

# Pseudo-3D Trajectories: An Effective Approach for Motion Representation in Depth Data

Chien-Quang LE

*The Graduate University for Advanced Studies*

Duy-Dinh LE

*National Institute of Informatics*

Shin'ichi Satoh

*National Institute of Informatics*

---

## Abstract

Leveraging the motion information of trajectories shows the effectiveness to the human action recognition in 2D video. However, the issue is that this approach direction is effective or not when represents motions in 3D video is not still answered. In this paper, we will deal with this issue by conducting experiments based on 2D trajectory features to present motion information from one 3D video representation. Beside, in order to ensure including depth information, we propose a method based on compensating motion information from other representations. Evaluated on the benchmark datasets, our method significantly outperforms the 3D SoA methods.

*Keywords:* **Trajectory**, action recognition, depth, feature representation

---

## 1. Introduction

*Background and Challenges.* Gần đây, với sự phát triển của RGB-D camera như Kinect, depth data đã mở ra nhiều hướng nghiên cứu tiềm năng cho bài toán Human Action Recognition. So sánh với intensity images thông thường, depth maps hỗ trợ nhiều advantages hơn. Ví dụ, depth maps cung cấp các thông tin về shape rõ ràng hơn so với intensity images. Hơn thế nữa, depth data ít bị ảnh hưởng bởi những thay đổi của ánh sáng. Tuy nhiên, các phương pháp dựa trên intensity liệu có hiệu quả trên depth data hay không vẫn chưa được quan tâm nhiều.

*Existing approaches and drawbacks.* Trong bài toán action recognition, để adapt các phương pháp dựa trên intensity cho depth data có 2 yếu tố chính. Thứ nhất, để capture motion information hiệu quả việc chọn lựa a robust feature representation là rất quan trọng. Thứ hai, để đảm bảo motion là đầy đủ thông tin trong depth video, việc bổ sung thông tin depth vào feature representation là yêu cầu không thể thiếu. Tuy nhiên, các phương pháp được đề xuất gần đây vẫn chưa hội tụ đủ 2 yếu tố này. Một số phương pháp như [DMM-HOG, DSTIP-DCSF] xem xét depth value như là intensity value và adapt các intensity-based techniques. Mặc dù, chúng có thể đạt được những kết quả hợp lý, nhưng tất cả chúng đều phải đối mặt với nhiều hạn chế. [DMM-HOG] có thể tận dụng thông tin depth từ các phép chiếu của depth maps. Nhưng its feature representation dựa trên global motion như HOG sẽ dễ gây nhầm lẫn bởi những similar postures. [DSTIP-DCSF] có thể đảm bảo depth information trong việc tính toán features. Nhưng cách tiếp cận này không đảm bảo được sự tin cậy khi extract các local points, do bởi textureless data and depth noise. Ngoài hướng tiếp cận trên, các phương pháp như [LOP, HON4D] chỉ tập trung khai thác depth information nên không tận dụng được sức mạnh của các intensity-based features. Do đó, hướng nghiên cứu của chúng tôi là propose một phương pháp có thể đáp ứng đầy đủ cả 2 yếu tố nêu trên.

*Proposal, Idea and Steps.* Trong bài báo này, chúng tôi sử dụng một feature  
30 representation dựa trên dense trajectories của [Heng Wang], do bởi hiệu quả  
của cách tiếp cận này trong nhiều bài toán, including activity recognition and  
multimedia event detection. Các trajectories thu được bằng cách tracking các  
sampled points densely sử dụng optical flow fields. Sau khi extract trajecto-  
ries, các trajectory-aligned descriptors sẽ được adopted. Sau đó, features tính  
35 toán được từ các descriptors này sẽ được sử dụng cho việc biểu diễn motion  
information trong video.

Tuy nhiên, việc thiếu sót depth information trong feature representation có  
thể gây ra các trường hợp bị confused, như được chỉ ra trong Figure 1a. Do  
đó, để đảm bảo việc không bỏ sót thông tin depth, ý tưởng cơ bản là combine  
40 thông tin chuyển động từ nhiều góc nhìn khác nhau. Các biểu diễn từ nhiều góc  
nhìn có thể đạt được bằng cách chiếu depth maps lên trên các mặt phẳng tương  
ứng. Việc chiếu này dễ dàng thực hiện được bởi những thuận lợi mà depth data  
mang lại.

*Experiments and Results.* Chúng tôi tiến hành các experiments trên challenging  
45 benchmark datasets, các kết quả thí nghiệm chỉ ra rằng phương pháp của chúng  
tôi đánh bại the SoA methods trên depth data. Các kết quả này đã cho thấy  
những contributions của our method: (1) We propose an adaptive method for  
3D video representation by using 2D features. (2) We thực hiện comprehensive  
experiments on the state-of-the-art MSR Action 3D dataset and show that our  
50 method is the best when compared with the state-of-the-art 3D methods.

*Paper structure.* After a brief review of the related work in Section 2, the pro-  
posed method is described in Section 3. Section 4 presents the experimental  
results and their concerned discussions. The summaries of our work are given  
in Section 5.

## 55 2. Related Works

Tìm hiểu các thành phần của một hệ thống HAR hiện là một trong những hướng nghiên cứu quan trọng của CV. Feature representation là 1 trong số các thành phần thu hút được sự chú ý của cộng đồng nghiên cứu.

