

差分ステレオを用いた集団トラッキング手法の提案

○ 星川佑磨[†], 橋本優希[†], モロ・アレサンドロ[†], 寺林賢司[†], 梅田和昇[†]

○ Yuma HOSHIKAWA[†], Yuki HASHIMOTO[†], Alessandro MORO[†],
Kenji TERABAYASHI and Kazunori UMEDA[†]

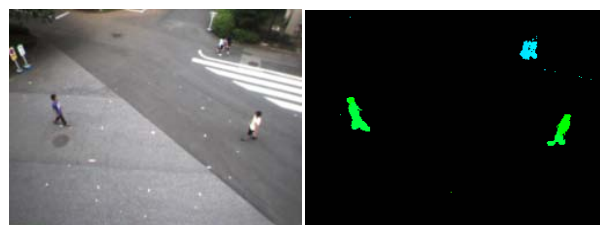
[†]: 中央大学/JST CREST, hoshika@sensor.mech.chuo-u.ac.jp, umeda@mech.chuo-u.ac.jp

我々は、背景差分により検出した前景領域に対しステレオマッチングを行う差分ステレオ法を提案している[1]. 本研究では、差分ステレオ法と KLT 特徴点追跡法から得られる三次元での特徴点の集合を、カルマンフィルタを用いトラッキングし、同一のトラッカーに属する観測点群を集団として定義する手法を提案する.
＜キーワード＞ 三次元計測, トラッキング, 人物の集団

1. 序論

近年、様々な環境下においてカメラを用いて人物の計測を行う研究が盛んに行われている[2]. この環境の中には、人物が集団を多数形成するようなものもある. 例えば、人が多数歩いている街中や、集団登校をする小学校の通学路などが挙げられる. このような環境では、形成された集団ごとの行動を計測することで、その場で起きていることをシステムが理解することが期待できる. そのため、個々の集団を認識し、トラッキングする技術が不可欠となる.

人物を様々な環境下にて計測するに当たり、多くの人物のトラッキング手法が提案されている. Rodriguez らは Correlated Topic Model を用いて環境モデルを作成することで、非常に混み合っていて人物が一様な方向に動いていないシーンにおいて多人数のトラッキングを実現した[3]. 杉村らは、KLT 特徴点追跡法から得られる人物が歩行する際に生じる画像上での周期的な上下運動と、色情報を用いた多人数トラッキング手法を提案し、非常に混み合った環境下において安定したトラッキングを実現した[4]. また近年では、簡易に三次元計測ができるステレオカメラを用いた人物の計測、トラッキングに関する研究も行われている[5, 6, 7]. 中でも、ステレオマッチングをロバスト化、高速化する差分ステレオという手法が提案されており、本手法とカルマンフィルタ[8]を用いた人物トラッキング手法が提案されている[9]. しかし、これらの研究では、三次元情報をを用いた集団のトラッキングや、具体的な集団の定義は行っていない. そこで本研究では、Gennari らにより提案された二次元での集団のトラ



(a) Input Image (b) Output Disparity Image

Fig.1 Result Obtained by Subtraction Stereo

ッキング手法[10]を応用し、差分ステレオにより得られる三次元情報と組み合わせることで集団を三次元でトラッキングし、トラッキングする観測点群の位置関係により集団を定義する手法を提案する.

2. 差分ステレオ

差分ステレオでは、まず左右の画像それぞれにおいて、背景差分により前景領域を抽出し、対応点探索の領域を前景に限定しマッチングを行うことで視差画像を得る. 本手法で得られる視差画像から、検出物体の三次元座標と幅、高さを取得することが出来る. 得られる視差画像の例を Fig.1 に示す.

3. 集団のトラッキング手法

3.1 本手法の概要

本手法では、KLT 特徴点追跡法[11, 12]により得られる人物の特徴点の三次元位置と速度を、カルマンフィルタをモデルとするトラッカーを用いトラッキングする. そして同一トラッカーに対応づけられた観測点の集合を集団として定義する.

