

RGB-D センサによる距離と濃淡画像のLBP 共起性を利用した空間識別

鄭 好政 (九州大学) マルティネス モゾス オスカル (リンカーン大学)
岩下 友美 (九州大学) 倉爪 亮 (九州大学)

1. はじめに

ヒトと共生する生活支援ロボットにとって、周囲環境の種別（台所、居間、廊下など）の識別は、状況に応じた適切なサービスの提供には必要な機能である。もしロボットが搭載した外界センサにより環境種別を正確に識別できれば、ヒトとのインタラクションにおける曖昧性を低減でき、ロボットのコミュニケーション能力の向上が期待できる [1, 2]。また、近年、RGB-D センサが安価に提供されており、これを用いれば、カラー情報および距離情報のマルチモーダルな情報が同時に取得できる。

そこで本稿では、この RGB-D センサから得られるマルチモーダル情報（カラー / 濃淡情報と距離情報）を用いた、生活支援ロボットのための環境識別手法を提案する。特に提案手法では、同一対象を同一時刻で撮影して得られるカラー / 濃淡画像と距離画像には、空間的な相関関係が存在することに着目する。この相関関係を表現するために、本手法では局所画像特徴量である Local Binary Pattern (LBP)[3] を利用し、濃淡画像と距離画像の両者から得られる LBP を 2 次元ヒストグラムで表現した新たな局所画像特徴量 Co-LBP (Co-occurrence of LBP) を提案する。この Co-LBP は、画像情報および距離情報からなるマルチモーダル情報に対し、両者の同時生起確率、すなわち共起性を表現したものである。ただし、得られた 2 次元ヒストグラムは次元数が非常に高くなることから、特異値分解により次元数を削減し、新たな特徴ベクトルを生成する。この特徴ベクトルを Support Vector Machine (SVM) により学習し、空間識別を行う。

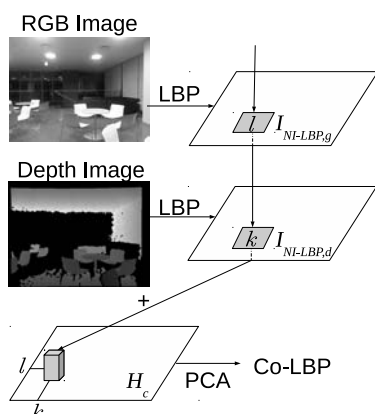


図 1 Co-LBP : 濃淡・距離画像の対する共起性を考慮した特徴記述子

2. Co-LBP

RGB-D センサで同一対象が同一時刻に撮影されると、得られたマルチモーダル情報にはお互いに空間的な相関関係が存在する。従って、それぞれの情報に個々に LBP を適用して得られた情報にも、何らかの相関が存在すると考えられる。ここで、 I_g および I_d を RGB-D センサから得られた濃淡画像および距離画像とし、これに LBP[3] を適用して LBP 画像 $I_{LBP,g}$ および $I_{LBP,d}$ を得る。

著者らは既に [4] において、濃淡画像および距離画像それぞれで得られた LBP 画像から、個別に LBP 値のヒストグラムである h_g, h_d を求め、それを単純につなぎ合わせて新たな特徴ベクトル $h_c = \{h_g, h_d\}$ を得る手法を提案している。しかしこの手法では、両モダリティの空間的な相関関係は考慮されていなかった。

そこで本稿では、両モダリティの空間的な相関関係を考慮した、新たな特徴ベクトルを提案する。提案手法の概要を図 1 に示す。まず、濃淡画像および距離画像から、それぞれ個別に LBP 画像を作成する。ここで、それぞれの画像で計算される LBP 値を $1 \sim L$ および $1 \sim K$ とする。次に、 $L \times K$ からなる 2 次元ヒストグラムを用意し、それぞれの LBP 画像で同一の画素に着目し、それぞれの値に応じて 2 次元ヒストグラムの対応するビンに投票する。すなわち、もし濃淡画像から得られた LBP 画像のある画素の値が l 、同じ位置にある距離画像の LBP 値が k であるなら、 $H_c(l, k)$ に 1 つ投票する。これを LBP 画像のすべての画素で行うことにより、濃淡・距離画像の LBP 値の 2 次元ヒストグラムが作成できる。より正確には

$$H_c(l, k) = \sum_{\forall i} \begin{cases} 1 & \text{if } I_{LBP,g}(i) = l \text{ and } I_{LBP,d}(i) = k \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

となる。ここで、 $I_{LBP,g}(i), I_{LBP,d}(i)$ は位置 i における濃淡、距離画像の LBP 値である。

次に、得られた 2 次元ヒストグラム H_c を行ごとにつなぎ合わせ、1 次元の特徴ベクトル h_c を生成する。最後に、特異値分解 (SVD) により特徴ベクトルの次元を削減することで、最終的な特徴ベクトル \tilde{h}_c を得る。この特徴ベクトル \tilde{h}_c は、マルチモーダルな LBP 画像の相関関係を表したものであり、これを Co-LBP (Co-occurrence LBP) と呼ぶ。

