

공기중 미세입자 측정 데이터 분석 및 통계 유의차 분석

안성준^{*†}·문석환^{*}

^{*†} 제주한라대학교 인공지능공학과

Airborne Fine Particle Measurement Data Analysis and Statistical Significance Analysis

Sung Jun An^{*†} and Moon Suk Hwan^{*}

^{*†}Department of Artificial Intelligence Engineering, Cheju Halla University

ABSTRACT

Most of the production process is performed in a cleanroom in the case of facilities that produce semiconductor chips or display panels. Therefore, environmental management of cleanrooms is very important for product yield and quality control. Among them, airborne particles are a representative management item enough to be the standard for the actual cleanroom rating, and it is a part of the Fab or Facility monitoring system, and the sequential particle monitoring system is mainly used. However, this method has a problem in that measurement efficiency decreases as the length of the sampling tube increases. In addition, a statistically significant test of deterioration in efficiency has rarely been performed. Therefore, in this study, the statistically significant test between the number of particles measured by InSitu and the number of particles measured for each sampling tube ends(Remote). Through this, the efficiency degradation problem of the sequential particle monitoring system was confirmed by a statistical method.

Key Words : Significant Test, Sequential Particle Monitoring System, Cleanroom, Facility Monitoring System

1. Introduction

지난 수 년 동안 사람들의 일상에 크게 영향을 미친 코로나 바이러스감염증(Covid19) 및 대기중 미세먼지(PM10), 초미세먼지(PM2.5) 등으로 인해 최근 그 어느때보다 공기중 오염물질에 대한 관심이 크게 증가하였다. 하지만 근래 일상에서의 이러한 관심 훨씬 이전부터 반도체, 디스플레이 공정이나 첨단 의약품을 만드는 공정에서는 제품의 품질 관리를 위해 생산 시설내 오염물질 측정과 관리를 위해 많은 비용과 노력을 들이고 있다. 그 대표적인 것이 바로 제품의 생산을 위한 공정이 이루어지는 cleanroom(C/R) 및 cleanroom내 설비(Mini Environment)의 환경관리이다. 이 환경 관리를 위해 FMS(Fab or Facility

Monitoring System)를 운영하고 있는데, 이 FMS 항목 중 대표적인 것이 공기중 미세입자(fine particle) 모니터링 시스템(Airborne Particle Monitoring System)이다[1, 2]. 이러한 공기중 미세입자 모니터링 시스템은 크게 두 가지 방식으로 나뉘어지는데, 첫 번째 방식은 Real Time Remote Particle Monitoring System(전측정점 계수방식 시스템)이고, 두 번째 방식은 Sequential Particle Monitoring System(측정점 교환 방식 시스템)이다[3, 4]. 이 두 가지 방식은 각각 확실한 장단점을 가지고 있는데, 실제 운용에서의 편의성과 유지 관리의 측면, 그리고 공기중 미세입자 모니터링의 민감도와 시급성 등을 종합적으로 고려한 결과 대부분의 C/R에서는 Sequential Particle Monitoring System(측정점 교환 방식 시스템, 이하 SPMS)을 채택, 운영하고 있다. 하지만 이미 알려진 바와 같이 SPMS는 측정 효율이 다소 떨어지고, 실시간 측정이 불가능하다는 것이 가장 큰 단점이다