本手法の流れを Fig.2 に示す. まず、差分ステレオにより得られた前景領域に対し、KLT 特徴点追跡

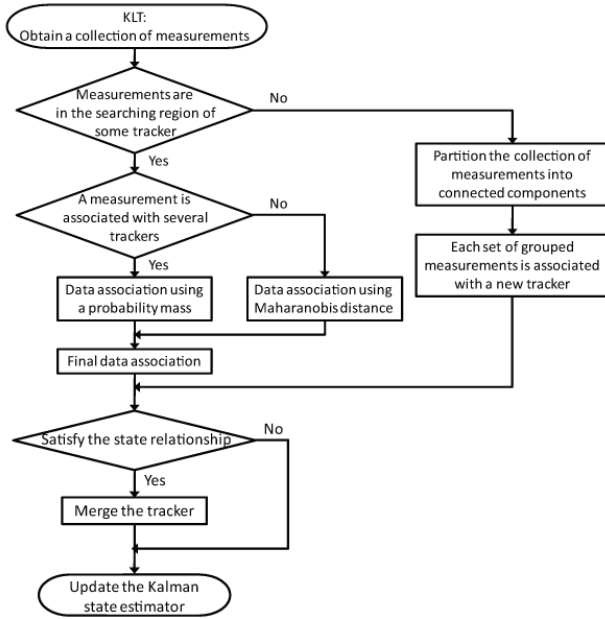


Fig.2 Flow of the Group Tracking Method

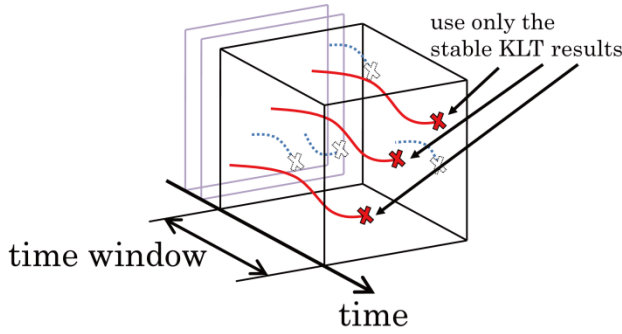


Fig.3 Example of the KLT Result

法を行う。そして、差分ステレオにより計測される三次元情報と組み合わせることで、特徴点の三次元座標を取得する。ここで、Fig.3に示すようにKLTの結果を時系列で見る。Time Window内で対応付けが出来ず消失する特徴点は除去し、軌跡がTime Windowを満たす特徴点のみを観測点として用いる。

次に、観測点群の位置関係の初期条件を用いて、観測点群の初期グルーピングを行う。そして、この観測点の集合にトラッカーを割り振り、トラッキングを開始する。初期グルーピング以降は、毎フレーム得られた各観測点をトラッカーと対応付ける。対応付けには、トラッカーからの位置に対するマハラノビス距離の二乗の値を用いる。また、集団同士の交差など、複数の集団が近づく際には、複数のトラッカーと対応付けされる観測点が検出されることがある。このような観測点は、対応付けられた各トラッカーに対する位置と速度を考慮した対応付けの確率を計算し、この確率が最も高いトラッカーに対応

付ける。最後に、各トラッカー間の位置、速度の状態の類似性を用い、集団同士の結合を判定する。

3.2 トラッカーの状態モデル

トラッカーに用いるカルマンフィルタは、等速直線運動を運動モデルとしたものとする。状態変数 $\mathbf{x}$ は、カメラ座標系における対象の三次元位置 (x, y, z) 、および速度 $(\dot{x}, \dot{y}, \dot{z})$ を用い、次式のように表わされる。

$$\mathbf{x} = [x \ \dot{x} \ y \ \dot{y} \ z \ \dot{z}]^T \quad (1)$$

また、状態方程式は次式で与えられる。

$$\mathbf{x}_{k+1} = \Phi \mathbf{x}_k + \Gamma_k \boldsymbol{\omega}_k \quad (2)$$

