社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2008－UBI－18 (10)
2008／5／15

Toward Real-Time Extraction of Pedestrian Contexts with Stereo Camera

鈴木 慧[1] 高汐 一紀[2] 徳田 英幸[1,2] 和田 正樹[3] 梅田 和昇[3] 松木優介[3]

[1] 慶應義塾大学大学院政策・メディア研究科
[2] 慶應義塾大学環境情報学部 / CREST, JST
[3] 中央大学理工学部 / CREST, JST

本研究の目的は街角に設置した定点ステレオカメラから不審者や不穏状況を抽出することである．近年，ビデオカメラの普及や機械学習の進化により，ビデオカメラで撮影した映像から歩行者を検知する研究が行われている．昨年，我々は運動注視型ステレオカメラシステムを用いることで人の3次元空間的な移動を検知する手法を提案した．本稿ではステレオカメラから個々の歩行者とグループの複数コンテクストを検出するプロトタイプを実装し，その実験によりリアルタイム性における有用性を示した．本プロトタイプシステムはリアルタイムにコンテクストを解析可能であり，コンテクスト抽出の正確性と頑強性を満足するためにベイジアンモデルを適用した．

Toward Real-Time Extraction of Pedestrian Contexts with Stereo Camera

Kei Suzuki[1], Kazunori Takashio[2], Hideyuki Tokuda[1,2], Masaki Wada[3], Kazunori Umeda[3], Yusuke Matsuki[3]

[1]Graduate School of Media and Governance, Keio University
[2]Faculty of Environment and Information Studies, Keio University / CREST, JST
[3]Dept. Precision Mechanics, Faculty of Science and Engineering Chuo Univ. / CREST, JST

We extract the mood of disquiet on street corners in real-time with stereo video camera systems. In recent years, with the advance of video camera technology and the evolution of machine learning, many projects on extraction of individual pedestrian's activities from video data have been conducted. Last year we proposed a novel stereo measurement algorithm to detect moving people, which was focusing on moving region in video data. In this paper, we report our prototype of probabilistic inference engine that can detect contexts of individual pedestrian and groups of pedestrians. We applied the Bayesian Network Model to real-time context analysis and increased the accuracy and robustness of contexts extraction.

1 はじめに

近年，ビデオカメラの普及や機械学習の進化により，ビデオカメラで撮影した映像から歩行者を検知する研究が行われている [1],[2],[3]．ビデオカメラは街角や小売り店舗内などに設置され，不審者や不穏な状況を監視するために使用される．ビデオカメラデータを解析することで，そのような異常状態の自動検知を可能にし，安全・安心社会を実現できる．

しかし，既存の歩行者検知システムは異常状態を検知できるが，より高度な歩行者コンテクストをリアルタイムに抽出することはできない．例えば，検知された異常が小競り合いなのか，ひったくりなのかを判断することはできない．そのため，実世界で起こっている状況を詳細に理解できないという問題がある．実世界の具体的な異常状態に応じた，ビデオデータの記録や警備への通報を行うアプリケーションを作成することはできない．

本研究のゴールは，街角に設置した定点ステレオカメラから不審者や不穏状況を抽出することである．本研究は，歩行者検知を行うだけでなく，高次コンテクストをリアルタイムに抽出することを目標とする．運動領域注視型ステレオカメラシステムを用いることで，複数人の3次元空間的な移動を検知可能である．そのため，人の移動情報に着目し，コンテクストの抽出を行う．しかし，ビデオカメラに映る歩行者数が増加するに従い，計算量の問題から高次コンテクストのリアルタイム抽出は難しくなる．計算量を減らすために，監視エリアを絞り込み局所化し，監視エリアを優先してコンテクストを抽出する手法が考えられる．

本稿では，個々の歩行者とグループの複数コンテクストを，リアルタイムに検出するプロトタイプを作成した．本プロトタイプはベイジアンネットを用いて歩行者コンテクストを推定する．監視エリアの絞込みを行うために，プロトタイプでは歩行者の群れを仲間同士の群れか否かに分類する．仲間同士の群れは無害

