

Analysis of Strategies for Using Limited Data to Recognize Abnormal Movements in Crowds

Da-Jeong Seo*, Yoo-Sung Kim**

*Undergraduate Student, Dept. of Artificial Intelligence, Inha University, Incheon, Korea

**Professor, Dept. of Artificial Intelligence, Inha University, Incheon, Korea

[Abstract]

This study investigates strategies for effectively utilizing limited data to recognize abnormal movements in crowd scenarios, based on a newly constructed crowd movement video dataset. Existing datasets such as Crowd-11 dataset focus on general flow patterns and lacks suitability for detecting abnormal behaviors. To address this, we redefined classification criteria and expanded the dataset with additional video samples. We tested various data composition strategies by changing the number of sequences per video and data splitting methods, and evaluated their impacts on recognition performance. Using optical flow for motion feature extraction and a Bidirectional LSTM model, we observed that more sequences led to improved performance, while fixed data splits showed high accuracy but risked overfitting. The proposed dataset and findings provide a solid foundation for developing abnormal crowd behavior detection models and offer practical guidance for data design in real-world applications.

▶ **Key words:** Abnormal Movement Recognition, Construction of Crowd Movement Dataset, Definition of Movement Classes, Data Usage Strategy

[요 약]

본 연구는 군중 상황에서 이상 이동을 인식하기 위한 제한된 데이터 활용 전략을 수립하고, 새롭게 구축한 군중 이동 비디오 데이터셋을 기반으로 이를 분석한다. 기존 데이터셋은 군중 이동 패턴 분류에는 유용하지만, 비정상적인 움직임 탐지에는 한계가 있다. 이에 따라 이상 이동 인식을 위한 분류 기준을 재정립하고, 해당 기준에 따라 군중 영상 샘플을 추가 수집하여 데이터셋을 확장하였다. 확장된 데이터셋을 바탕으로, 비디오로부터 시퀀스를 추출하는 방법 및 수, 데이터 분할 방식을 달리한 다양한 데이터 활용 전략을 수립하고, 각 전략이 이상 이동 인식 성능에 미치는 영향을 비교·분석하였다. 입력에는 Optical Flow 기법을 적용해 시공간 움직임 정보를 추출하고, 이를 Bidirectional LSTM 모델에 입력하여 이상 이동을 분류하도록 설계하였다. 실험 결과, 시퀀스 수가 많을수록 인식 성능이 향상되었으며, 고정된 데이터 분할 방식은 높은 정확도를 보였으나 과적합의 위험성도 함께 나타났다. 제한된 데이터셋과 분석 결과는 비정상 군중 행동 탐지 모델 개발을 위한 기초 자료로 활용 가능하며, 효율적이고 현실적인 데이터 구성 전략 수립을 위한 실질적인 가이드를 제공한다.

▶ **주제어:** 이상 이동 감지, 군중 이동 영상 데이터셋 구축, 이동 분류 정의, 데이터 사용 전략

-
- First Author: Da-Jeong Seo, Corresponding Author: Yoo-Sung Kim
 - *Da-Jeong Seo (7803jung@naver.com), Dept. of Artificial Intelligence, Inha University
 - **Yoo-Sung Kim (yskim@inha.ac.kr), Dept. of Artificial Intelligence, Inha University
 - Received: 2025. 07. 03, Revised: 2025. 09. 17, Accepted: 2025. 09. 17.

I. Introduction

군중영상 감시(video surveillance) 기술의 발전과 보편화에 따라, 군중의 생성 및 집합적 행동을 자동으로 인지하고 분석하려는 연구가 활발히 진행되고 있다[1-4]. 이를 위해 군중 내 보행자 인식, 이동 분석 등 다양한 측면에서 군중 상황을 다루는 기술들이 제안되고 있으며, 최근에는 딥러닝을 활용한 접근법을 통해 이러한 분석의 정밀도와 효율성을 높이고 있다[5-8]. 이에 따라 군중 상황을 보다 정확하고 효율적으로 이해하려는 시도는 다양한 형태로 확장되고 있으며, 그 활용 가능성 또한 점차 넓어지고 있다.

그중에 이상 이동 인식은 잠재적 위험 상황을 조기에 탐지하여 사회 안전을 증진시킬 수 있다는 점에서 중요한 과제이다. 군중 상황에서 이상 이동을 감지하기 위해서는 정상 이동과 이상 이동을 구분하기 위한 다양한 유형의 학습 데이터가 필요하다. 실제 인식 시스템에서 활용될 데이터셋 또한 이러한 유형을 기반으로 적절히 구성되어야 하며, 이를 통해 다양한 분석 기법들이 이상 이동을 효과적으로 인식할 수 있도록 설계되어야 한다.

군중 상황을 이동 형태에 따라 체계적으로 분류하려는 대표적인 시도로는 Crowd-11 데이터셋[9]이 있다. Crowd-11 데이터셋은 군중의 집합적 움직임 양상과 흐름 구조에 따라 11가지 유형으로 구분 정의하고 분류를 위한 기본 모델과 실험도 동시에 제시하였다. 해당 데이터셋은 이후 다양한 군중 분석 연구에서 널리 활용되었으며, 그중에 Choi et al.[10]은 3D convolution을 적용해 시공간 정보를 더욱 정밀하게 반영한 분류 모델인 3D-AIM을 제안하였다.

Abdurasulov[11]은 Crowd-11 데이터셋을 이상 이동 인식의 목적에 맞게 선별하여 재분류한 데이터셋과, optical flow와 궤적 정보를 통해 군중의 이동 정보를 추출하고 이를 시계열로 입력받는 LSTM 기반 인식기를 함께 제안하였다. 그러나 구성된 데이터의 수가 충분하지 않았기에, 각 비디오내에서 프레임 번호 끝자리가 같은 프레임들을 모아 10개의 시퀀스를 생성하여 모두 활용하는 방식을 채택하였다. 이 방식은 데이터 수가 제한된 환경에서 유용하지만, 동일한 비디오에서 생성한 10개 시퀀스 간의 유사성으로 인해 모델의 일반화 성능에 영향을 줄 수 있다.

본 연구는 군중의 이상 이동 인식을 주된 목표로 하는 데이터셋을 새롭게 구성하고, 시퀀스 활용 방식을 다양하게 구성함으로써 효과적인 데이터 구성 및 활용 방안에 따른 효과를 분석하는 것을 목적으로 두 가지 방향에서 기여

한다. 첫째, 기존 연구의 분류 기준을 보완하고 전체 비디오 샘플 수를 기존 80개에서 190개로 확장하여 학습 데이터의 다양성을 향상시켰다. 둘째, 단일 비디오에서 추출된 시퀀스 일부를 선택적으로 활용하여 시퀀스 간 유사성을 낮추고, 각 분할 사용 전략에 따른 군중 이상 이동 인식의 정확도 비교를 통해 경향성을 분석하였다.