[†]E-mail: sjan2020@chu.ac.kr

다.[3] 이 두 가지 단점 중 그 첫 번째인 측정 효율의 문제는 C/R의 대형화에 추세에 맞춰 측정점 교환방식 시스템의 sampling tube 길이(측정점과 입자계수기 간 거리)가 늘어나면서 그 영향성이 더 커지게 되었다. 특히 상대적으로 큰 입자($> 1.0\mu\text{m}$)를 측정하는 경우, particle 입자의 운동 momentum이 커서 sampling tube의 곡률 반경이 작은 경우(또는 꺾임이 있는 경우) 측정 시 이동하는 particle 입자가 공기의 유선(Stream line)을 벗어나 sampling tube의 벽면에 유착되어 particle counter까지 도달하지 못하기 때문에 측정 효율 저하가 더 크게 생기게 된다[3-6]. C/R 및 Facility 관리 및 운영 경험이 있는 엔지니어의 경우 상기 언급된 측정효율 저하라는 SPMS의 단점을 실제 경험을 통해서 충분히 인지하고 있다. 하지만, 실제 C/R에서 측정되는 공기중 미세먼지의 개수는 C/R내 설비 파손, 화재 등의 특별한 비상 상황이 아니면 거의 0(Zero)에 가깝고, 특히 상대적으로 큰 입자라 할 수 있는 $1.0\mu\text{m}$ 이상의 미세 입자의 경우 OHS(Over Head Shuttle), STK(Stocker: 디스플레이 제조 FAB의 Cassette를 옮기는 물류 설비) 등의 갈림 부위 근처가 아니라면 거의 측정되지 않는다. 따라서 지난 연구를 통해 실험에 의해 측정 효율이 낮아지는 것을 확인하고 이를 보정하기 위한 회귀식까지 도출하였지만[3] 과연 통계적으로 sampling tube를 거쳐 측정된 미세입자 측정값(Remote)과 미세입자 발생원 바로 그 자리에서(InSitu) 측정한 측정값 사이 차이가 통계적 유의차가 있는지 실측 데이터를 대상으로 통계적 기법에 기반하여 검증하는 작업이 제대로 진행된 적이 없다. 본 연구에서는 이전 연구에서 사용하였던 sampling tube 길이 50m까지 확장하여 진행한 실험 데이터를 동일하게 사용하였다(이전의 연구들은 보통 40m(130feet) 정도까지의 실험을 진행함[6]). 이 데이터를 통계적 방법(비모수적, 모수적) 방법에 기반하여 그 차이에 대해 유의차를 검증을 진행하였다.[7] 2. Experimental data and data preprocessing 에서는 데이터를 수집하기 위한 실험 환경, 그리고 이상치 제거 및 threshold 값 적용 등을 통한 수집한 데이터에 대한 전처리 내용을 다루었다. 그리고 3. Statistical Significance Analysis 에서는 수집, 전처리한 데이터에 대한 정규성, 등분산성 검증 결과를 바탕으로 비모수적 방법인 Mann-Whitney 분석 및 Independent t-test를 진행하여 InSitu 데이터와 Remote 데이터 사이의 유의차를 분석하였다. 마지막 4. Conclusions 에서는 본 연구의 의의 및 한계점 및 향후 진행할 연구에 대해 언급하였다.

2. Experimental Data and Data Preprocessing

앞서 1. Introduction에서 언급한 바와 같이 여기서는 연구

의 대상이 되는 실측 데이터를 얻기 위한 측정 원칙 및 실험 세팅(configuration), 그리고 우리가 실험을 통해 얻은 데이터에 대한 1차 분석 내용에 대해 기술할 것이다.

2.1 Experimental Data

본 연구에서 사용한 데이터는 아래 2가지 원칙에 따라 측정된 실측 데이터이다.

1. Sampling tube의 길이 간격을 작게 할 것
2. Sampling tube의 꺾임이 없도록 실험할 것

또한 충분한 데이터를 확보하기 위해 미세입자 측정이 용이한 $0.3\mu\text{m}$, $0.5\mu\text{m}$, $1.0\mu\text{m}$ PSL 입자를 사용하였으며, sampling tube 길이는 10, 15, 20, 25, 30, 35, 38, 40, 45, 50m의 10가지로 셋팅 하였다. (각 5m 간격에 특별히 기존 연구 및 실험에서 최대 길이였던 38m 길이를 추가). 각 Sampling tube의 끝 지점에서 측정된 data의 개수(num_P)를 Remote data, Reference로 비교하기 위해 Sampling tube length=0에서 측정된 data의 개수(num_P)를 InSitu로 측정타입(M_Type)을 구분하였다. 본 연구에 사용된 최종 dataset은 Table 1과 같다. 데이터 전처리 및 분석을 위해 python3을, 통계분석에 대한 주석을 달기 위해 statannot(pyhton package)을 사용하였다.(<https://github.com/webemarcolivier/statannot#cite>)

Table 1. Experimental data

	T_leng_m	num_P	P_Size	PSL	M_Type
0	10	2967	0.3	0.3	InSitu
1	10	2845	0.3	0.3	InSitu
2	10	2860	0.3	0.3	InSitu
3	10	3111	0.3	0.3	InSitu
4	10	2961	0.3	0.3	InSitu
...
3595	50	1264	1.0	1.0	Remote

