

ビジョン型のペットロボット・インタラクション

A Vision-based Pet Robot Interaction

三原 功雄, 山内 康晋, 土井 美和子

Isao Mihara, Yasunobu Yamauchi and Miwako Doi

(株) 東芝 研究開発センター ヒューマンインターフェースラボラトリー

TOSHIBA Corporate Research & Development Center Human Interface Laboratory

{isao.mihara, yasunobu.yamauchi, miwako.doi}@toshiba.co.jp

ABSTRACT: We developed a new vision-based interface that can detect the human motion and control a pet robot interaction system. This vision-based interface is composed of the “*Motion Processor (MP)*” and the “*Region Of Interest (ROI)*”. The *MP* is a new image input device. The *MP* can exclude background part and capture only the object’s shape, motion and depth information in real time. The *ROI* is a concept in order to distinguish objects in a captured image. For detecting the *ROI*, the proposed method used depth information of the captured image and offered a fast and robust searching method. It enables the *MP* to sense a variety of object’s shapes, motions, positions and so on. Therefore, our vision-based interface using the *MP* with the *ROI* can easily recognize object’s motions as effective computer eyes. The user can interact with the pet robot (constructed by LEGO MINDSTORMS™) using this vision-based interface.

1. はじめに

人間は日常生活では、言語によるコミュニケーションだけでなく、身振り・手振りといった動作を自然に用いて、理解の助けにしている [1]。そこで、人間の動作を認識するための様々な研究が行われてきた [2-5]。しかし、従来手法では、画像から複雑な背景部分を除去し、安定的に認識対象のみを得る(図1参照)ことは非常に困難で、また、計算コストを要していた。



背景部分の分離 動作の認識
図1: 従来手法の処理の流れ

そこで我々は、このような問題を回避すべく、新しい方法で対象物だけの画像を取得することが可能なデバイス「モーションプロセッサ (Motion Processor, 以下 MP)」を開発した [6,7]。さらに、MPで取得した画像の奥行き情報を用いることで、画像内から任意の注目領域 (Region Of Interest, 以下 ROI) を検出する手法を開発し、動作のリアルタイムセンシングを実現した。この結果、モーションプロセッサをコンピュータの目として実用化することができた。

本論文では、まずこれらについて説明し、この機構

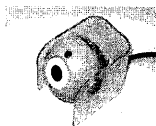


図2: モーションプロセッサ



図3: 撮像画像

を「目」として持つペットロボットとのインタラクション・システムを紹介する。

2. モーションプロセッサ (MP) とは?

図2に、MPの試作機を示す。MPは、光の反射の原理を用いて撮像画像から背景部分を除去し、対象物部分の画像のみを取り出すという新規画像処理手段を採用した画像入力装置である [6]。MPの撮像画像の例を図3に示す。手の部分のみが検出されており、背景部分は撮像されていないことが分かる。またMPは、対象物の奥行き情報も取得することができる。表1に試作機の性能を示す。

表1: MP試作機の性能

解像度	64 x 64 pixels
奥行き深度	7bit
動作レート	50 frames / sec
本体サイズ	75 x 75 x 78mm (W x H x D)
撮像距離	30-90cm, 40 square cm (at 40 cm)

3. 注目領域 (ROI) の検出 — ロボットの目としての利用

MPをコンピュータの目として用いる場合、比較的遠い距離で、広い範囲で対象物の動きなどを捉える必要がある。このような状況では、撮像範囲内に頭、左手、右手、身体などというように、複数の対象が存在する (図4参照)。このような画像から、特定の部位の動きを捉え、認識などを行うため、本論文では注目領域 (Region Of Interest, ROI) という概念を導入し、画像内から特定の領域を取得するための手法について説明する。