본 논문의 구성은 다음과 같다. 2절에서는 군중 상황에서 이상 이동 인식을 위한 대표적인 데이터셋과 이동 패턴 분류 및 이상 이동 인식과 관련된 기존 연구를 소개한다. 3절에서는 본 연구에서 제안하는 이상 이동 인식용 데이터셋의 구조와 구성 과정을 설명하며, 4절에서는 실험 설정과 결과를 통해 제안한 데이터 활용 전략의 효과를 분석한다. 마지막으로 5절에서는 본 연구를 요약하고 향후 연구 방향을 제시한다.

II. Related Works

본 절에서는 군중 상황을 다루는 대표적인 데이터셋인 Crowd-11[9]과 이를 활용한 모델인 3D-AIM[10], 그리고 이상 이동 인식을 목적으로 제안된 Abdurasulov[11]의 접근을 각각 소개한다. 더불어 이들이 군중 상황에서 이상 이동을 감지하는 데 있어 갖는 한계 또한 함께 살펴보고자 한다.

2.1. Crowd-11

Crowd-11은 다양한 군중 흐름의 방향성과 공간적 구조를 기준으로 구분한 데이터셋으로, 총 11개의 클래스로 구성되며 6,000개 이상의 비디오 샘플을 포함하고 있다. 각 클래스는 군중의 집합적인 패턴에 기반해 정의되며, 군중 상황 분류, 밀집도 분석, 흐름 변화 감지 등 다양한 연구에서 널리 활용되고 있다[12-14].

그러나 해당 데이터셋은 움직임의 방향성이나 밀도 변화 등 군중의 구조적 특성에 따라 군중의 이동 형태를 분류하는 데 목적을 두고 있으며, 이상 이동을 명확하게 정의하거나 구분하지는 않았다. 예를 들어, 빠른 흐름이나 밀집된 움직임은 상황에 따라 질서 있는 이동으로도, 혹은 공황 상태를 반영하는 이상 이동으로도 해석될 수 있어 구분 기준이 모호하다. 이러한 분류 기준은 군중 흐름의 일반적인 양상을 학습하는 데는 유용하지만, 정상과 이상을 명확히 구분해야 하는 인식 과제에는 적합하지 않다.

Crowd-11을 제안한 연구에서는 군중 흐름 유형을 분류하기 위한 기본 모델도 함께 제시하였으며, 이후 분류

성능 향상을 목적으로 다양한 후속 모델들이 개발되었다. 그중 하나로 Choi et al.[10]은 3D convolution 기반의 3D-AIM 모델을 제안하였다. 3D-AIM은 다양한 구성 요소를 결합하여 시공간적 특징을 정밀하게 추출하고 클래스 간 분리를 극대화함으로써, Crowd-11의 11개 클래스를 높은 정확도로 분류할 수 있도록 설계되었다. 그러나 이 모델 역시 군중 흐름 유형 간의 구조적 차이를 효과적으로 분류하는데 초점을 맞추고 있기 때문에, 특정한 행동의 목적이나 맥락을 기반으로 이상 이동을 인식하는 데에는 제한적이다.

2.2. Abnormal Movement Dataset and Two-Stream LSTM Model

Abdurasulov[11]은 군중 상황 내 이상 이동 인식을 위해 Crowd-11 데이터셋을 목적에 맞게 선별 및 재구성하고, 이를 기반으로 한 LSTM 분류 모델을 함께 제안하였다. 구성된 데이터셋은 총 80개의 비디오 샘플로 이루어져 있으며, 정상 상황을 나타내는 Linear paths, Individual movement, Static, No crowd의 네 가지 클래스와 이상 이동 상황에 해당하는 Fighting, Scattering, Opposite movement의 세 가지 클래스로 총 일곱 개의 클래스로 구성된다.

Abdurasulov[11]은 해당 데이터셋을 기반으로 이상 이동을 분류하기 위한 2-stream LSTM 모델을 함께 제안하였다. Fig.1에 나타난 바와 같이, 분류기는 군중의 이상 이동 양상을 효과적으로 식별하기 위해 optical flow와 궤적(trajecory) 정보를 각각 입력으로 활용하는 two-stream 구조로 설계되었다. Optical flow는 프레임 간 시공간적

움직임 변화를, 궤적 정보는 객체의 연속적인 이동 경로를 통해 장기적인 행동 패턴을 파악하는 데 사용된다. 이처럼 서로 다른 정보를 통합함으로써, 모델은 복합적인 군중 행동을 보다 정밀하게 인식할 수 있도록 구성되었다.

이러한 시계열 기반 딥러닝 모델은 이상 이동의 발생 시점을 민감하게 포착할 수 있다는 장점이 있지만, 일반적으로 충분한 학습 데이터를 전제로 하기 때문에 데이터 구성 방식이 모델의 성능과 일반화 가능성에 크게 영향을 미친다. 예를 들어, 동일한 비디오에서 생성된 시퀀스 간의 유사성이 높을 경우, 모델이 특정 패턴에 과적합하여 실제 상황에서 이상 이동을 정확히 인식하지 못할 수 있다. 따라서 시퀀스 간 구분 가능성과 데이터의 다양성은 모델 성능을 결정짓는 핵심 요소로 작용한다.

III. The Proposed Dataset

Abdurasulov[11]은 이상 이동 인식 모델 학습을 목적으로 Crowd-11 데이터셋을 재구성한 소규모 데이터셋을 제안하였으나, 전체 데이터 규모가 제한적이라는 점에서 일반화된 학습에 어려움이 존재한다. 이에 본 연구는 이러한 기존 연구의 한계를 극복하고, 이상 이동과 정상 이동 간의 경계를 보다 명확히 규정하기 위해 클래스 기준의 일관성과 해석 가능성을 향상시켰고 이에 따라 데이터셋을 정비하였다.