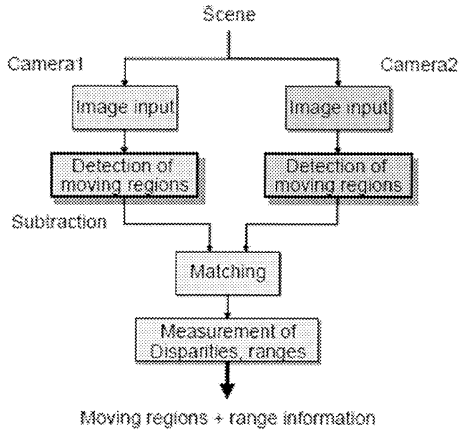


図 1: ステレオアルゴリズムの流れ

図 2: シーン例

(a) With motion detection(proposed method)

(b) With motion detection(prior method)

図 3: ステレオアルゴリズム適応例: 図 2 への適応

な群れとして認識し，群れの監視対象から除く．ステレオカメラシステムは最大 18times/sec の頻度で運動領域情報を取得する．歩行者の移動速度は徒歩で秒速 1.56m 程度なので，運動領域の取得レートは歩行者の行動を十分追跡できる．

本稿では，第 2 章において運動領域注視型ステレオカメラシステムを紹介し，第 3 章ではステレオカメラデータから歩行者コンテクストの抽出について述べる．第 4 章では歩行者コンテクストのリアルタイム解析について，第 5 章では実験と考察について述べる．第 6 章では今後の予定を述べ，第 7 章でまとめについて，第 8 章で謝辞について述べる．

2 運動領域注視型ステレオカメラシステム

本研究で利用するステレオカメラシステムの基本アルゴリズムを 図 1 に示す．通常のステレオ手法では，左右カメラの画像をマッチングすることで視差画像を得る．これに対し，本研究で提案する手法では，左右カメラそれぞれでまず運動領域を抽出し，その後に抽出された運動領域を左右画像でマッチングする．運動領域の抽出には差分処理を用いる．

図 2 のシーンに対して，我々の手法と従来のステレオ手法とで得られる視差画像を，それぞれ 図 3(a),(b) に示す．図 3(a) での運動領域抽出には単純な背景差分を用いている．図 3(a) と 図 3(b) を比べると，提案手法では視差画像の算出が運動領域のみに限定され，ステレオマッチングが運動領域に限定されることによって，左右画像の誤対応が抑制されることが期待される．また，監視用途などを考えれば，提案手法は必要十分な領域のみに対して結果を出していると言え，その後の処理が容易になると考えられる．さらに，ステレオマッチングを行う領域が限定されることから，計算コストの抑制も期待できる．

この手法は，運動している領域にしか適用できないという欠点がある反面，対応点探索を左右カメラの運動領域に大きく限定することができる，また，マッチングのための情報に元画像だけでなく運動情報も利用

可能である等の理由から，マッチングを大幅にロバスト化することができると考えている．

同アルゴリズムによるステレオカメラシステムが出力するデータは，運動領域の特徴量（距離，重心座標値，高さ，幅），タイムスタンプ，ラベル番号である．タイムスタンプはミリ秒精度で計測する．ラベル番号は，抽出された運動領域に番号付けした値である．同一ラベル番号は，同一の移動体を示す．また，運動領域の特徴量から猫なのか人なのか，大人なのか子供なのか，何名程度の人がいるのか，などの情報を得ることが容易になると期待できる．

3 ステレオカメラデータから歩行者コンテクストの抽出

本章では，最初に歩行者コンテクストについて整理し，次に高次コンテクスト抽出手法としてベイジアンネットの利用について述べる．最後に，本プロトタイプシステムで抽出した，具体的な高次コンテクストをベイズモデルに基づいて紹介する．

3.1 歩行者コンテクストの特徴

本稿では歩行者の運動領域データ (3 次元座標や幅，高さ) からコンテクスト抽出を行うため，最初に街角における歩行者の移動の特徴について具体的に述べる．知人同士はグループとして行動し，近い距離を移動する．一方，他人同士は通常は離れて行動するが，すれ違う時は肩が当たるほど，接近することもある．また，歩行者がその場にしゃがみこんだ場合，休憩を取っているのか転倒しているのか，ある時間の状態から得られるデータだけでは判別することは難しい．そのため，歩行者のコンテクスト抽出は時間経過に伴う変化を考慮し，コンテクスト抽出時には時系列データとして運動領域データを解析する必要がある．例えば，移動中に突然停止し，しゃがみこんだ事を検知した場合，転んだ可能性が高いと考えられる．本プロトタイプでは，取得するコンテクスト毎に異なる間隔で時分割した運動領域データから，歩行者コンテクスト抽出を行っている．