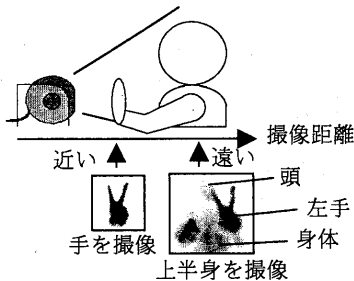


図4: 撮影距離と撮像物の関係

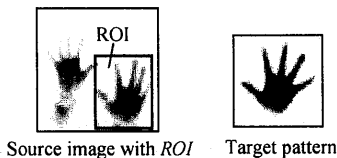


図5: 左手部分を注目領域として検出している様子

図5は、画像中の左手の部分ROIとして検出しているところである。従来法では、色情報などを用いて、特定領域の検出を行っていたが、これは環境条件に強く依存するため、ロバストではない。そこで、本手法では、MPで得られた画像の奥行き情報と物体の形状情報を用いて実現した。以降に、我々のROI検出手法を示す。(1) 特定の奥行き値(depth-level)よりも値が小さな部分を除去した画像を生成する。例えば、図6のような4つの画像を生成する。(2) それらに含まれる全ての領域を取得する。図6における R_n ($n=1, \dots, 11$) がこれにあたる。(3) その中から、検索条件に合わない領域(例えば、小さすぎる、など)を除く。 R_n ($n=1, 3, 4, 6, 7, 9$) が条件を満たすとする。(4) これらの領域のうち、Target Pattern (図5) と最も類似度の高い領域の部分

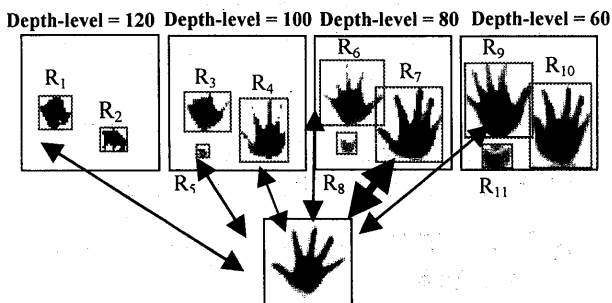


図6: Depth-level の変化に応じた領域の検出

表2: 実験結果

L2 \ L1	35 cm	40 cm	50 cm	60 cm
15 cm	48%	38%	2%	—
20 cm	93%	98%	45%	—
25 cm	97%	97%	76%	4%
30 cm	—	99%	94%	36%
35 cm	—	—	98%	98%
40 cm	—	—	97%	99%

ROIとする。この場合、 R_7 が検出された。

4. 評価実験

上半身を撮像した画像から、手の部分を検出する。この際、MPと身体との距離 $L1$ と、手と身体との距離 $L2$ とをそれぞれ変化させて検出率を測定する。表2が実験結果である。自然な動作を行う範囲では、非常に高い検出率が得られることがわかる。

5. ペットロボットとのインタラクション

図7は、LEGO MINDSTORMS™ Robotics Invention Systemで作成したロボットである。このロボットには、MPとマイクが装着してある。MPはロボットの目として機能し、ROI検出技術により、注目領域部分の動作のセンシングを行う。また、マイクから入ってきた声は、音声認識され、ロボットの耳として働く。これにより、例えば、「こっちおいで」などと言われると、人間がいる方向を探し、そちらに近づいていき止まるなどといった、ペットロボットとのインタラクションが実現される。

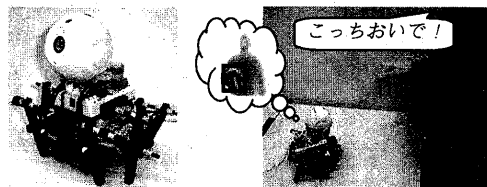


図7: MPを搭載したロボットとそのインタラクションの様子

参考文献

1. P.Ekman et al., "The repertoire of nonverbal behavior-categories, origins, usage, and coding", *Semiotic*, pp.49-98, 1969.
2. M.Kass et al., "Snakes: Active Contour Models", *International Journal of Computer Vision*, pp.321-331, 1988.
3. R.Cipolla et al., "Qualitative visual interpretation of 3D hand gestures using motion parallax", *Proc. of MVA '92*, pp.477-482, 1992.
4. T.J.Darell et al., "Space-time gestures", *Proc. of IJCAI'93*, 1993.
5. A.Wexelblat, "An Approach to Natural Gesture in Virtual Environments", *Computer-Human Interaction*, Vol.2, No.3, pp.179-200, 1995.
6. S.Numazaki et al., "A Kinetic and 3D Image Input Device", *CHI '98*, pp.237-238, 1998.
7. N.Umeki, et al., "A Motional Interface Approach Based on User's Tempo", *CHI '99*, 1999.