최종적으로 구성된 데이터셋은 총 190개의 비디오 샘플로 구성되며, 모든 비디오는 30fps로 구성되었으며 80~100개의 프레임을 포함한다. 전체 데이터셋은 [11]에

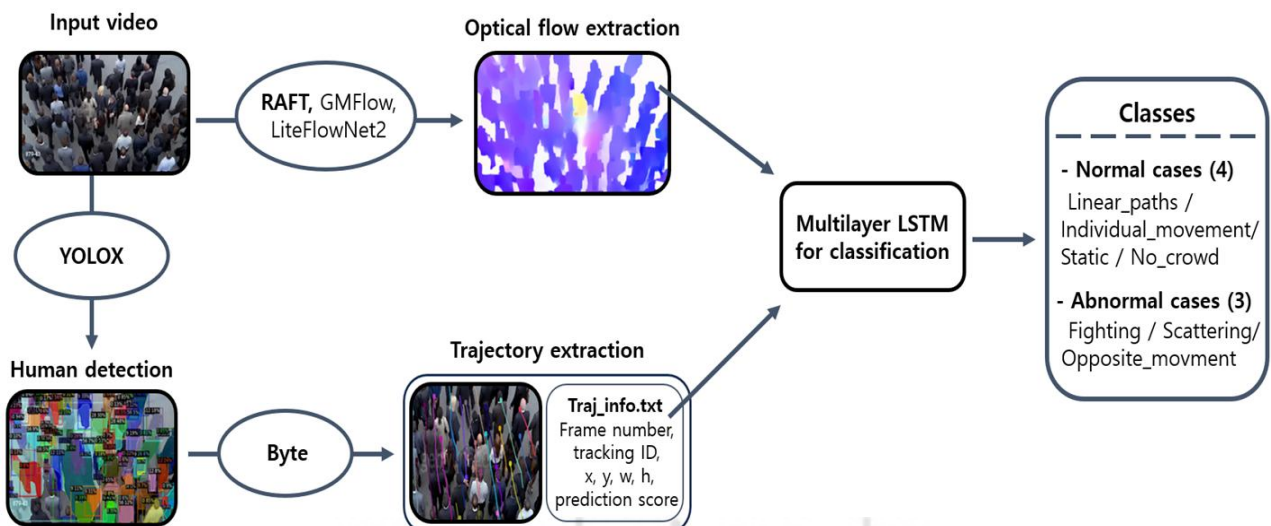


Fig. 1. Baseline classification pipeline proposed by Abdurasulov[11]

Table 1. Dataset composition by source and class (number of videos).

| | Crowd-11 | POND5 | WWW | Getty Images | Violent flow | CUHK | Youtube | PETS | UMN | count |
|---------------------|-----------|-----------|-----------|--------------|--------------|-----------|----------|----------|----------|------------|
| Linear_paths | 4 | 3 | 1 | 10 | 0 | 3 | 2 | 0 | 0 | 23 |
| Individual_movement | 12 | 12 | 1 | 7 | 0 | 19 | 0 | 2 | 5 | 58 |
| Static | 10 | 5 | 0 | 2 | 0 | 0 | 2 | 2 | 1 | 22 |
| Opposite_movement | 14 | 8 | 2 | 17 | 0 | 5 | 0 | 0 | 0 | 46 |
| Scattering | 14 | 0 | 0 | 1 | 0 | 0 | 2 | 2 | 0 | 19 |
| Fighting | 3 | 2 | 6 | 6 | 5 | 0 | 0 | 0 | 0 | 22 |
| count | 57 | 30 | 10 | 43 | 5 | 27 | 6 | 6 | 6 | 190 |

서와 같이 Crowd-11 기반의 샘플을 포함하며, 추가로 POND5[17], CUHK[18], Getty Images[19] 등 다양한 출처에서 수집된 데이터를 함께 활용하였다. 각 출처별, 클래스별 샘플 수의 구체적인 분포는 Table 1에 제시하였다. 본 데이터셋은 서로 다른 환경, 군중 밀집도, 카메라 시점을 포함하고 있어 이상 이동 인식 성능의 일반화 가능성을 높인다.

제안된 데이터셋은 Table 2에 나타난 바와 같이 총 6개의 클래스로 구성되었으며, 정상 상황(normal case)을 나타내는 3개, 이상 상황(abnormal case)을 나타내는 3개의 클래스로 구분하였다. 클래스 정의는 Abdurasulov의 기존 연구[11]에서 제시한 7개 클래스 분류 기준을 기반으로 하되, 이상 이동 인식의 목적에 보다 적합하도록 다음과 같은 방식으로 수정 및 보완하였다.

우선 기존의 No crowd 클래스는 장면 내에 인물이 존재하지 않거나 움직임 자체가 거의 없어, 이상 이동 인식과 관련이 적어서 본 연구에서는 분석 대상에서 제외하였다. 그리고 Opposite movement 클래스는 Fig.2에서 볼 수 있듯이, 기존에는 소수 개인이 군중 흐름과 반대 방향으로 이동하는 상황만을 포함하도록 정의되었으나, 본 연

구에서는 두 개의 뚜렷한 이동 방향이 존재하여 서로 반대 되는 흐름이 발생하는 상황으로 명확하게 정의하였다. 이를 통해 이상 이동을 개별 행위가 아닌, 군중 전체의 이동 구조 차원에서 해석할 수 있도록 하였다. 또한, Individual movement는 보행자들이 서로 충돌하지는 않지만, 주된 이동 방향이 세 개 이상(≥ 3)으로 분산되어 일관된 흐름을 형성하지 못하는 상황을 의미하지만, Opposite movement는 장면 내에 두 개(=2)의 뚜렷한 흐름이 존재하며, 이들이 서로 반대 방향으로 교차하거나 충돌하는 상황을 나타낸다. 두 클래스의 차이를 직관적으로 확인할 수 있도록, 대표적인 영상 프레임을 예시로 제시하였으며, 이를 Fig.2에 나타내었다.

이외의 Linear paths, Static, Scattering 클래스는 기존 정의의 핵심 틀을 유지하되, 장면 전체의 이동 방향성, 지속 시간, 움직임의 공간적 분포 양상 등을 보다 명확히 기술함으로써 기존의 일관성과 해석 가능성을 높였다. 데이터 라벨링은 사전에 정의한 기준에 따라 세 명의 연구원이 각 샘플을 독립적으로 판별하고, 다수결로 클래스를 결정하였다. 이 과정을 총 세 차례 반복함으로써 라벨의 일관성을 확보하였다. 본 데이터셋은 정상/이상 이동을 명확

Table 2. Annotation criteria and sample distribution in proposed dataset.

| Class | | Definition | Number of samples | Number of frames (average/min) | fps |
|----------|---------------------|---|-------------------|--------------------------------|-----|
| Normal | Linear_paths | The overall movement in the scene follows a single dominant direction. | 23 | 99.89 / 99 | 30 |
| | Static | No significant movement is observed throughout the entire video. | 22 | 99.59 / 91 | 30 |
| | Individual_movement | A situation in which multiple(≥ 3) separate flows are dispersed, with individuals moving independently without collisions. | 58 | 99.96 / 98 | 30 |
| Abnormal | Opposite_movement | Two(=2) dominant movement directions are presents, flowing in opposite directions. | 46 | 99.83 / 92 | 30 |
| | Scattering | Little movement is observed at the beginning, followed by a sudden emergence of rapid motion. | 19 | 100 / 100 | 30 |
| | Fighting | The scene exhibits fighting behaviors between two or more groups. | 22 | 98.32 / 94 | 30 |

| Class | Individual_movement | | Opposite_movement | |
|------------------|---------------------|---|-------------------|---|
| Sample images | | | | |
| By [11] criteria | 0 | 0 | 0 | X |
| By our criteria | 0 | 0 | 0 | 0 |

Fig. 2. Representative samples and criteria comparison for Individual movement and Opposite movement.