Φ は状態遷移行列であり

$$\Phi = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

で与えられる。 $\boldsymbol{\omega}_k$ はプロセスノイズである。観測変数には、

$$\mathbf{z} = [u \ v \ d]^T \quad (4)$$

を用いる。 u, v は画像上の座標、 d は視差である。観測変数と状態変数の関係は次式で表わされる。

$$\mathbf{z}_k = \mathbf{f}(\mathbf{x}_k) + \mathbf{v}_k \quad (5)$$

$$\mathbf{f}(\mathbf{x}_k) = \begin{bmatrix} x_k \cdot f & y_k \cdot f & b \cdot f \\ z_k & z_k & z_k \end{bmatrix}^T \quad (6)$$

f はカメラの焦点距離、 b は基線長、 $\mathbf{v}_k$ は観測ノイズである。

時刻 k におけるトラッカーの定義する集団の大きさ、速度のばらつき具合を表す共分散行列 $\boldsymbol{\Sigma}(k)$ の更新には、毎フレーム観測される観測点群のばらつき具合を反映するために、以下の式(7)を用いる。

$$\boldsymbol{\Sigma}(k+1) = \alpha_s \cdot \boldsymbol{\Sigma}(k) + (1 - \alpha_s) \cdot \text{cov}(\mathbf{C}_y) \quad (7)$$

ここに $\mathbf{C}_y$ は、観測点群の状態を表す共分散行列を表し、 α_s は重み定数を表す。以上の式からトラッカーの状態を更新する。

3.3 各観測点とトラッカーの対応付け

KLT 特徴点追跡法により得られた各観測点を、動作している各トラッカーと対応付けることで集団を定義、トラッキングする。ここで用いる各観測点は、

文献[10]では画像上での人物の二次元での位置だが、本手法では全て三次元での位置を用いる。対応付けには以下の式(8)を用いる。

$$SR_{\xi} = \{y_i | (y_i - \xi_j)^T \cdot \Sigma^{-1} \cdot (y_i - \xi_j) < \gamma\} \quad (8)$$

ここで、 y_i は観測点の三次元位置、 ξ_j は前フレームの状態から推定されたトラックの三次元平均位置、 Σ はトラックの位置に関する共分散行列を表す。式(8) γ のを満たすトラックに観測点を対応付ける。なお、 γ には実験的に求めた最適な値を用いる。

集団同士が近づく場合には、ある観測点に対し、式(8)を満たすトラックが複数存在することがある。 i 番目の観測値と j 番目のトラックの対応付けの組み合わせを θ_{ij} と表し、上記のような場合の対応付けを考える際には、以下の式(9)、(10)、(11)式を用い、集団の大きさと速度を確率的に考慮する。

$$m_p(\theta_{ij}) = k_p \cdot p(y_i | \xi_j, P_{\xi_j}) \quad (9)$$

$$m_v(\theta_{ij}) = k_v \cdot p(\lambda_i | v_j, P_{v_j}) \quad (10)$$

$$m_{Total}(\theta_{ij}) = m_p(\theta_{ij}) \cdot m_v(\theta_{ij}) \quad (11)$$

ここで、 y_i 、 λ_i はそれぞれ i 番目の観測点の位置と速度を表す。また、 ξ_j 、 v_j はそれぞれ j 番目のトラックの平均位置と速度を表し、 P_{ξ_j} と P_{v_j} はそれぞれに対応する共分散行列とする。 k_p 、 k_v はそれぞれ重み定数を表す。この(11)式の確率が最大となるトラックが、観測値と対応付けられる。

以上の方法より同一トラックに対応付けられた観測点群の平均位置を求め、これを観測値としてトラックの状態を更新する。また、同一トラックに対応付けられた観測点の数が閾値以下だった場合、トラックは観測点を用いずに状態を更新する。

3.4 観測点群の初期グルーピング

どのトラックにも属さない観測点群をグルーピングする。グルーピングには、各観測点群の位置関係を用いる。各観測点の位置関係は、以下の条件式により表される。

$$y_i R_0 y_j \Leftrightarrow SR_{y_i} \cap SR_{y_j} \neq \emptyset \quad (12)$$

ここに、 SR_{y_i} は観測値 y_i の周りに別の観測値があるかを探索探索領域を表す。式(12)の条件を満たす観測点群をグルーピングし、観測点の数が設定した閾値以上ならばこれを集団の初期状態とする。そしてこの観測点群の平均の位置と速度、共分散行列を求める。ここで求めた各成分を初期値としてトラックに与え、トラッキングを開始する。