*Works trích chọn features từ depth data.* - Hướng xem depth value như intensity  
60 value. - Hướng sử dụng real depth value và skeleton information.

*Điểm khác biệt của phương pháp hiện tại với các phương pháp trước.* - Hướng sử dụng 2d trajectories cho 3d data.

*Works in combining many types of features and sự khác biệt với our work.* -  
Works gần đây sử dụng early fusion sheme. - Our method sử dụng late fusion  
65 sheme.

## 3. Proposed Method

### 3.1. Dense trajectories

- Giới thiệu khái quát về dense trajectories - Giới thiệu sơ lược về các descriptors: HOG, HOF, MBH

### 70 3.2. Pseudo-3D trajectory-based approach for motion feature

- Hình - Illustration of proposed method - Hình - Framework overview for our system - Step 1 - Project depth maps onto orthogonal planes - Step 2 - Extract dense trajectories from scaled videos - Step 3 - Fuse motion information from different views

## 75 4. Experimental Settings

### 4.1. Dataset

- Giới thiệu về MSR Action 3D dataset - Table - 20 actions - Table - 3 subsets
- Mô tả đặc điểm của 3 subsets - Settings trong thí nghiệm

### 4.2. Evaluation Method

- 80
- Sử dụng thư viện online để extract DT - Mô tả các settings trong framework: Number of codewords, assignment method, BoW model, SVM ở 2 pha phân lớp - Thư viện libsvm - Đánh giá dựa trên accuracy

## 5. Experimental Results

### 5.1. Recognize actions from single-view

- 85
- Mô tả thí nghiệm trên 1 view - Table - Kết quả trên front view - Nhận xét các trường hợp bị confused, lý giải dựa trên trajectory shape - Figure - Ví dụ về các trường hợp bị confused

### 5.2. Fuse motion information from all views

- 90
- Mô tả thí nghiệm trên 3 view - Table - So sánh với SoA - Đánh giá lại các trường hợp bị confused - Figure - Ví dụ về trajectories trên các views khác

### 5.3. The impact of our method on descriptors

- Mô tả thí nghiệm trên 3 descriptors: HOG, HOF, MBH - Nhận xét trước khi fusion - Nhận xét sau khi fusion: + Compensate information + Keep salient information

95 **6. Conclusions**

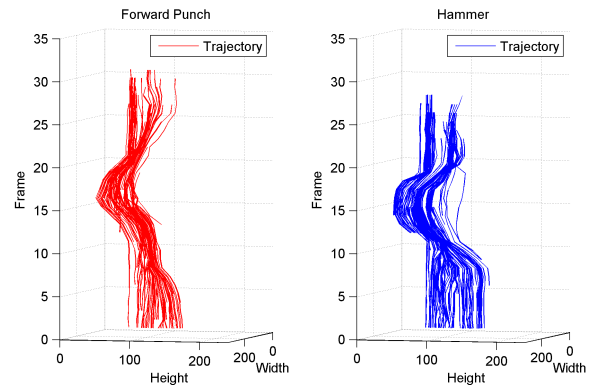
*Tóm tắt our work.*

*Đề xuất cho future work.*

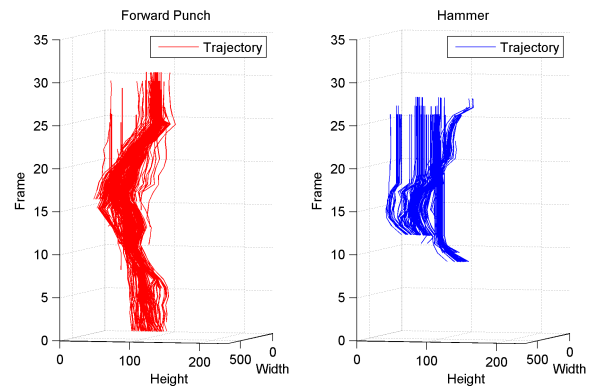
**References**

*Cách cite ref.* Here are two sample references: [1, 2].

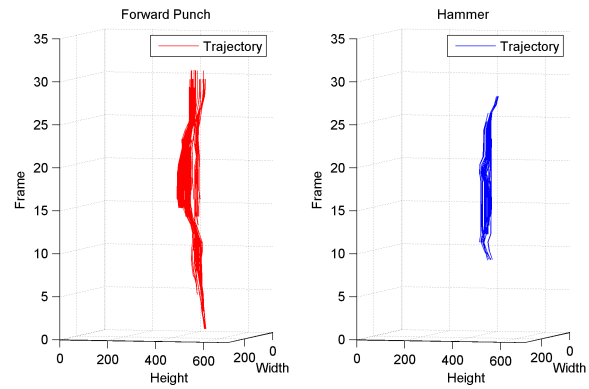
- 100 [1] R. Feynman, F. Vernon Jr., The theory of a general quantum system interacting with a linear dissipative system, *Annals of Physics* 24 (1963) 118–173. doi:10.1016/0003-4916(63)90068-X.
- [2] P. Dirac, The lorentz transformation and absolute time, *Physica* 19 (1–12) (1953) 888–896. doi:10.1016/S0031-8914(53)80099-6.



(a) From front view



(b) From side view



(c) From top view

Hình 1: Minh họa sự tương tự giữa phần lớn các Trajectories của 2 actions: Forward Punch & Hammer.