本プロトタイプでは，歩行者コンテクストとして歩行者個人のコンテクストと，歩行者グループのコンテクストの 2 種類に分類し，対象を絞る．単独で行動している人として，普通に歩いている人や，急いでいて小走りに歩いている人などが考えられる．集団で行動している人は，無害な群れ（仲間同士，共通の意図で行動する集団），偶発的な群れ（たまたま同一方向に移動する集団，一時的な人だかり），挙動不審な群れなどが考えられる．

3.2 ベイジアンネットによるアプローチ

歩行者の運動領域データ解析モデルとして，本研究ではベイジアンネット [5],[4] を用いる．ベイジアンネットは，グラフ構造によって結合確率分布を効率的に表現できるだけでなく，視覚的に事象間の依存関係をモデル化できる利点がある．また，従来のルールベースのモデルに比べ，確率を用いることで不確実な要因下でも妥当な推論を実行できる利点がある．ステレオカメラの運動領域データは，光源の影響により一時的に情報を取得できないことやノイズが入る可能性があるため，不確かで複雑な確率分布を扱うことのできるベイジアンネットが適している．また，歩行者モデルは時系列データで表す必用があり，複雑なデータモデルになるため視覚的に時系列モデルを理解しやすい．

歩行者の時系列データを用いてコンテクストの事後確率を求めるベイズ確率推論アルゴリズムは，大きく分けて厳密推論アルゴリズムと近似推論アルゴリズムに分類できる [7]．本研究は，複数コンテクストのリアルタイム推定を行うため，応答性に優れる近似推論を行う．近似推論アルゴリズムは，コンピュータシミュレーションによって近似的に事後確率を求めるアルゴリズムである．任意の精度で推論結果を出力できるため，リアルタイム性と精度を満足する妥当な計算時間に抑えることができる．本プロトタイプでは，代表的な近似推論アルゴリズムである，Likelihood Weighting アルゴリズム [6] による推定エンジンを用いる．

3.3 歩行者コンテクストのベイズモデル

ここでは，本研究のプロトタイプシステムで取得したコンテクストを個人で移動している歩行者と，グループで移動している歩行者に分類して述べる．本プロトタイプでは，歩行者個人のコンテクストとして，転んだ状態を抽出し，歩行者グループコンテクストとして，仲間同士のグループを抽出する．

3.3.1 歩行者個人のコンテクスト

歩行者個人コンテクストとして，”転んだ”コンテクストを抽出するためのベイジアンネットについて図 4 に示す．それぞれの楕円がコンテクスト，運動領域データを表す確率変数にそれぞれ対応する．tumble が”転んだ”コンテクストに対応し，xs1, ys1, z1 は順に歩行者個人における 1 秒間の平均 X 成分速度, Y 成分速度, Z 座標値に対応する．xs2, ys2, z2 はそれぞれ xs1, ys1, z1 の一秒前のデータに対応する．歩行者が転んだ場合，一般的に歩行者の歩行速度は急に下がり，倒れこむため z 座標値も急に小さくなる．各確率変数は，これら転んだ事象の特徴を良く表すために用いられる．これらの運動領域データに対応する確率変数は，tumble と直接的な依存関係を持ち，依存関係は二つの確率変数間を結ぶ有向リンクとして表現される．これらの依存関係は各確率変数ごとに定義された条件付き確率分布表によって重み付けられる．

3.3.2 歩行者グループや群のコンテクスト

歩行者グループコンテクストとして，”仲間同士”コンテクストを抽出するためのベイジアンネットを図 5 に示す．それぞれの楕円がコンテクスト，運動領域データを表す確率変数にそれぞれ対応する．”group”