3.5 集団同士の統合

複数の集団が統合し、新しい集団を形成する場合を考える。本手法では、トラック間の状態の類似性を用いてトラックの統合を行う。トラックの類似性には、各トラック間の位置と速度に関するマハラノビス距離を用いる。マハラノビス距離が以下の式(13)を満たす場合、トラックの統合を行う。

$$x_i R x_j \Leftrightarrow MR_{x_i} \cap MR_{x_j} \neq \emptyset \quad (13)$$

ここに、 x_i は i 番目のトラックの状態を、 MR_{x_i} は i 番目のトラックの共分散行列を考慮した際のマハラノビス距離での探索範囲を表す。この式(13)を満たすトラックの集合を I_n 、その要素（各トラック）を G_i 、この各要素を統合してできる新たな集団を $G_{m,j}$ と表記する。この集合 I_n に関し、以下に示す式(14)、(15)を用いて各要素 G_i を統合し、新たな状態を与える。

$$x_{m,j} = \sum_{G_i \in I_n} \overline{N_{i,j}} \cdot x_i \quad (14)$$

$$\Sigma_{m,j} = \sum_{G_i \in I_n} [\overline{N_{i,j}} \cdot \{ \Sigma_i + (x_i - x_{m,j}) \cdot (x_i - x_{m,j})^T \}] \quad (15)$$

ここに、 $\overline{N_{i,j}} = N_i / \sum_{G_i \in I_n} N_i$ であり、 N_i 、 Σ_i はそれぞれ i 番目のトラックが含む観測値の数と共分散行列を表す。

4. 実験

4.1 集団の交差実験

集団が画像中で横方向に接近し交差するシーンでは、集団に部分的なオクルージョンが生じ観測点が少なくなる。また、集団同士が接近するため、観測点のトラックへの対応付けが不安定になる問題がある。このようなシーンに対し、本手法の有効性を検証するため被験者実験を行った。

ステレオカメラは Bumblebee2(Point Grey Research, カラー, $f=3.8\text{mm}$, $b=120\text{mm}$, VGA, 48fps)を使用した。カメラの位置姿勢は地面から高さ 5.1[m]、仰角 40°とした。トラックに設定したプロセスノイズと観測ノイズには、実験的に求めた最適な値を設定した。また、本実験では、初期グルーピングの際の探索領域を観測点からのユークリッド距離で 1m とし、トラックは 2 つ以上の観測点に対応づけた際には観測点を用いて状態を更新する。

実験は、Fig.4(a)に示す様に、画像の右側から歩く 2 人の集団と左側から歩く 3 人の集団が画像中央付近で交差し、画像の反対側まで歩くシーンにおいて行った。実験結果を Fig.4(b)、三次元での集団の

軌跡を世界座標系で表したものを Fig.5 に示す. 実験結果において, 集団上にプロットされている点群が観測点としている KLT の結果を表し, 楕円に囲まれているのが一つの集団として認識された結果である. 結果より, 部分的なオクルージョンが生じる集団に対し, 正しくトラッキングが行えている事が分かる.

4.2 集団の統合実験

本手法を用いることにより, 集団の統合が正しく行われるかを検証するために被験者実験を行った. 実験は, Fig.6(a)に示す様に, 画像の左端から 2 人の集団が, 右端から 3 人の集団がそれぞれ画像の中央に向けて歩き, 画像の中央付近で集団が合体し一つになるようなシーンにて行った. 本実験で使用したステレオカメラ, 及びカメラの設置条件は, 4.1 節の実験と同様のカメラと設置条件とした.