3. サポートベクトルマシン SVM を用いた分類

周囲環境の種別の識別を行うために、サポートベクトルマシン [5] を用いる。ただし、カーネルにはガウ

シアンカーネル (RBF) を用い、1対1方式 [6] により多クラス分類を行った。SVMの入力は Co-LBP である \tilde{h}_c であり、SVM には LIBSVM ライブラリ [7] を用いた。RBF のパラメータ C, γ の最適値は、総当たり法により $C \in [2^{-5}, \dots, 2^{15}]$ 、および $\gamma \in [2^{-12}, \dots, 2^3]$ の範囲で求めた。

4. 屋内空間での識別実験

提案する Co-LBP の性能を評価するため、公開されている屋内空間データベース [8] を用いて識別実験を行った。このデータベースは5つの異なる空間 (廊下, 台所, 研究室, 学習室, オフィス) において、RGB-D センサである Kinect を用いて撮影された多数枚のカラー画像、および距離画像からなる。今回の実験では、このうち表 1 に示す計 1228 組のカラー画像と距離画像を用いた。ただしカラー画像は濃淡画像に変換して用いた。識別性能の評価では、各種別に含まれる複数の撮影場所を、重複しないように学習データとテストデータに分け、複数の場所で撮影された学習データで学習した SVM の識別性能を、異なる場所で撮影された画像のテストデータで評価した。実験は学習データとテストデータの組み合わせを変えて 10 回行い、正答率の平均値を求めた。

表 1 屋内空間データベース (計 1228 組)

種別	場所	カラー & 距離画像の組
廊下	廊下 1	68
	廊下 2	42
	廊下 3	70
	廊下 4	99
	計	279
台所	台所 1	73
	台所 2	65
	台所 3	53
	計	191
研究室	研究室 1	99
	研究室 2	99
	研究室 3	81
	研究室 4	78
	計	357
学習室	学習室 1	71
	学習室 2	70
	学習室 3	49
	学習室 4	62
	計	252
オフィス	オフィス 1	57
	オフィス 2	45
	オフィス 3	47
	計	149

表 2 濃淡・距離画像の単純な結合による空間識別

Feature type	Dim	CCR (%)
LBP [9]	512	88.25
LBP ^{u4} [11]	398	89.14
NI-LBP [10]	512	90.02
NI-LBP ^{u4}	398	91.35
LM filter [12]	512	73.07
MR filter [13]	512	76.74
SIFT [14]	200	86.86
SIFT [14]	400	86.75
SURF [15]	200	64.16
SURF [15]	400	89.51

5. 実験結果

まず、LBP を含む様々な識別手法を実装し、上述したデータベースを用いて識別性能を評価した。表 2 に、各特徴ベクトルの次元、およびそれぞれの識別性能を示す。実験では、一般的な LBP [9] の他に、LBP を次元削減した LBP^{u4} ($U = 4$ の Short LBP) [9]、LBP の改良である NI-LBP [10] と NI-LBP^{u4} を示している。このうち、文献 [11] では、LBP^{u4} が最も高い識別率を示している。また、これらに加えて、一般的なテクスチャ記述子である Texton [12], [13]、SIFT [14]、および SURF [15] も比較した。実験の結果、表 2 に示すように NI-LBP^{u4} が最も高い識別率を示した。

次に、最も性能の高かった NI-LBP^{u4} と、本項で提案した Co-LBP (Co-NI-LBP^{u4}) に対し、SVD で削減する次元数を変化させながら性能を比較した。結果を図 2 に示す。このうち、NI-LBP^{u4} は、濃淡画像と距離画像から得られる LBP ($U = 4$ の Short LBP) を単純に結合し、SVD で次元削減を行ったものである。一方、Co-NI-LBP^{u4} は、濃淡画像と距離画像から得られる LBP ($U = 4$ の Short LBP) から 2 次元ヒストグラムを生成し、SVD で次元削減を行ったものである。実験の結果、同一の次元数まで次元削減を行った場合には、提案する Co-NI-LBP^{u4} の識別率が高く、特に Co-NI-LBP^{u4} は次元数が 25 の時に最大識別率 89.39% が得られた。

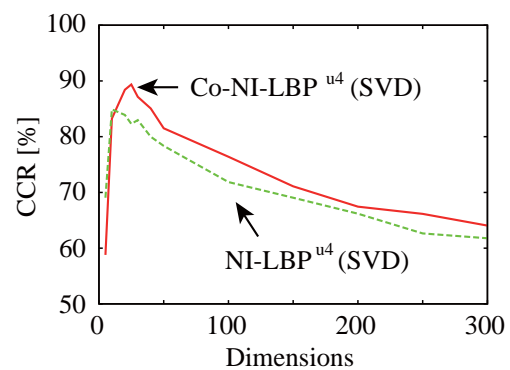


図 2 Co-NI-LBP^{u4} と NI-LBP^{u4} の比較

表3 Co-NI-LBP^{u4}の混合行列 (%)