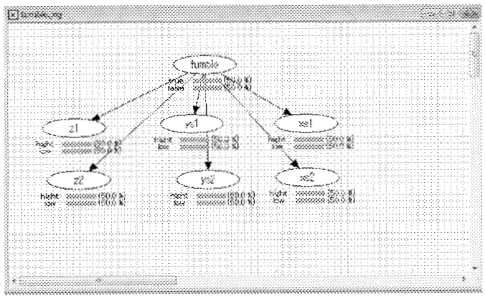


図 4: 転んだのベイズモデル

が仲間同士コンテクストに対応し，vs1, dist1, angle1 は順に 2 人の間における 1 秒間の速度ベクトルの類似度，距離，ベクトル間の狭角に対応する．vs2, dist2, angle2 はそれぞれ vs1, dist1, angle1 の一秒前のデータに対応し，vs3, dist3, angle3 はそれぞれ vs2, dist2, angle2 の 1 秒前のデータに対応する．歩行者が仲間同士でグループを作る場合，一般的に歩行者間の歩行速度と進行方向は類似し，歩行者間の距離は近い．また，基本的に隣り合って行動し，後ろや前を歩く歩行者は他人である可能性が高い．

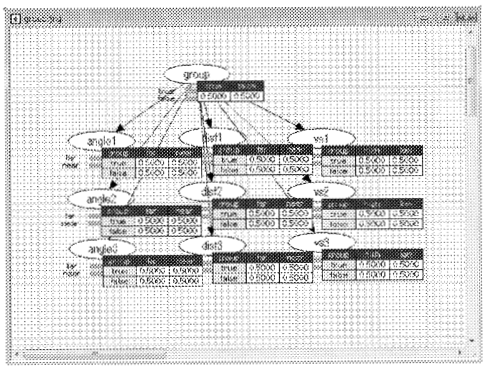


図 5: 仲間同士グループのベイズモデル

4　歩行者コンテクストのリアルタイム解析

歩行者コンテクスト抽出における全体の流れを図 6 に示す．コンテクスト抽出時には，時系列データとして運動領域データを解析する必要があるため，スライディングウィンドウ方式でステレオカメラからの運動領域データを管理した．各歩行者コンテクスト毎に，異なった時間単位で運動領域データを利用するため，バッファのスライド時には，全てのコンテクストが解析し終わったデータのみバッファをスライドさせる．ビデオカメラに映る歩行者数が増加するに従い，計算量の問題から高次コンテクストの複数リアルタイム抽出は難しくなる．計算量を減らすために，監視対象やエリアを絞り込み局所化し，優先してコンテクストを抽出する．本プロトタイプでは，仲間同士で歩いているグループは監視対象から外した．監視対象を優先するために，ラベル番号を格納する優先キューを用いた．本プロトタイプは優先キューに格納されたラベル番号が示す歩行者対象から優先してコンテクスト抽出処理を行うことができる．また，ベイズエンジンによるコンテクスト推定は，システムで最も計算処理負荷の高い部分である．ステレオカメラの運動領域取得レートを考えると，全てのデータに対してリアルタイムに複数コンテクスト推定を行うのは監視対象が増加した場合に難しくなる．そのため，本プロトタイプでは，抽出するコンテクスト毎にイベント検知を行い，あるイベントを検出した場合に限ってコンテクスト推定を行った．

本プロトタイプでは，具体的なイベントとして，転んだという歩行者個人のコンテクストでは，突然ユーザの z 座標値がある一定以上の割合で減った場合とし，仲間同士という歩行者グループコンテクストでは，歩行者間の位置座標距離がある一定以下になった場合とした．前者は，急に歩行者がしゃがみこんだ，あるいは倒れた状態を示し，後者はユーザ同士が接近している状態を示している．