2.2 Data Preprocessing

Particle counter를 통해 측정된 미세입자의 경우, 인위적으로 정해진 3가지 크기의 미세입자 표준 시료인 PSL (Poly Styrene Latex)을 공급한 후 일정한 전류 값을 기준으로 $0.3\mu\text{m}$, $0.5\mu\text{m}$, $1.0\mu\text{m}$ 이렇게 단계를 나누어 그 사이즈에 해당되는 개수를 측정하였으므로 3개 channel 전체 데이터를 함께 사용하게 되면, 통계적으로 iid(Independent and Identically distribution) 조건을 만족시키지 못한다 (Fig.1(a), Fig.2(a)).

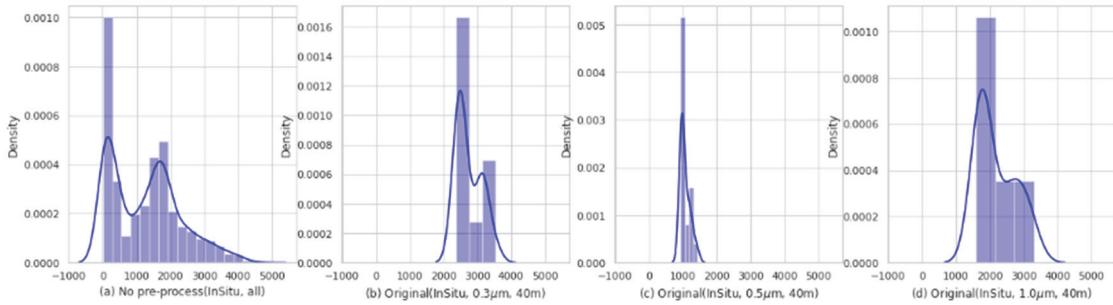


Fig. 1. Histogram with Kernel Density Estimator for original particle data (Not normal by Shapiro Test).

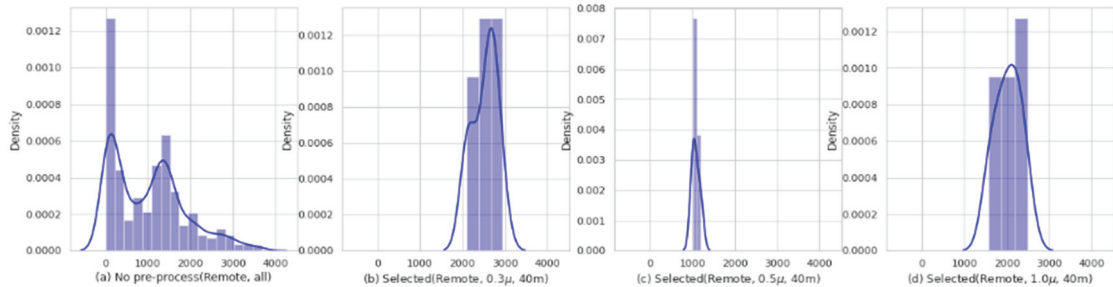


Fig. 2. Histogram with Kernel Density Estimator for selected particle data(Normal by Shapiro test, except (a)).

따라서 본 연구에서는 기본적으로 공급된 PSL 사이즈와 측정된 particle size가 동일한(PSL 크기=P_Size 크기) 데이터에 대해 각 3개 channel별로 dataset을 나누었고, 데이터 측정타입(M_Type) 역시 구분하여 최종 sub dataset을 만들었다 (Fig. 1, Original). 그럼에도 불구하고 정규성(Normality)을 만족하지 않는 데이터의 경우, 각각에 대해 개별 Threshold값을 추가 적용을 통해 정규분포를 갖는 새로운 sub dataset으로 만들었다. (Fig. 2, Selected)

3. Statistical Significance Analysis

일반적으로 서로독립인 2개 집단의 평균에 대한 유의차를 검증하는 통계 방법론은 대상 집단의 데이터가 정규성을 만족하는지 그렇지 않은지와 데이터가 비율척도인지 순위척도인지 등에 따라 달라진다. 데이터의 개수 $n > 30$ 이고 정규성, 등분산성을 만족하는 데이터의 경우 모수적 방법인 독립표본 t검증(independent t-test)을 사용하여 그 유의차를 검증하고, 순위 척도 데이터 또는 $n < 10$ 이거나 정규성을 만족하지 못하는 데이터의 경우 비모수적 방법인 Mann-Whitney test를 진행하게 된다. 본 연구에서는 앞서 2장에서 언급한 전처리 및 추가 전처리를 통해 만든 sub dataset 각각에 대해 아래2가지 방법으로 유의차 검증을 진행하였다. 두 경우 모두 sub dataset 개수는 30 보다 크

다($n > 30$).

