의 단어는 스마트폰 카메라의 대표적인 스펙 사양인 Lens, Resolution(Pixel), Auto focusing, Features(부가기능), Body(외관)의 5가지 항목으로 분류하였고, 분류된 단어들 중 상위5개의 코사인 유사도 수치에 대해 평균값을 구하여 정량화 함으로써, 해당 EC항목의 상대적 중요도로 활용할 수 있었다. 그 결과, 미국 소비자 분석을 통해 파악된 CA항목의 중요도 순위는 Resolution(22.8), Lens(20.8), Body(20.2), AF(18.9), Features(17.5)로 확인할 수 있었다. 전통적으로 몇메가 화소의, 어떤 렌즈의 카메라냐 라는 것이 가장 먼저 와닿게 되는 카메라 사양임을 알 수 있을 때, 해당 상대적 중요도 순위를 이해할 수 있었고, 해당 중요도 우선순위에 대해 상대적 중요도를 빅데이터 분석을 통해 수치화했다는 것이 의미있었다. EC항목들은 DXO Mark社의 화질 평가 항목으로 선정하였고, 해당 EC와 CA와의 관계 매트릭스는, 각각의 EC와 관련된 미국 특허들을 CPC(Cooperative Patent Clas-

sification) code 분류를 통해 수집 후, 수집된 특허 데이터에 대해 텍스트 마이닝 분석을 수행하였다. 그 후, 각각의 CA항목들과 관련있는 단어들의 빈도테이블(Word frequency table) 확인을 통해 CA와 EC간의 연관 관계 강도를 측정하였고, Importance 수치를 산출하여 HOQ level 2에 이용하였다.

HOQ level 2에서는 전문가 설문을 통해 기술적 요구 사항과 EC - EC 간의 관계 행렬, 이미지 센서 특성 기술들의 우선순위, 3년후, 5년후의 기술 중요도 변화 등을 확인하였다. 이를 위해 HOQ level 1에서 산출된 weight를 기술적 요구사항에 적용하고, 특성 기술들간의 관계행렬과 전문가 점수를 통해 나타난 기술 우선순위를 확인하였고, 함께 3년후, 5년후에 대한 중요도 변화 수치를 반영해 미래 중요도 변화를 예상하였다. 그 결과 현재는 0.7um 픽셀 크기에 따른 슬림사이즈, 1억화소의 초고해상도, 저조도 밝기를 확보하기 위한 나노셀, 더 빠르고 정확한 오토포커스를 위한 듀얼픽셀 등의 순으로 9가지 기술에 대한 우선순위가 나왔으며, 그 중 슬림사이즈 기술이 초고해상도 기술보다 미세하게 우세하였다. 하지만, 3년후 시간이 갈수록 초고해상도와 3D sensor에 대한 중요도가 점점 커지면서 초고해상도 항목과 3D sensor 항목은 우선순위가 각각 2위에서 1위, 7위에서 5위로 상승하며 점점 중요도가 커지고 있음을 알 수 있었다. 이를 통해 1억화소, 또는 그 이상의 해상도 기술과 활용 범위가 점점 늘어나고 있는 3D sensor에 대한 소비자 상대적 중요도 비중이 커짐을 알 수 있었다.

## 5.2 연구 성과 및 기대 효과

실험 결과 HOQ level 1 CA 항목 중 Resolution이나 level 2의 EC 항목 중 Ultra high resolution 과 같이 Resolution(해상도)와 관계된 카메라와

센서 특성들의 소비자 중요도가 높게 나타나고 있다. 다른 이미지 화질 특성들 대비 소비자들에게 직관적으로 어필이 되는 스펙이기 때문이라고 판단된다. 해상도의 중요성이 높다 하더라도 무조건적으로 해상도를 높이기에는 기술적 어려움이 많다. 제한된 센서 크기에서 해상도가 증가하면, 그에 따라 고려해야 하는 트레이드 오프(trade off) 관계의 센서 특성들이 있기 때문이다. 화소수가 늘어나면서 더욱 더 디테일한 이미지 표현이 가능하겠지만, 같은 크기의 센서에 더욱 더 많은 픽셀들이 자리함에 따라 픽셀이 점점 미세화 되면서 받을 수 있는 빛의 양이 줄어들게 되고, 그에 따라 SNR(Signal to Noise Ratio) 측면에서도 불리한 점이 많아 진다. 화소수가 늘어남에 따라 나타나는 단점들을 보완하는 대안 기술들도 개발되고 있고, 미세공정 기술력이 발전하면서 적어진 빛 신호에 대한 손실을 최소화 하면서, 적은 노이즈로 디지털 신호로 변환하게 하거나, ISP(Image Signal Processing) 측면에서도 많은 부분을 개선하면서 보완할 수도 있다. 그리하여 화소수를 늘리기 보다 센서 성능을 더욱 개발하면서 이미지 화질을 개선해온 스마트폰 제조사들도 있는 반면에, 보완점들을 적용하여 해상도를 증가시키면서 디테일한 표현을 더욱 더 중시한 제품을 출시하는 제조사들도 있다. 본 연구 결과를 통해 신제품 개발시, 여러가지 센서 특성에 대한 트레이드 오프가 발생할 경우 고객지향적인 관점에서는 어떤 것을 우선해야 하는가에 대한 마켓 센싱 자료로 활용할 수 있을 것이라 판단된다.