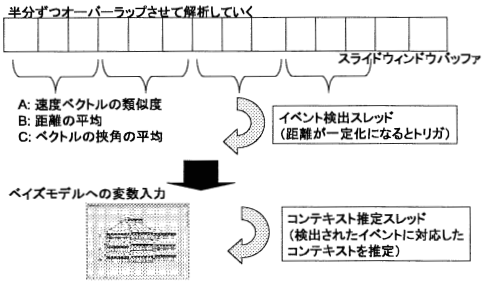


図 6: 仲間同士グループコンテクスト抽出の流れ

5　システム設計

本プロトタイプのハードウェア構成図を図 7 に、ソフトウェア構成図を図 6 に示す．ステレオカメラデータの解析とベイジアンネットによるコンテクスト推定は，どちらも処理負荷が高いため，図 7 の様に別々の PC に分けて実行させる．ステレオカメラシステムと歩行者コンテクストリアルタイム推定部の間は，ネットワーク経由で接続される．本プロトタイプでは，2 種類のワーカースレッドを用いた．歩行者コンテクスト毎のイベント検知スレッドと抽出する歩行者コンテクスト毎の推定スレッドである．イベント検知スレッドは，優先キューを参照し，登録されている歩行者の運動領域データから順にイベント発生を確認する．イベントを検知した場合，イベントに対応した歩行者コ

ンテクスト推定スレッドに推定させる．イベント検知スレッドは，運動領域データバッファを共有し利用する．推定スレッドは，ベイズエンジンを用いてコンテクスト推定を実行する役割を持つ．

実装は C++で行い，GUI に Qt Library[8] を利用した．

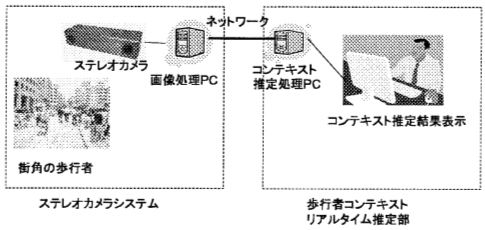


図 7: ハードウェア構成図

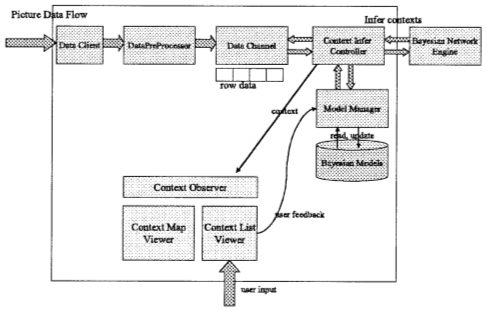
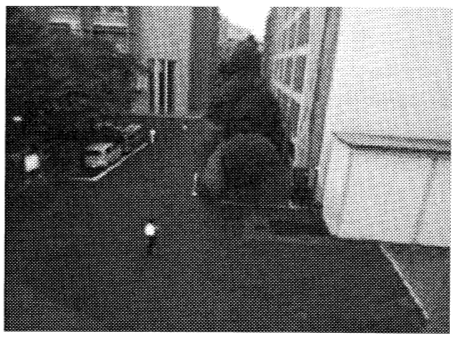
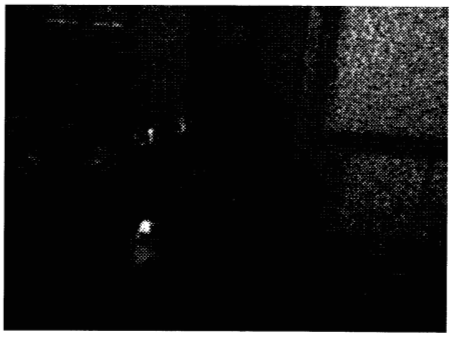


図 8: システム構成図



(a) Experimental scene



(b) Subtraction image

図 9: 屋外シーン例

6 実験と考察

ステレオカメラシステムと本プロトタイプを用いた複数歩行者コンテクストのリアルタイム抽出を行った．ステレオカメラシステムは，市販のステレオカメラを用いて，提案した基本アルゴリズムの実装を行った．カメラには，Point Grey Research 製の Bumblebee2（カラー，f=3.8mm, XGA）を用いた．実験での画像サイズは 320 × 240 画素とした．運動領域の検出には純な背景差分を用いた．ステレオカメラシステムと本プロトタイプは，ネットワーク経由でデータを送受信する．本プロトタイプの実行環境は，CPU : Core2Duo E6850(2.6GHz 2 core)，OS : Windows Vista である．

最初に，ステレオカメラシステムの実験の様子を図 9(a),9(b) 示す．この例では，人物がかなり遠くて小さく，困難なシーンであるが，人物領域の検出と距離計測が実現できている．計測精度の比較などは現段階では行っていないが，少なくとも監視用途などを考えれば，通常の距離画像よりも提案手法で得られる距離画像の方が必要十分な情報のみを含んでおり有用であると考えている．処理速度に関しては，Bumblebee2 の最高速度である 18fps をほぼ実現している．