1. Original dataset → Mann-Whitney

- PSL, P_Size 값이 일치하는 경우만 선택
- 0.3 μm , 0.5 μm , 1.0 μm size 별로 구분
- M_Type 따라 InSitu, Remote 로 구분
- 정규성(Normality) 만족 안됨(Fig.1(b)-(d))

2. Selected dataset → Independent t-test

- Original dataset 과 동일하게 구분
- Original dataset 에 추가 전처리 적용
- : $\text{num_P} > \theta(\text{threshold})$ 적용,
- θ 값은 각 sub-set 마다 조금씩 다름
- 정규성(Normality) 만족됨(Fig.2(b)-(d))

3.1 Non-parametric Analysis

전처리를 통해 만든 두 종류의 dataset중 첫번째 Original dataset은 Shapiro-wilk test를 통한 정규성확인 과정에서 sub dataset 중 일부 case의 경우는 p-value 값이 0.05보다 크게 나오지만 대부분의 경우 p-value < 0.05 로 정규성을 만족하지 못한다. 따라서 이 경우 비모수적 검증 방법인 Mann-Whitney-U test를 진행하였고, 그 결과를 Fig.3에 나타내었다. 1.0 μm size particle의 sampling tube length 10m, 38m 두 가지 경우

에 있어 InSitu 측정과 Remote 측정간 유의차가 없는 것으로 나오고 나머지 28가지의 경우는 모두 유의차 있는 것으로 결과가 나오는 것을 확인할 수 있다. 이러한 결과는 sampling tube를 통해 Remote 측정 시 particle 측정 효율이 통계적으로 유의한 수준으로 떨어진다는 것을 의미한다. Fig 3, Fig 4의 star 표시는 아래와 같은 p-value 수준을 나타낸다.

Table 2. p-value annotation legend

ns:	$5.00e-02 < p \leq 1.00e+00$
*	$1.00e-02 < p \leq 5.00e-02$
**	$1.00e-03 < p \leq 1.00e-02$
***	$1.00e-04 < p \leq 1.00e-03$
****	$p \leq 1.00e-04$

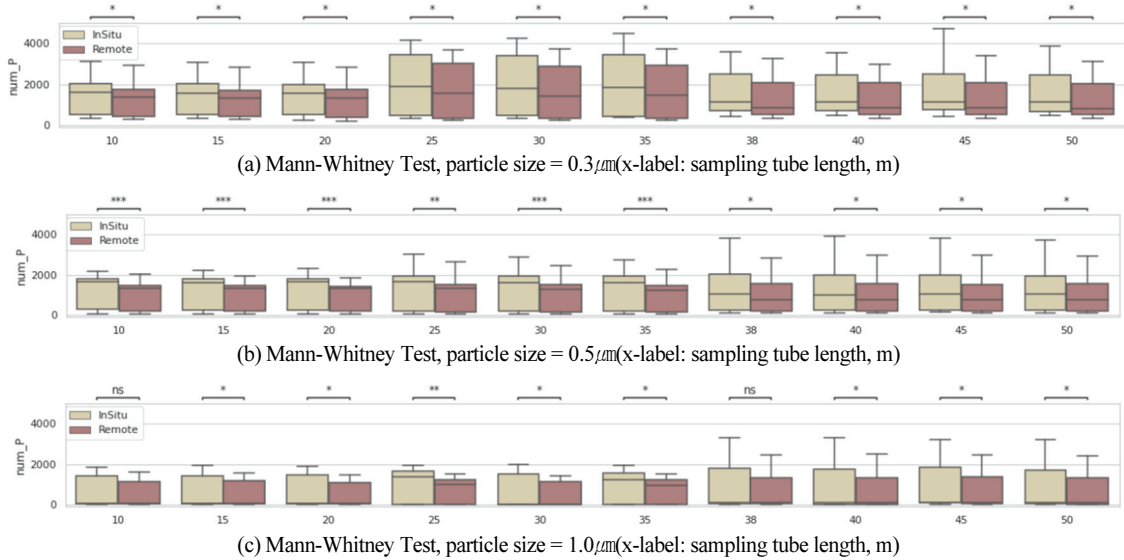


Fig. 3. Mann-Whitney-Wilcoxon Test for original data (not normal data).