	廊下	台所	研究室	学習室	オフィス
廊下	96.06	0	1.27	2.54	0.13
台所	2.05	79.35	1.88	16.55	0.17
研究室	0	0	98.27	1.50	0.23
学習室	0.59	3.71	14.99	79.82	0.89
オフィス	0	2.51	4.63	4.83	88.03

また、表3にCo-NI-LBP^{u4}に対する混合行列を示す。これより、特に台所と学習室の識別率が低いことがわかる。これは、研究室と学習室は共に机や椅子が多くみられることや、台所には壁取り付けタイプや島タイプなど異なるタイプが含まれていたためと考えられる。

最後に、これまでで最も高い識別率を示しNI-LBP^{u4}と、提案したCo-NI-LBP^{u4}を組み合わせさせた場合の結果を表4に示す。ただし、NI-LBP^{u4}は398次元、Co-NI-LBP^{u4}は25次元であり、両者を単純につなぎ合わせた423次元の特徴ベクトルを用いた。この結果、識別率は91.70%であり、従来手法を上回る最も高い識別率を得ることができた。

表4 NI-LBP^{u4}とCo-NI-LBP^{u4}を組み合わせさせた場合の識別率

Feature type	Dim	CCR (%)
LBP ^{u4} [11]	398	89.14
NI-LBP ^{u4} +Co-NI-LBP ^{u4}	423	91.70

6. まとめ

本稿では、RGB-D センサを用い、異なるモダリティ間の相関関係を考慮した新たな周囲環境の種別手法を提案した。この手法は、局所特徴記述子であるLBPに対し、モダリティ間の共起性を2次元ヒストグラムで表し、得られた特徴ベクトルをSVDにより次元削減するものである。実験の結果、従来手法より少ない次元数で従来手法と同等の識別率が得られた。これは、今後の大規模データを用いた空間識別では、重要な特徴となると考えられる。また従来手法と組み合わせ使用した結果、これまでで最も高い識別性能が得られた。

7. 謝辞

本研究は文部科学省科学研究費補助金基盤研究(A)(課題番号26249029)の支援を受けた。

参考文献

- [1] H. Zender, O. M. Mozos, P. Jensfelt, G.-J. M. Kruijff, and W. Burgard, "Conceptual spatial representations for indoor mobile robots," *Robotics and Autonomous Systems (RAS)*, vol. 56, no. 6, pp. 493–502, June 2008.
- [2] A. Pronobis and P. Jensfelt, "Large-scale semantic mapping and reasoning with heterogeneous modalities," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Saint Paul, MN, USA, May 2012. [Online]. Available: <http://www.pronobis.pro/publications/pronobis2012icra>
- [3] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," in *Pattern Recognition, 1994. Vol. 1 - Conference A: Computer Vision and Image Processing., Proceedings of the 12th IAPR International Conference on*, vol. 1, 1994, pp. 582–585 vol.1.
- [4] O. M. Mozos, H. Mizutani, R. Kurazume, and T. Hasegawa, "Categorization of indoor places using the kinect sensor," *Sensors*, vol. 12, no. 5, pp. 6695–6711, May 2012.
- [5] C. Cortes and V. Vapnik, "Support-vector network," *Machine Learning*, vol. 20, pp. 273–297, 1995.
- [6] S. Knerr, L. Personnaz, and G. Dreyfus, "Single-layer learning revisited: a stepwise procedure for building and training a neural network," in *Neurocomputing: Algorithms, Architectures and Applications*, J. Fogelman, Ed. Springer-Verlag, 1990.
- [7] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [8] "九州大学 kinect 空間識別データベース," <http://robotics.ait.kyushu-u.ac.jp/~kurazume/research-j.php?content=db#d02>.
- [9] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.
- [10] L. Liu, L. Zhao, Y. Long, G. Kuang, and P. Fieguth, "Extended local binary patterns for texture classification," *Image and Vision Computing*, vol. 30, no. 2, pp. 86 – 99, 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0262885612000066>
- [11] O. M. Mozos, H. Mizutani, R. Kurazume, and T. Hasegawa, "Categorization of indoor places using the kinect sensor," *Sensors*, vol. 12, no. 6, pp. 6695–6711, 2012.

- [12] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *Int. J. Comput. Vision*, vol. 43, no. 1, pp. 29-44, Jun. 2001. [Online]. Available: <http://dx.doi.org/10.1023/A:1011126920638>
- [13] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *International Journal of Computer Vision*, vol. 62, no. 1-2, pp. 61-81, 2005.
- [14] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91-110, 2004.
- [15] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346-359, 2008.