次に，本プロトタイプの実行画面を図 10 に示す．実験時には，ステレオカメラからの運動領域データレートは 7times/sec であり，時間によってデータの欠損が起こった．実験は屋外で行ったため，太陽光などの影響によって映像データにノイズが入ったことが考えられる．ベイジアンネットはこのような不確かな確立変数のモデルに適している．また，ベイズエンジンによるコンテクスト推定は，平均 58msec 程度の計算時間を必用としたが，本プロトタイプシステムは，イベントトリガ型を用いているため，リアルタイム性を実現できた．また，実験データから，仲間同士コンテクストと転んだコンテクストを抽出でき，ベイジアンネットの適用と，定義したモデルの妥当性を確認できた．本稿では，十分に定量的評価を行っていないが，歩行者コンテクストをリアルタイムに複数同時抽出を行う有効性を確認できた．

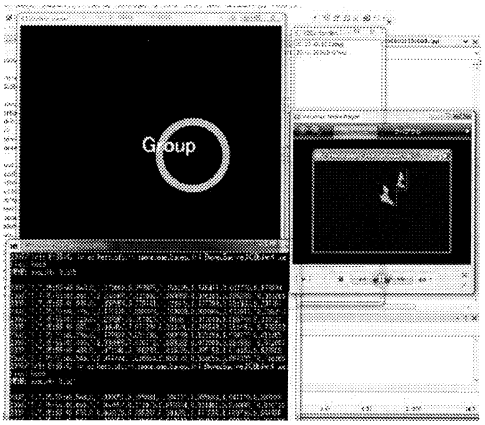


図 10: プロトタイプシステムの実行画面

7 今後の予定

今後の発展として，最初にベイジアンネットの学習による提案ベイズモデルの妥当性の検証とモデルの改善を行い，新たに抽出するコンテクストの検討を行う．次に，ベイジアンネットの学習を行い，閾値による単純なコンテクスト推定など他の手法と比較した本手法の優位性を検証する．また，現在のシステムでは歩行者コンテクスト毎にベイズモデルを記述し学習させる必要があり，抽出するコンテクストの種類が増加に従って負担が大きくなる．そのため，歩行者フローに基づいて歩行者移動モデルを自動生成するシステムが必要になる．最後に，監視対象や街角に設置する定点ステレオカメラの増加すると，コンテクスト推定の計算量だけでなく，データ量自体も現在の集約型では対応できなくなる．そのため，コンテクスト推定に特化した分散処理型のタスクシステムを考えている．

8 まとめ

運動領域注視型ステレオカメラシステムを用いた歩行者コンテクストのリアルタイム抽出について，プロトタイプ実装を行い実験を行い有効性を確認した．具体的には，歩行者のグループコンテンツとして，仲間グループ検出，個人コンテンツとして転んだことを，ベイズエンジンを利用し複数同時推定した．

9 謝辞

本研究は JST CREST の一部として行われた．

参考文献

[1] M. Bertozzi and E. Binelli and A. Broggi and M. Del Rose, Stereo Vision-based approaches for Pedestrian Detection, CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops, 2005.

[2] Z. Liang and C. Thorpe, Stereo- and Neural Network-Based Pedestrian Detection, Proceedings of the IEEE Intelligent Transportation Systems Conference, Tokyo, Japan, Spring, 1999.

[3] F. Xu and K. Fujimura, Pedestrian detection and tracking with night vision, In IEEE Intelligent Vehicles Symposium, 2002. 212.

[4] R. E. Neapolitan. Learning Bayesian Networks. Prentice Hall, 2003.

[5] S. Russel and P. Norvig. Artificiall Intelligence Modern Approach Second Edition. Prentice Hall, 2002.

[6] R. Fung, Chang, and K.-C. Weighting and integrating evidence for stochasitic simulation in bayesian networks. In Fifth Conference on Uncertainty in Artificial Intelligence, 1989.

[7] H. Guo and W. Hsu. A survey of algorithms for real-time bayesian network inference, 2002.

[8] Tolltech, http://trolltech.com/products/qt.