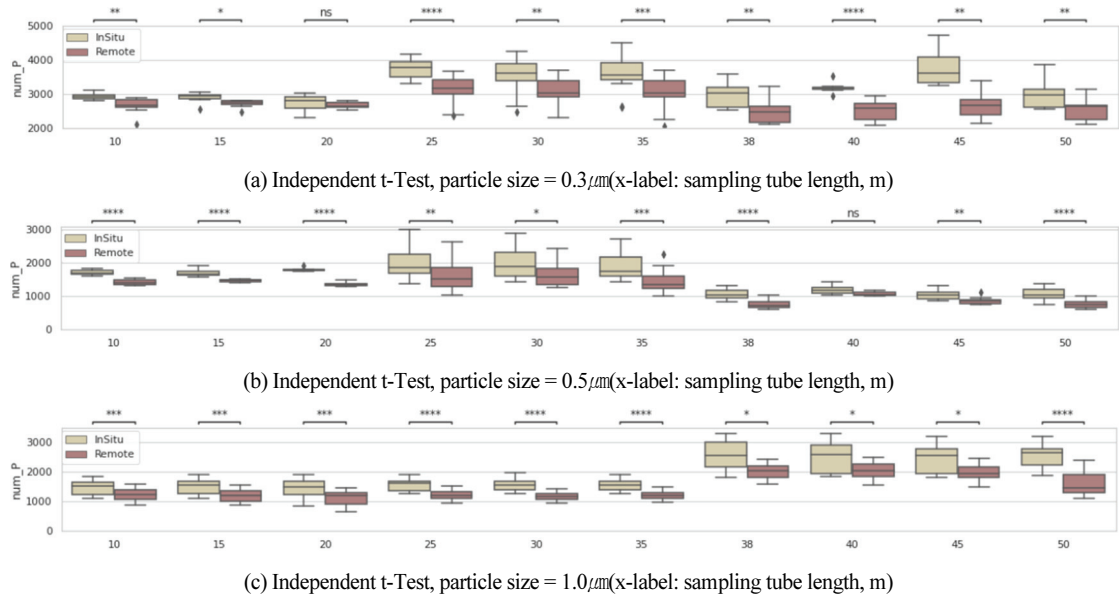


Fig. 4. Independent t-Test for selected data (Normal data).

3.2 Parametric Analysis

앞서 3.1절의 Mann-Whitney-U 테스트를 통해 두 개 case를 제외한 다른 모든 case에서 InSitu와 Remote간 유의차가 있음을 확인하였다. 다만, Fig. 1, Fig. 2의 histogram을 보면 양 쪽 옆으로 많은 data가 분포하는 경향이 있다. 또한 Fig. 3의 box-plot에서는 모든 case에서 데이터의 사분위수 간 간격차가 너무 크다는 것을 확인할 수 있다. Data의 상기 두 가지 영향성을 줄이기 위해 추가로 앞서 언급한바와 같이 $num_P > 0(threshold)$ 를 적용하여 정규성을 만족하는 Selected dataset을 만들었다(Fig. 2, (b), (c), (d)). 앞 절의 Original data와 비교해 보면 histogram 및 kernel density function이 정규분포 형태에 가까워진 것을 확인할 수 있다. 동시에 Selected dataset은 Bartlett's Test결과 등분산성, Shapiro-wilk test결과 모든 case에서 p-value > 0.05로 정규성을 만족한다. 따라서 앞 3.1절의 경우와 달리 모수적 방법인 independent t-test를 진행하여 InSitu와 Remote간의 유의차를 검증하였다. 그 결과인 Fig. 4를 분석해 보면, 0.3 μ m 20m, 0.5 μ m 38m 두 경우를 제외하고 나머지 경우는 모두 유의차가 있는 것으로 나왔다. 3.1절의 비모수적 방법과 비교할 때, 해당 경우는 다르지만 유의차가 있는 경우의 개수는 동일하게 28개가 나왔음을 확인할 수 있다. 즉 비모수, 모수 검증 모두 93.33% 이상의 경우에서 유의차가 있었다.