히 구분할 수 있도록 설계되었으며, 실제 환경에서 발생 가능한 다양한 상황을 고려하여 구성되었기에 이상 이동 인식 모델의 학습 및 평가에 효과적으로 활용될 수 있다.

IV. Experiments

본 절에서는 군중 상황에서 이상 이동 인식을 위한 비디오 데이터셋으로부터 학습 데이터셋을 다양한 방식으로 구성하고 활용한 여섯 가지 데이터 활용 전략을 소개한다. 각 전략은 비디오에서 추출하는 시퀀스 수와 데이터 분할 방식 등을 달리하여 설계되었으며, 모든 설정에서 optical flow 기반 입력으로 분류 정확도를 평가하였다.

모든 실험은 총 190개의 비디오 샘플로 구성된 데이터셋을 기반으로 수행하였다. 시퀀스는 프레임 번호 끝자리가 같은 프레임들을 연결하여 10개로 구성하는 Abdurasulov[11]의 방식을 따르며, 구체적인 절차와 도식은 4.1에 정리하였다. 이후 실험 입력에서는 유사 장면의 중복을 줄이기 위해 0번 시퀀스를 포함하고 나머지는 균등 간격으로 선택하는 규칙을 적용하였으며, 세부 설명은 4.2에 제시한다.

모델 측면에서, 이상 이동 인식을 위해 [11]에서 제안된 LSTM 기반 분류 모델을 사용하되, 기존 방식과 달리 궤적 (trajectory) 정보를 사용하지 않고 optical flow 정보만을 입력으로 사용하였다. 이는 기존 연구에서 optical flow는 이동을 반영하는 주된 입력으로 사용되어왔으며, 데이터 분할 전략의 효과를 명확히 분석하기 위함이다. 또한 [11]의 단방향 LSTM과 달리 Bidirectional LSTM을 적용하여 시계열 정보의 양방향성을 반영하였다. Optical flow는 기존[11]에 사용된 RAFT[15] 뿐 아니라, 군중 상황의 복잡한 흐름을 효과적으로 반영할 수 있는 MemFlow-T[16] 방식도 함께 적용하였다. 전체 실험 과정은 Fig.3에 제시된 분류 파이프라인을 기반으로 하며, 각 비디오에서 10개의 시퀀스를 생성한 뒤 optical flow를 추출하고, 이 중, 일부 D개 시퀀스를 선택하여 Bidirectional LSTM 분류기의 입력으로 사용한다.

학습 과정에서는 검증(validation) 정확도를 기준으로 조기 종료(early stopping)을 적용하였고, patience는 7로 설정하여 일정 에폭(epoch) 동안 성능 개선이 없으면 학습이 조기 종료되도록 설정하였다.

모든 실험은 PyTorch 프레임워크를 기반으로 수행되었으며, Intel i9 9980XE CPU, NVIDIA TITAN RTX GPU

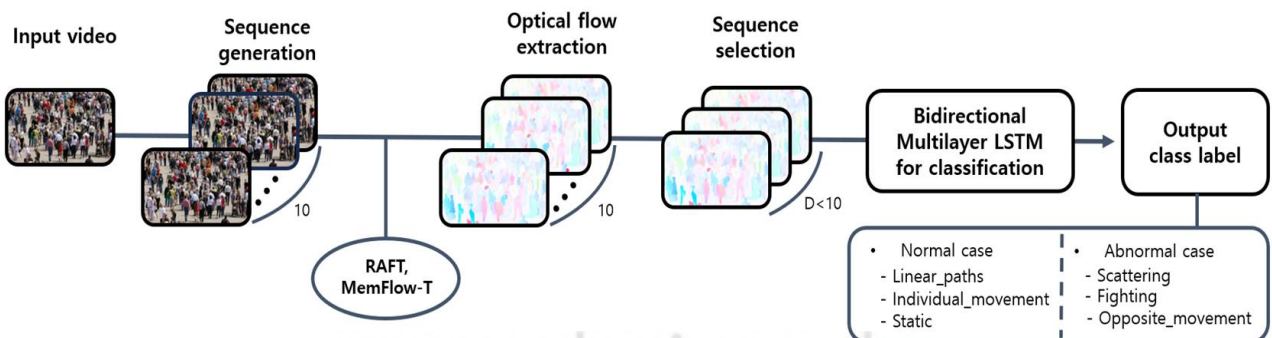


Fig. 3. Overview of the proposed classification pipeline.

Table 3. Hyper-parameter values for the Bidirectional LSTM model used for classification.

| Multilayers | Input size | Hidden size | Dropout rate | Learning rate | Max epoch | Patience | Batch size | Normalization |
|-------------|------------|-------------|--------------|---------------|-----------|----------|------------|---------------|
| 2 | 256 | 512 | 0.5 | 0.001 | 100 | 7 | 32 | batch |

Table 4. Summary of data split and usage strategies.

| | Strategy 1 | Strategy 2 | Strategy 3 | Strategy 4 | Strategy 5 | Strategy 6 |
|---|--------------------------------------|------------|------------|------------|---------------|--|
| Number of using sequences per video | 1 | 2 | 3 | 4 | 5 | 5 |
| Using sequences (Seq #) | 0 | 0, 5 | 0, 3, 7 | 0, 2, 5, 7 | 0, 2, 4, 6, 8 | 0, 2, 4, 6, 8 |
| Data split method (Train/Validation/Test) | 6:2:2 random split within each video | | | | | Train : 0, 4, 8 Val : 2 Test : 6 |

2개, 125GB 메모리로 구성된 시스템에서 실행되었다. 분류에 사용된 Bidirectional LSTM 모델의 주요 하이퍼파라미터 설정은 Table 3에 정리되어 있다. 본 연구는 Abdurasulov[11]의 설정을 우선적으로 참고하였으나, multilayer(1, 2, 3, 4), hidden size(256, 700, 800), dropout rate(0.1, 0.3, 0.5), learning rate(0.01, 0.001, 0.002, 0.003, 0.004, 0.005), patience(5, 7, 10) 등 주요 하이퍼파라미터에 대해서는 후보군을 설정하여 비교 실험을 진행하였다. 그 결과 최적의 성능을 보인 값들이 Abdurasulov[11]의 설정과 대부분 일치하거나 유사하게 나타났다. 이를 통해, 기존 [11]의 설정이 본 연구의 데이터셋에서도 타당함을 확인하였다. 단, 배치 크기의 경우, batch normalization을 적용하더라도 전체 데이터를 충분히 활용할 수 있도록 16에서 32로 조정하여 실험을 진행하였다. 진행된 여섯 가지 실험내용은 Table 4에 요약되어 있다.