実験結果を Fig.6(b)に示す. トラッキングした集団の三次元での軌跡を Fig.7 に示す. 実験結果より, 右から歩く Group0 と左から歩く Group1 が画像中央上付近で統合され, 新たな集団 Group2 を生成されていることが分かる. また, 本実験のように集団同士の挙動が似ている際には集団が統合されるが, 4.1 節の実験シーンのように, 集団同士が近づいても挙動が似ていない際には集団は統合されない.

4.3 立体環境下における集団のトラッキング実験

立体環境下でのトラッキングの検証を行うために, 階段を含むシーンにおいて被験者実験を行った. 実験では, Fig.8(a)に示す様に, 平面のフロアを歩いている集団が, 階段を上るシーンにおいて集団をトラッキングし, 三次元の集団の軌跡を取得する. ステレオカメラは 4.1 節に示したカメラと同様のものを使用し, 高さ 2.1m, 仰角 15°に設置した.

実験結果を Fig.8(b), 計測した集団の三次元での平均の位置を Fig.9 に示す. 実験結果より, 集団が三次元で正しくトラッキング出来ていることが分かる. また, 本実験で設定したような階段を含むような立体環境では, 単眼カメラでは三次元での計測データを取得するのが困難となることから, 立体環境下での本手法の有用性を示せた.

5. 結論

KLT 特徴点追跡法とカルマンフィルタをモデルとしたトラッカーを用い, 集団をトラッキング, 定義する手法を提案した. そして, トラッキングが困難な部分的なオクルージョンが生じるシーンと, 集団が統合するシーン, 単眼カメラでは計測が難しい

立体環境での被験者実験により, 本手法の有用性を検証した. 今後は, 実環境下での本手法の有用性について検証する.

参考文献

- [1] 梅田和昇, 和田正樹, 松木優介: “運動領域注視型ステレオカメラシステムの提案第 1 報: 基本コンセプトの提案”, 第 25 回日本ロボット学会学術講演予稿集, 1N31, 2007.
- [2] 橋本学: “映像監視における CV 実用技術の現状と課題”, 情報処理学会研究報告(2005-CVIM-151), Vol.2005, No.112, pp.55-62, 2005.
- [3] Mikel Rodriguez, Saad Ali, Takeo Kanade: “Tracking in Unstructured Crowded Scenes”, Proc. of the IEEE ICCV2009, pp.1389-1396, 2009.
- [4] 杉村大介, 木谷クリス真実, 岡部孝弘, 佐藤洋一, 杉本晃宏: “歩容特徴と局所的な見えを用いた特徴点軌跡クラスタリングによる混雑環境下人物追跡”, 画像の認識・理解シンポジウム(MIRU2009), OS5-2, pp135-142, 2009.
- [5] 鈴木達也, 岩崎慎介, 佐藤洋一, 杉本晃宏: “環境モデルの導入による人物追跡の安定化”, 画像の認識・理解シンポジウム MIRU2004 講演論文集, Vol.1, pp.243-248, 2004.
- [6] Bahadori, S., et al.: “Real-time people localization and tracking through fixed stereo vision”, Applied Intelligence, Vol.26, No.2, pp.83-97, 2007.
- [7] Zhao, T., et al.: “Toward a sentient environment: real-time wide area multiple human tracking with identities”, Machine Vision and Application, Vol.19, No.5, pp.301-314, 2008.
- [8] Sebastian Thrun, Wolfram Burgard, Dieter Fox, 上田隆一 訳: “確率ロボティクス”, 毎日コミュニケーションズ, 2007.
- [9] 中西達也, 寺林賢司, 橋本優希, 星川佑磨, 梅田和昇: “差分ステレオを用いた複数人物計測”, 日本機械学会ロボティクス・メカトロニクス講演会 2009 講演論文集, 1A1-D16, 2009.
- [10] G. Gennari, et al.: “Probabilistic Data Association Methods in Visual Tracking of Groups”, Proc. of the IEEE CVPR2004, VOL 2, pp.876-881, 2004.
- [11] C. Tomasi and T. Kanade: “Detection and tracking of point features”, Technical Report CMU-CS-91-132, 1991.
- [12] C. Tomasi and J. Shi: “Good features to track”, Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition, pp. 593-600, 1994.