대표적인 B2B 산업 중 하나인 반도체 산업에 대해, 최종 소비자들의 제품 리뷰에 대한 빅데이터 분석을 바탕으로 하여, 제품 개발 과정에서 활용할 수 있도록, 실제 스마트폰 카메라 사용자의 VOC를 QFD모형에 반영해 보았다. 전통적으로 B2B 산업의 특징은 B2C와는 달리 B2B제품의 수요는 구매 고객으로부터의 직접 수요라기 보다는 고객의 고객을 거쳐

간접적으로 발생하는 파생수요(derived demand)의 성격을 가지고 있다. 따라서 제품의 품질과 납기, 가격을 바탕으로 기업 고객과의 장기적인 거래 관계가 매우 중요한 요소가 될 뿐만 아니라 마켓 센싱(Market sensing)과 수요 예측의 중요성이 강조된다[22]. 본 연구를 통해 미국 시장 스마트폰 소비자의 요구사항을 기술 개발에 직접적으로 연계시켜, 특성 기술들의 우선순위를 도출해 보았다. 이를 통해 최종 소비자 니즈가 반영된 CIS 신기술 우선순위를 마켓 센싱 자료로써 기술 로드맵 수립에 활용 가능할 것으로 판단된다. 뿐만 아니라 본 연구의 모형은 CMOS image sensor 반도체 분야 외에도 빅데이터 분석을 통해 도출된 최종 소비자 니즈가 반영된 기술 요소 우선 순위를 활용하고자 하는 타 산업 기술 분야에서도 적용 가능 할 것으로 기대된다.

한 편 본 연구는 몇가지 한계점이 있다. 2016년 아마존 고객 데이터를 2016 ~ 2018년 특허데이터와 이미지 센서 최신 기술에 연계하였다. 매년 급속도로 기술 발전이 일어나는 IT기술에 대해 최신의 기술에 대한 소비자 니즈라 보기는 어렵다. 2016년 이후의 최신 제품에 대한 고객 리뷰 데이터까지 포함하여 분석을 진행하였다면, 조금 더 최근의 소비자 니즈를 반영한 기술 선호 우선순위가 될 수 있었을 것이다. 또한, HOQ level 2의 전문가 설문에서 제한된 범위 안에서 전문가를 선정하여, 다양한 시각의 전문가의 의견을 반영하지 못했다는 한계를 가진다. 우리나라 특정기업에 국한된 전문가로 제한되어 미국이나 중국 같은 거대 소비시장의 소비자 선호 변화를 정확히 반영하지 못할 수도 있다. 다양한 전문가층을 대상으로 설문을 진행하거나, 또는 HOQ level 1과 같이 HOQ level 2에서도 텍스트 마이닝 기술 분석이나 머신 러닝을 통해 특허 데이터를 활용하여 전문가 설문을 대신하였다면, 빅데이터를 통한 좀 더 객관화된 연구가 될 수 있었을 거로 예상된다.

## 참고 문헌

- [1] 백영민, “R을 이용한 텍스트 마이닝 = text-mining using r / 백영민 지음,” 2017.
- [2] M. Gentzkow, B. Kelly, and M. Taddy, “Text as data,” *Journal of Economic Literature*, vol. 57, no. 3, pp. 535–74, 2019.
- [3] G. Miner, J. Elder IV, A. Fast, T. Hill, R. Nisbet, and D. Delen, *Practical text mining and statistical analysis for non-structured text data applications*. Academic Press, 2012.
- [4] 이기창, *한국어 임베딩*. 서울: 에이콘출판사, 2019.
- [5] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” *arXiv preprint arXiv:1301.3781*, 2013.
- [6] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” pp. 3111–3119, 2013.
- [7] 이상복, *QFD(품질기능전개) 이론과 사례*. 군포: 이레테크, 2007.
- [8] M. S. Yadav and P. A. Pavlou, “Marketing in computer-mediated environments: Research synthesis and new directions,” *Journal of Marketing*, 2014.
- [9] 김인규, “텍스트마이닝과 머신러닝을 이용한 모바일앱 구매반응 분석,” vol. , pp. 116–124, 한국경영정보학회, 2016.
- [10] N. Archak, A. Ghose, and P. G. Ipeirotis, “Deriving the pricing power of product features by mining consumer reviews,” *Management science*, vol. 57, no. 8, pp. 1485–1509, 2011.