4.1. Sequence Construction

본 연구에서는 각 비디오(약 80-100 프레임)에서 프레임 번호의 끝자리가 같은 프레임을 묶어 10개의 시퀀스

(seq 0-9)를 생성하였다. 이 방식은 시간 축을 균등 간격으로 표본화하여 비디오 전 구간의 움직임을 고르게 반영하는 간결하고 일관된 방법이다. 구체적인 절차는 Fig. 4에 제시하였으며, 이렇게 생성한 시퀀스를 바탕으로 학습, 검증, 평가 세트를 구성하였다.

이 구성 방식의 타당성을 확인하기 위해, 동일 길이(10 프레임) 조건에서의 시작 위치에 따른 실험을 수행하였다. 즉, 연속 10프레임 하나를 사용하는 경우를 세 가지(초반 0-9프레임, 중반 40-49프레임, 후반 70-79프레임)로 구분하고, 이를 제안 방식으로 생성한 시퀀스 중 하나만 사용하는 경우인 Strategy 1과 비교하였다. 모든 결과는 MemFlow-T 기반이며, 5-폴드 교차 검증(5-fold cross validation)의 평균 정확도를 기준으로 평가한다.

Table 5에 정리된 결과를 보면 초반/중반/후반 연속 프레임 구성 방식 대비 제안 방식은 일관된 성능 개선을 보였다. 이는 전체 비디오에서 일정 간격으로 10개의 프레임을 선정하는 것이 국소 구간에서만 동일 개수의 프레임을 선정하는 것보다 정보 효율 측면에서 유리함을 시사한다.

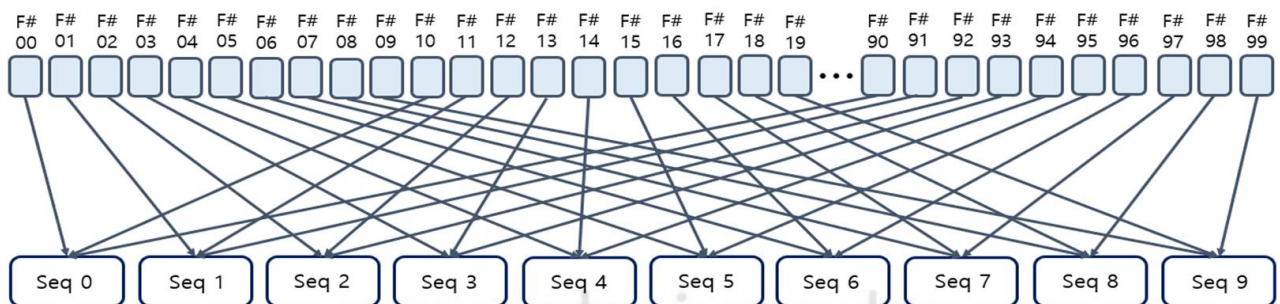


Fig. 4. Sequence generation process from video frames.

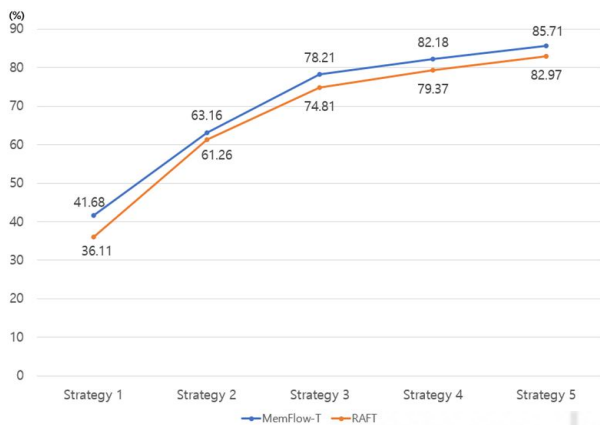
Table 5. Start-position ablation with 10-frame sequence.

| Method | Accuracy |
|-----------------------------------|----------|
| Strategy 1 | 41.68% |
| Beginning part (frame 000~009) | 36.95% |
| Middle part (frame 040~049) | 36.11% |
| End part (frame 070~079) | 33.26% |

4.2. Experiments by Number of Sequences

본 실험은 하나의 비디오에서 추출하여 사용하는 시퀀스 수인 D 를 점차적으로 증가시켜, 입력 정보의 양과 다양성이 모델 성능에 어떤 영향을 미치는지를 관찰하기 위해 설계되었다. 이때 시퀀스는 유사 장면의 중복을 최소화하고 다양한 장면 구성을 확보하기 위해, 항상 0번 시퀀스를 포함한 뒤 나머지는 균등한 간격을 두어 일정하게 선택하였다. 각 실험에서는 선택된 시퀀스에 원본, 수평 반전, $\pm 10^\circ$ 회전, $1.1\times$ 확대를 적용해 총 5개로 증강하였다. 5-폴드 교차 검증을 수행하여 이에 대한 평균 정확도를 기준으로 결과를 분석하였다.

가장 단순한 조건인 시퀀스 1개를 사용하는 Strategy 1에서는 RAFT 기준 36.11%, MemFlow-T 기준 41.68%의 정확도를 기록하며 가장 낮은 성능을 나타냈다. 시퀀스 수가 2개로 늘어난 Strategy 2에서는 정확도가 급격히 상승하였고, RAFT는 Strategy 1 대비 25.15%p 상승한 61.26%를, MemFlow-T는 21.48%p 상승한 63.16%를 기록하였다. 이후 시퀀스를 3개, 4개로 점진적으로 늘린 Strategy 3과 Strategy 4에서도 성능은 안정적으로 향상되었으며, 5개 시퀀스를 사용한 Strategy 5 조건에서는 RAFT가 82.97%의 정확도를 보였고, MemFlow-T가 85.71%로 전체 실험 중 가장 높은 성능을 기록하였다. 각 전략에 따른 경향성은 Fig.5에서 확인할 수 있다.

Fig. 5. Performance trends by sequence count ($D=1\sim 5$).

시퀀스 수 증가에 따른 정확도 향상은 일관된 패턴을 보였다. 특히 시퀀스를 2개 이상 사용할 경우부터는 정확도가 뚜렷하게 개선되었는데, 이는 시퀀스 수 증가가 시간적 패턴 학습에 유리하게 작용하고, 더 다양한 장면 구성을 포함함으로써 모델의 표현력을 효과적으로 높였기 때문으로 해석할 수 있다. 반면, 시퀀스를 4개 이상 사용한 Strategy 4와 Strategy 5 조건에서는 성능 향상 폭이 완만해졌으며, 이는 일정 수준 이상의 정보가 확보된 이후에는 추가 시퀀스가 학습에 기여하는 효과가 줄어들 수 있음을 보여준다. 또한, 모든 실험에서 MemFlow-T가 RAFT보다 우수한 성능을 보였으며, 이는 optical flow 추출 방식 또한 모델 성능에 영향을 미친다는 점을 의미한다.