(a) Experimental Scene



Frame #140

Frame #200

Frame #260

Frame #340

(b) Experimental Result (Two Groups Cross Each Other)

Fig.4 Experimental Result in which Two Groups Cross Each Other with Partial Occlusion

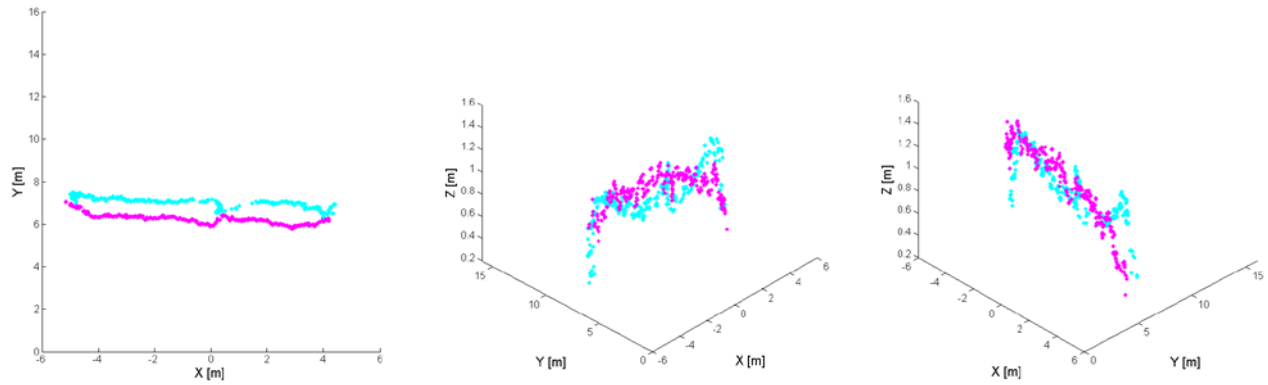


Fig.5 Measured Three Dimensional Trajectories Obtained from the Crossing Experiment
(Cyan : Group No.3, Magenta : Group No.4)



(a) Experimental Scene



Frame #100

Frame #180

Frame #250

Frame #330

(b) Experimental Result (Two Groups Merge)

Fig.6 Experimental Result in which Two Groups Merge Together

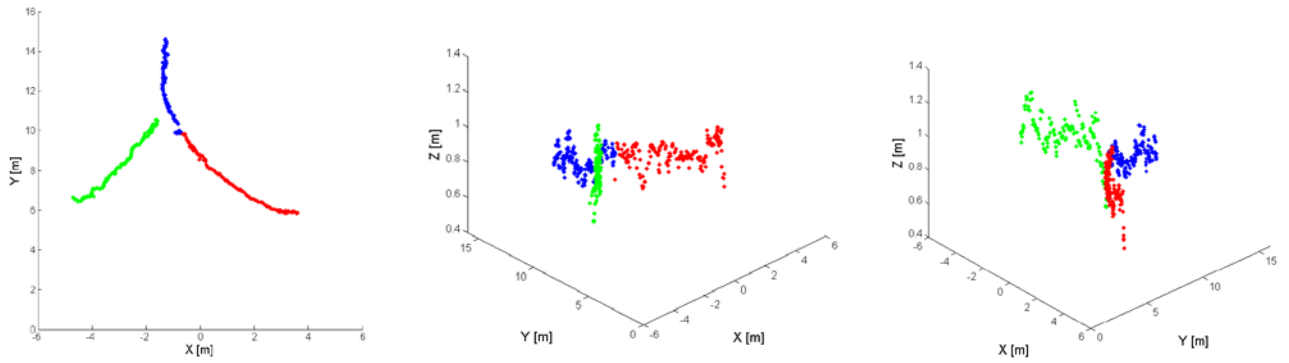


Fig.7 Measured Three Dimensional Trajectories Obtained from the Merging Experiment
(Red : Group No.0, Green : Group No.1, Blue : Merged Group No.2)