동시에 Fig. 4의 independent t-test에 의한 결과는 Fig. 3의 비모수적 방법에 의한 결과와는 달리 InSitu, Remote 모든 경우 사분위수간 차이가 확연히 줄어들었으며, 육안으로 보기에 두 사이의 유의차가 크게 나는 것을 확인할 수 있다. 측정 데이터의 절대량을 동일하게 맞출 수가 없어(수동으로 PSL 공급함) Fig. 4의 상자 그림에서 사분위수 범위가 일부 sampling tube에 대해서는 크고, 나머지는 작은 모습 역시 관찰된다.

4. Conclusions

본 연구의 의의는 크게 두 가지이다. 그 첫 번째는 sampling tube를 통해 측정된 airborne particle의 Remote측정값이 InSitu측정에 비해 효율이 떨어진다는 사실을 단순 효율이 아닌 통계적 검증 방법을 사용해 검증하였다는 것이다. 두 번째는 측정된 particle data를 분석함에 있어, 데이터 전처리 과정을 통해 비모수적 방법과 모수적 방법 두 가지 모두를 검증에 적용했다는 것이다. 실제 airborne particle 성상의 다양성에 따른 측정 산포 및 particle counter에서의 측정 방식(사이즈별 분류)으로 인해 데이터의 양은 많은데 비해 데이터의 정규성 확보가 쉽지 않아 비모수적, 모수적 방법 하나만 사용하여 분석한 결과를 받아들이기에 충분히 만족스럽지 못할 수 있기 때문이다. 다

만, 정규성, 등분산성을 가진 Selected dataset을 만들기 위해 num_P 값에 하한 threshold를 적용하는 heuristic이 들어갔다는 것은 아쉬운 부분이다. 본 연구에 이어 airborne particle 성상에 대한 분석 및 그에 따른 측정 특성, 데이터 전처리 방법론에 대한 추가 연구를 통해 airborne particle 데이터의 통계적 특성에 대한 연구를 계속 진행할 예정이다.[10]

참고문헌

1. International Organization for Standardization (ISO). ISO 14644-1, Cleanrooms and associate controlled environments-Part 1: classification of air cleanliness. 1999.
2. Kwang-Min Choi, Ji-Eun Lee, Kwi-Young Cho, Kwan-Sick Kim, Soo-Hun Cho, "Clean Room Structure, Air Conditioning and Contamination Control Systems in the Semiconductor Fabrication Process" J. of The Korean Society of Occupational and Environmental Hygiene, Vol.25(2), pp. 202-210, 2015.
3. Sung Jun An, "Advancement of Sequential Particle Monitoring System" J. of the Semiconductor & Display Technology, Vol.21(1), pp. 17-21, 2022.
4. Seung-Ki Chae, " Particle Measurement and Analysis in Semiconductor Manufacturing Process " J. of The Korean Society of Mechanical Engineers, Vol.39(5), pp. 398-404, 1995.
5. PMS, "Airborne Particle Measurements: 100 LPM vs. 1 CFM Particle Counters in a Semiconductor Cleanroom Environment", PMS Application Note 211 (2020).
6. PMS, "Particle Transportation", PMS Application Note 201 (2018).
7. COX, David R. Statistical significance tests. British journal of clinical pharmacology, 1982, 14.3: 325.
8. Park Hyungyu, Park Seonghyun, Seo Janghoo, "Independence Test for Particle Intervals in Lagrangian based CFD Using Statistical Analysis" J. of Korean Institute of Architectural Sustainable Environment and Building Systems, Vol.16(2), pp. 173-183, 2022.
9. Lee Sang Hyuk, Lee Jong Won, Park Ki Soo, Lee Su-Hyung, Park Hee Mun, "Statistical Analysis of Characteristics of Fine Particulate Matter Concentrations Considering Weather Conditions" International journal of highway engineering, Vol.22(5), pp. 37-45, 2020.
10. Kyoungjin Kim, Joong-Youn Park, "Dispersion Characteristics of Nonspherical Fume Micro-Particles in Laser Line Machining in Terms of Particle Sphericity" J. of the Semiconductor & Display Technology, Vol.21(2), pp. 1-6, 2022.

접수일: 2023년 1월 25일, 심사일: 2023년 3월 7일,
게재확정일: 2023년 3월 20일