추가적으로 각 전략별 분류 성능의 세부 양상은 Fig.6의 분류 기준표(confusion matrix)에 제시하였다. 이는 5-폴드 교차 검증 중 평균 정확도와 가장 가까운 경우를 대표로 선택한 것이다. Fig.6을 보면 전반적으로 Strategy 1에서 5로 갈수록 대각선 성분이 강화되고 비대각의 오분류가 점진적으로 감소함을 확인할 수 있다. 특히 Individual_movement와 Opposite_movement 간 상호 혼동이 뚜렷이 줄어, 다중 시퀀스 입력이 상반된 흐름의 분리에 유효함을 확인하였다.

Linear_paths와 Static은 비교적 안정적인 정분류율을 보이며, 시퀀스 수가 늘어날수록 안정성이 추가로 향상된다. 한편 Scattering은 Individual_movement, Static과의 경계에서 일부 혼동이 잔존하고, Fighting은 근접 대처로 인한 양방향 움직임이 강조되면서 Opposite_movement로의 오분류가 발생하는 경향을 보인다.

4.3. Experiments by Sequence Splitting Method

앞선 실험(Strategy 1~5)에서는 시퀀스 수에 따른 성능 변화를 관찰하기 위해, 선택된 시퀀스들을 5-폴드로 구성할 때 무작위로 분할하여 학습, 검증, 테스트 데이터셋을 구성하였다. 본 절에서는 시퀀스를 고정된 기준에 따라 분할한 Strategy 6을 추가로 구성하고, 분할 방식의 차이가 모델 성능에 미치는 영향을 비교하고자 한다.

Strategy 6은 Strategy 5와 동일한 개수의 시퀀스를 사용하되, 분할 방식에 있어 각 비디오에서 생성된 10개의 시퀀스 중 학습에는 0번, 4번, 8번 시퀀스를, 검증에는 2번, 테스트에는 6번 시퀀스를 고정 배정하였다. 평가 방법의 경우, Strategy 5는 전체 데이터를 기반으로 5-폴드 교차 검증을 수행하여 평균 정확도를 산출하였으며, 고정 분할 방식이 적용된 Strategy 6은 동일한 방식의 교차 검증이 어려워 랜덤 시드(random seed)를 달리한 5회 반복 실험을 수행한 뒤 평균 정확도를 기준으로 결과를 분석하였다.

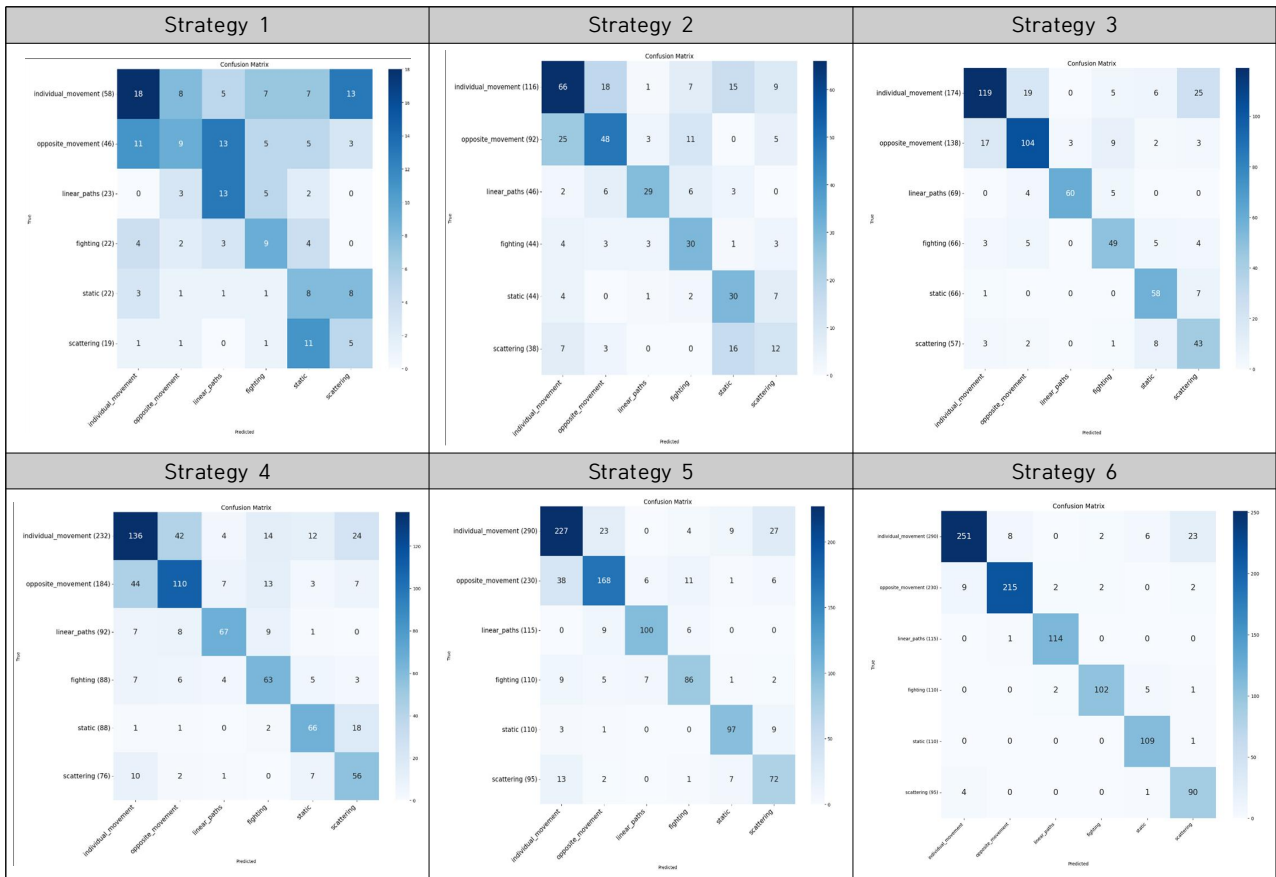


Fig. 6. Confusion matrices for Strategy 1~6 (MemFlow-T).

Fig.7에 나타난 실험 결과를 보면 고정 분할 방식인 Strategy 6이 무작위 분할을 적용한 Strategy 5보다 전반적으로 높은 정확도를 기록하였다. RAFT를 사용한 경우, Strategy 5는 82.97%의 정확도를 보였으며, Strategy 6에서는 88.55%로 5.58%p 향상되었다. MemFlow-T를 적용한 경우에도 Strategy 5의 정확도는 85.71%였고, Strategy 6에서는 92.02%로 6.31%p 상승하였다.