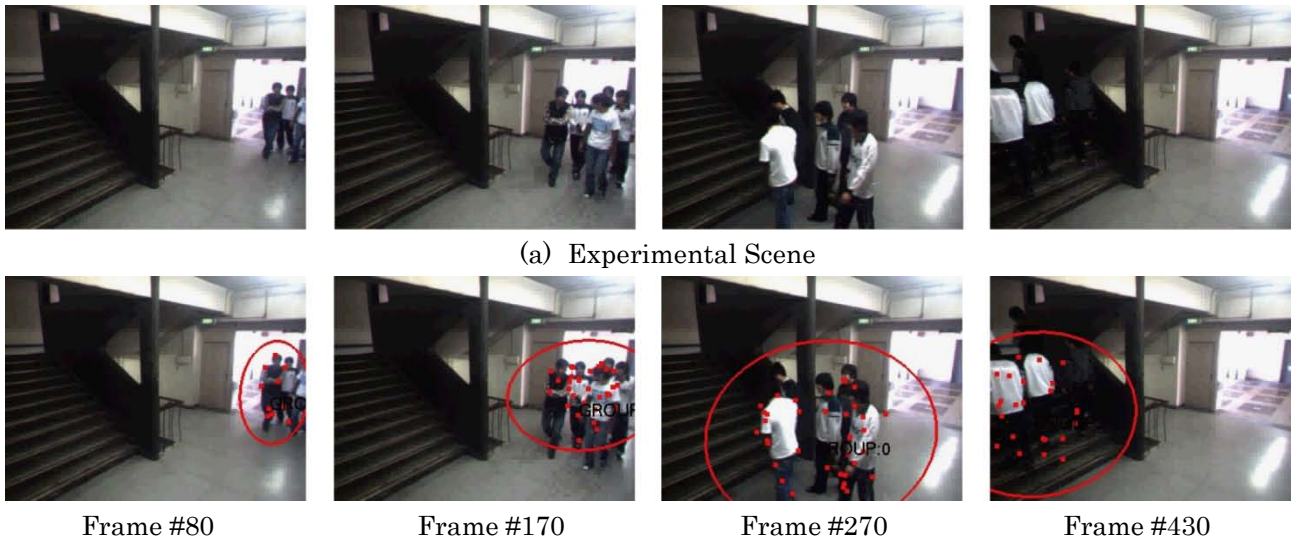


Fig.8 Experimental Result Whose Persons are Measured in Three Dimensional Environment

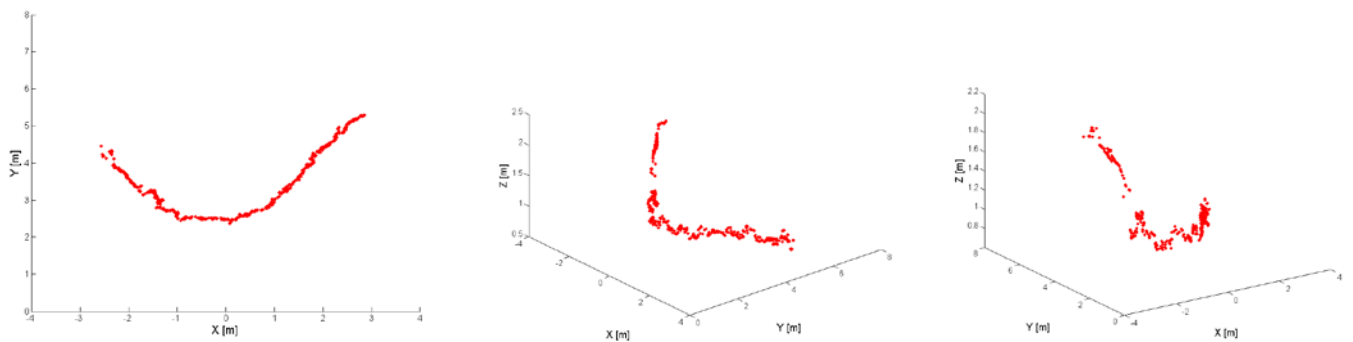


Fig.9 Measured Three Dimensional Trajectories of the Group which Goes Up the Stairs in the Scene