464–8601
464–8601
471–8571 1
E-mail: †ninomiyah@murase.m.is.nagoya-u.ac.jp, ††ddeguchi@nagoya-u.jp,
†{kawanishi,ide,murase}@is.nagoya-u.ac.jp, †††{norimasa_kobori,kunimatsu_hashimoto}@mail.toyota.co.jp

あらまし 3 … 2 … Deep Convolutional Neural Network DCNN … DCNN …

キーワード 3 …

Preliminary study on deep manifold embedding for 3D object pose estimation

Hiroshi NINOMIYA†, Yasutomo KAWANISHI†, Daisuke DEGUCHI††, Ichiro IDE†, Hiroshi MURASE†, Norimasa KOBORI†††, and Kunimatsu HASHIMOTO†††

† Graduate School of Informaion Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464–8601 Japan
†† Information Strategy Office, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464–8601 Japan
††† Toyota Motor Corpolation
1 Toyota-cho, Toyota-shi, Aichi, 471–8571 Japan
E-mail: †ninomiyah@murase.m.is.nagoya-u.ac.jp, ††ddeguchi@nagoya-u.jp,
†{kawanishi,ide,murase}@is.nagoya-u.ac.jp, †††{norimasa_kobori,kunimatsu_hashimoto}@mail.toyota.co.jp

Abstract Recently, 3D object pose estimation is being focused. The parametric eigenspace method is known as one of the fundamental methods. It represents the appearance change of an object caused by pose change with a manifold embedded in a low-dimentional subspace. It obtains features by PCA, which maximizes the appearance variation. However, there is not always a correlation between pose change and appearance change. So, there is a problem that the method cannot handle a pose change with a slight appearance change. In this report, we introduce deep manifold embedding which maximizes the pose variation. We construct a manifold from features extracted from Deep Convolutional Neural Networks (DCNNs) trained with pose information. Pose estimation with the proposed method achieved the best accuracy in experiments using a public dataset.

Key words 3D object, pose estimation, manifold, deep learning

1. まえがき

[illegible] Amazon Picking Challenge [1] [illegible] [2] [illegible]

[illegible] [3] [illegible] [4] [illegible]

[illegible] Murase [illegible] [4] [illegible]

[illegible] Deep

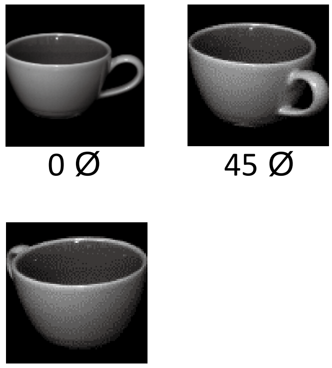


図 1 [illegible]

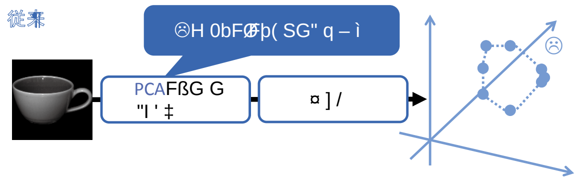


図 2 PCA [illegible]

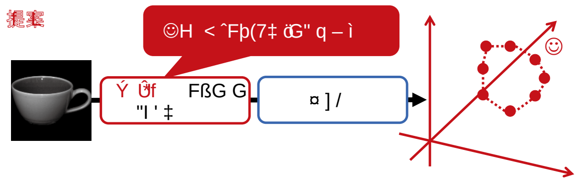
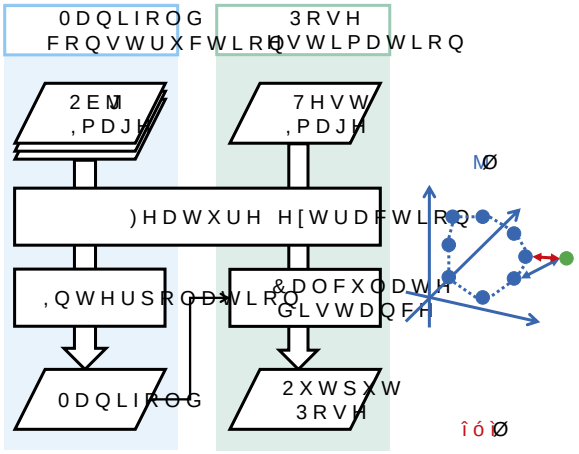