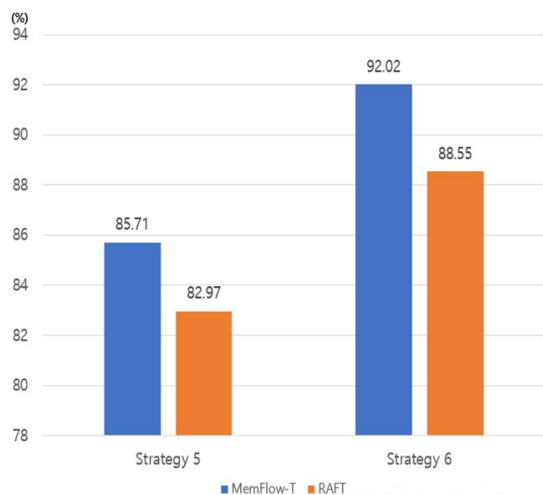


Fig. 7. Performance comparison by data split method.

Strategy 6에서는 각 비디오에 대해 테스트 시퀀스(6번)와 시간적으로 인접하여 유사성이 높을 수 있는 시퀀스들(4, 8번)이 학습 데이터에 항상 포함되도록 구성되어 있다. 반면, Strategy 5에서는 전체 시퀀스를 무작위로 폴드를 나누기 때문에 학습 데이터에 테스트 시퀀스와 인접한 프레임이 포함될 수도, 포함되지 않을 수도 있다. 이러한 차이로 인해 Strategy 6은 학습과 테스트 간 유사성이 일관되게 형성되어, 모델이 테스트 데이터에 익숙한 패턴을 잘 예측할 수 있는 조건을 제공하게 된다. 그러나 이와 같은 구조는 테스트 시퀀스와 시간적으로 가까운 정보가 학습에 포함되어 있다는 점에서 과적합의 가능성을 동반할 수 있다. 다만 이는 실제 과적합을 의미한다기보다 평가 결과가 학습 데이터셋과 평가 데이터셋을 구성하는 시퀀스 간 관계로 인해 높게 나타날 위험 요인이 존재함을 시사한다. 따라서 Strategy 6은 특정 목적에서는 유용할 수 있으나, 해석 시 데이터 분할 구조에 따른 영향을 함께 고려해야 한다.

추가적으로 Strategy 5와 Strategy 6의 클래스별 분류 결과를 비교하기 위해 분류 결과표를 앞의 Fig.6에 포함하여 제시하였다. 이를 통해 Strategy 6은 전반적으로 대각선 성분이 강화되었으며, 특히 Opposite movement에 대해 Strategy 5에 비해 정분류율이 많이 개선됨을 확인할 수 있다.

4.4. Experiments by Sequence Length and Position

본 절에서는 동일한 학습 및 조건 하에서 제안한 전략 Strategy 1과 전체 프레임을 활용하여 시퀀스를 구성하는 방법(all_frames)과 시퀀스에 포함되는 프레임의 수와 같은 개수의 프레임을 선정하는 위치에 따른 효과를 정량적으로 분석한다.

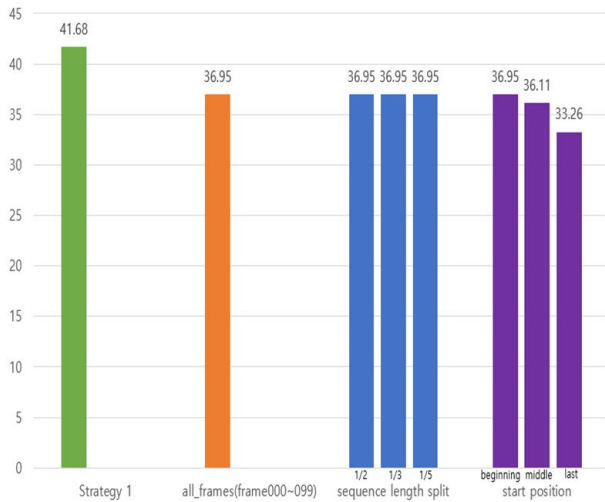


Fig. 8. Accuracy comparison : Strategy 1, all_frames, length splits, start positions.

여기서, all_frames은 각 비디오(최대 100프레임)를 연속된 단일 시퀀스(예: frame000-099)로 구성하여 입력하는 방식이며, 비교의 공정성을 위해 동일한 분류기와 평가 방식을 적용하고 optical flow를 MemFlow-T로 고정하여 실험하였다.

실험 결과, all_frames의 정확도는 36.95%로 Strategy 1(41.68%) 대비 4.73%p 낮았다. 원인 분석을 위해 (i) 길이 분할(100개의 프레임을 1/2, 1/3, 1/5로 분할)과 (ii) 시작 위치(초반 0-9프레임, 중반 40-49프레임, 후반 70-79프레임으로 구분) 실험을 추가 수행한 결과, 긴 연속 시퀀스의 정보가 전체 길이에 걸쳐 고르게 활용되지 않고 시퀀스의 초반 구간에 편중되는 경향을 확인하였다. 이는 모델의 시퀀스 길이(sequence length)와 주요 하이퍼파라미터가 본 연구의 10프레임 단위 구성에 최적화되어 있기 때문으로 해석된다.

V. Conclusion

본 연구는 군중 상황에서 이상 이동 인식을 위한 학습 데이터셋을 구성하고, 이를 바탕으로 다양한 데이터 활용

전략을 세워 각 전략의 효과를 체계적으로 분석하고자 하였다. 이를 위해 이상 이동 분류라는 목적에 맞게 기존 연구의 클래스 기준을 보완하고, 해당 기준에 부합하는 학습 데이터를 190개로 확대하여 구축하였다. 또한, 시퀀스 수, 분할 방식 등의 데이터 구성 조건을 조합하여 총 여섯 가지 사용 전략을 설계하고, 각 요소가 모델 성능에 미치는 영향을 정량적으로 비교하였다. 실험 결과, 시퀀스 수 증가에 따른 정확도 향상의 증강 적용의 긍정적 효과가 확인되었으며, 고정된 시퀀스 분할 방식은 높은 성능을 보였으나 과적합 가능성을 내포하고 있어 해석에 주의가 필요한 것으로 나타났다. 또한, optical flow 방식에 따라 성능 차이가 존재함도 관찰되었다. 이러한 결과는 데이터 구성 전략이 단순한 전처리 절차를 넘어, 모델 성능과 평가 신뢰도에 중대한 영향을 미치는 핵심 요소임을 보여준다. 본 연구에서 제시한 여섯 가지 데이터 활용 전략은 군중 이상 행동 감지 모델 개발에 있어 목적과 상황에 따라 유연하게 적용할 수 있는 실용적인 기준을 제공할 수 있을 것이다. 향후 연구에서는 사용한 하이퍼파라미터 이외에 대한 추가적인 최적화 및 LSTM 타임 스텝 크기 등에 따른 설계의 확장에 방향을 두고 학습 및 추론 시간 등 효율성까지 고려하여 진행할 예정이다. 또한 본 연구에서 제안한 데이터를 기반으로 optical flow에 다른 입력 스트림을 추가하여 이상 이동 감지 모델의 성능을 발전시키려고 한다.