- [11] 심영석 and 김홍범, “온라인 리뷰 빅데이터 기반의 word2vec 기법을 활용한 관광지 개성과 여행객 평점 간 구조적 관계 분석,” *관광학연구*, vol. 42, no. 8, pp. 165–189, 2018.
- [12] 안병대, 최진욱, 서용무, *et al.*, “기업개성이 직원의 직무만족과 기업 이직률의 관계에 미치는 영향,” *한국 IT 서비스학회지*, vol. 19, pp. 35–56, 2020.
- [13] 채인영, 김지영, and 유기윤, “트위터 데이터와 구글 리뷰를 이용한 장소 선호도 분석 방법,” *대한공간정보학회 학술대회*, pp. 29–30, 2017.
- [14] 남미우, “빅데이터 분석 방법을 적용한 인스타그램 뷰티관련 해시태그 게시글 분석에 관한 융합 연구,” *한국과학예술융합학회*, vol. 37, no. 4, pp. 93–106, 2019.
- [15] 조익천, 전경채, 허우, and 곽춘중, “Qfd 와 fuzzy 이론을 적용한 dslr 카메라 신제품 개발,” *한국생산관리학회지*, vol. 23, no. 1, pp. 43–69, 2012.
- [16] 황반 and 곽춘중, “Kano 모델을 기반으로 한 qfd 효율화: Dslr 카메라 사례,” *한국생산관리학회지*, vol. 30, no. 4, pp. 447–465, 2019.
- [17] 이슬기, 김아영, 홍상필, 이승제, and 이민아, “품질기능전개 (quality function deployment) 방법을 적용한 고추장 소스 콘셉트 개발: 미국과 중국 시장을 중심으로,” *한국식품영양과학회지*, vol. 44, no. 9, pp. 1388–1398, 2015.
- [18] 김현정 and 김수욱, “Ahp 와 qfd 를 이용한 철도기술 개선에 관한 실증적 연구,” *품질경영학회지*, vol. 41, no. 2, pp. 301–321, 2013.
- [19] J. Ni, J. Li, and J. McAuley, “Justifying recommendations using distantly-labeled reviews and fine-grained aspects,” in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 188–197, 2019.

- [20] 김방룡, “특허 정보를 활용한 IT 유망기술 도출에 관한 연구,” *한국통신학회논문지*, vol. 34, pp. 1021–1030, 2009.
- [21] 이철용 and 이종수, “기술 로드맵의 기초 작성을 위한 컨조인트 분석과 qfd,” *기술혁신연구*, vol. 15, no. 1, pp. 65–86, 2007.
- [22] 한상린, “국내 B2B마케팅의 연구 동향과 향후 연구방향,” *마케팅연구*, vol. 29, pp. 93–119, 2014.

# Abstract

## **A Study on Technology Roadmap for Smartphone CMOS Image Sensor: Based on US market on-line consumer review data**

Sungchul CHOI

Graduate School of Practical Engineering

Seoul National University

CMOS Image Sensor(CIS) market is continuously growing for years. The CIS is the eye of Smartphone camera, Autonomous vehicle, Industrial robot,,etc. So, Korean semiconductor companies which make the CMOS image sensor are trying to transform to 'first mover' from 'fast follower'. For that, through the QFD(Quality Function Deployment) model using consumer big data analysis, this study intends to establish a technical road map reflecting consumer-oriented technical factor priorities. The QFD converts CA(Customer Attribute) to EC(Engineering Characteristic) when developing a new product, and it makes possible to understand the priorities of each technology factor. When applying CA to HOQ(House Of Quality) of QFD, this study uses Word embedding(word2vec), a field of machine learning that converts natural language into vectors and expresses them so that computer can understand them, so that customer needs(CA) has been quantified. After that, a

relationship matrix between the CA and EC was made out using US patent data. In the second stage of the HOQ, the opinions of experts were reflected through expert questionnaires, and finally, connected them with the latest technology of CIS. In conclusion, this study aims to connect the latest technology factors of CIS with consumer needs based on big data analysis for clarifying the priorities of technology factors, and use them as reference materials for establishing CIS technology roadmap.

**Keywords :** CMOS image Sensor(CIS), Machine learning, Word2Vec, QFD