ACKNOWLEDGEMENT

“This research was supported by the MSIT (Ministry of Science, ICT), Korea, under the National Program for Excellence in SW, supervised by the IITP (Institute of Information & communications Technology Planning & Evaluation) in 2024”(2022-0-01127).

REFERENCES

- [1] M. Szczodrak, J. Kotus, K. Kopaczewski, K. Lopatka, A. Czyzewski, and H. Krawczyk, “Behavior analysis and dynamic crowd management in video surveillance system,” in Proc. 22nd Int. Workshop Database Expert Syst. Appl. (DEXA), Toulouse, France, 2011, pp. 371-375, doi:10.1109/DEXA.2011.16.
- [2] J. M. Grant and P. J. Flynn, “Crowd scene understanding from video: A survey,” ACM Trans. Multimedia Comput. Commun.

- Appl., vol. 13, no. 2, Art. no. 19, pp. 1-23, Mar. 2017, doi: 10.1145/3052930.
- [3] X. Zhang, Q. Yu, and H. Yu, "Physics inspired methods for crowd video surveillance and analysis: A survey," *IEEE Access*, vol. 6, pp. 66816-66830, Oct. 2018, doi: 10.1109/ACCESS.2018.2878733.
- [4] A. Tomar, S. Kumar, and B. Pant, "Crowd analysis in video surveillance: A review," *Proc. 2022 Int. Conf. Decision Aid Sci. Appl. (DASA)*, pp. 162-168, Chiangrai, Thailand, Mar. 2022, doi: 10.1109/DASA54658.2022.9765008.
- [5] B. Tyagi, S. Nigam, and R. A. Singh, "A review of deep learning techniques for crowd behavior analysis," *Arch. Computat. Methods Eng.*, vol. 29, pp. 5427-5455, Nov. 2022, doi:10.1007/s11831-022-09772-1.
- [6] N. Thakur, P. Nagraath, R. Jain, et al., "Autonomous pedestrian detection for crowd surveillance using deep learning framework," *Soft Comput.*, vol. 27, pp. 9383-9399, Jul. 2023, doi: 10.1007/s00500-023-08289-4.
- [7] R. Jiang et al., "DeepCrowd: A deep model for large-scale citywide crowd density and flow prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 1, pp. 276-290, Jan. 2023, doi:10.1109/TKDE.2021.3077056.
- [8] F. Abdullah et al., "Context aware crowd tracking and anomaly detection via deep learning and social force model," *IEEE Access*, vol. 11, pp. 75884-75898, Jul. 2023, doi:10.1109/ACCESS.2023.3293537.
- [9] C. Dupont, L. Tobias, and B. Luvion, "Crowd-11: A dataset for fine grained crowd behaviour analysis," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, pp. 9-16, Honolulu, HI, USA, Jul. 2017.
- [10] J.-H. Choi, J.-H. Kim, A. Nasridinov, and Y.-S. Kim, "Three-dimensional atrous inception module for crowd behavior classification," *Sci. Rep.*, vol. 14, Art. no. 14390, Jun. 2024, doi: 10.1038/s41598-024-74309-5.
- [11] J. B. Abdurasulov, "Using trajectory and motion information with multilayer LSTM to detect abnormal movements in crowd situations," M.S. thesis, Dept. Electrical and Computer Engineering, Inha Univ., Incheon, Korea, Feb. 2025.
- [12] M. Bendali-Braham, J. Weber, G. Forestier, L. Idoumghar, and P.-A. Muller, "Transfer learning for the classification of video-recorded crowd movements," in *Proc. 11th Int. Symp. Image Signal Process. Anal. (ISPA)*, Dubrovnik, Croatia, Sep. 2019, pp. 271-276, doi:10.1109/ISPA.2019.8868704.
- [13] Y. Zuo, A. Hamrouni, H. Ghazzai, and Y. Massoud, "V3Trans-Crowd: A video-based visual transformer for crowd management monitoring," in *Proc. 2023 IEEE Int. Conf. Smart Mobility (SM)*, Thuwal, Saudi Arabia, Mar. 2023, pp. 154-159, doi:10.1109/SM57895.2023.10112514.
- [14] M. Elmezain, A. S. Maklad, M. Alwateer, M. Farsi, and H. M. Ibrahim, "Analyzing crowd behavior in highly dense crowd videos using 3D ConvNet and multi-SVM," *Electronics*, vol. 13, Art. no. 4925, Dec. 2024, doi:10.3390/electronics13244925.
- [15] Z. Teed and J. Deng, "RAFT: Recurrent all-pairs field transforms for optical flow," *Proc. 16th European Conf. Computer Vision (ECCV)*, pp. 402-419, Glasgow, UK, Aug. 2020, doi:10.1007/978-3-030-58536-5_24
- [16] Q. Dong and Y. Fu, "MemFlow: Optical flow estimation and prediction with memory," *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 19068-19078, Seattle, USA, Jun. 2024, doi: 10.1109/CVPR52733.2024.01804.
- [17] Pond5, "Royalty-Free Stock Footage & Music," Pond5. [Online]. Available: <https://www.pond5.com>
- [18] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 fps in fixed-camera videos," *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2720-2727, Portland, USA, Jun. 2013, DOI: 10.1109/CVPR.2013.352
- [19] Getty Images, "Stock Photos, Royalty-Free Pictures, Illustrations & Videos," Getty Images. [Online]. Available: <http://www.gettyimages.com/>

Authors



Da-Jeong Seo is a undergraduate student of Department of Artificial Intelligence at Inha University. As a member of KMS Lab, she is doing researches on computer visions and big-data analysis for analyzing crowd behaviors.



Yoo-Sung Kim received the B.S. degree in Computer Science from Inha University, Korea, in 1986, his M.S. and Ph.D. degrees in Computer Science from Korea Advanced Institute of Science and Technology(KAIST),

Korea, in 1988, and 1992, respectively. Dr. Kim joined the faculty of the Department of Artificial Intelligence at Inha University, Incheon, Korea, in 1992. He is interested in Data Intelligence, Machine Learning, and Intelligent Software Systems.