図 3 [illegible]

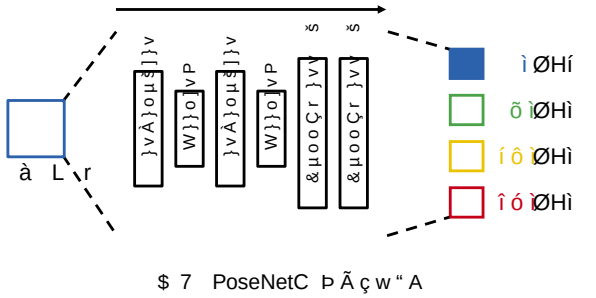
Convolutional Neural Network (DCNN) [5] [illegible] DCNN [illegible] [6] [illegible] 1 [illegible] DCNN [illegible] 2 [illegible] 3 [illegible] 4 [illegible] 5 [illegible]

2. 多様体に基づく姿勢推定

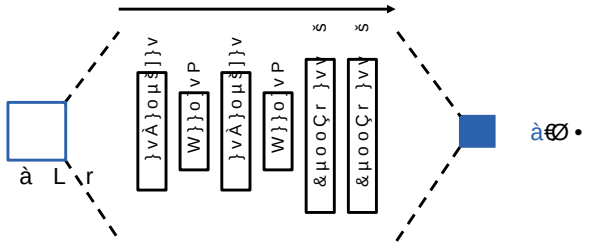
[illegible] 4 [illegible]



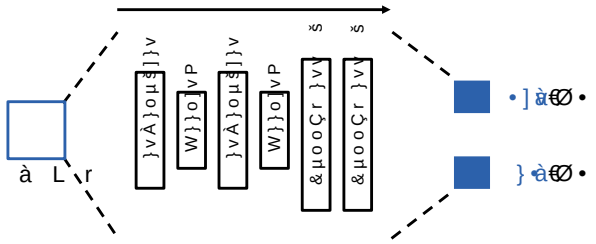
$ 4 7.Ï™t,nXŸN* rg q



$ 7 PoseNetC Þ Ã ç w“ A



$ 8 PoseNetR Þ Ã ç w“ A



$ 9 TriNetR Þ Ã ç w“ A

$ 5 ŸN›-£ô ø q`h DCNN w¶6 q

$ 6 ¶6A^ DCNN T’w ›Ã¨Z q

U ÉpK”}f\p| 7.Ï™tŸN›-£ô ø q`h-£K“¶6 O›;M”\qUßQ’•”}
fw¤p|f‘x Ú¶6t‘”›Ã¨Zt£è`h} Ú¶6qx;•¶6 Ow 1mpK“|›Ã¨ZT’Ý ‡p›° `o¶6pV” OpK”}\•t‘“¶6^•”›Ã” x pƒ-^•h›Ã”qz±`oôM¯qó—›Ëj|wÝ +w¶6tSMo‹ ®pK”\qUŒ’•oM” [7]}‡h|ú.wíòwÝÝw‘Os| t‘”›Ãƒ-UÉ` MðJtSMo‹| Ú¶6t‘“¨Z^•h›Ã”xôMQó›Ôb [8]}\w\qT’|ŸN›-£ô ø q`o¶6`hÚ¶6ÞÃç›;M”\qp|ŸNwümQwôM›Ã¨ZUpV”qßQ’•”}f`o|fw›Ã”›;MoÏ™`h7.s’y| Rüüst‘” 7.px0 pVsM_Qw!=U–^MŸNw§M›à b”\qUpV”}
Ú¶6›;Mh 7.Ï™›æsOhŠtx|‡c Ú¶6ÞÃçw 1mpK” DCNN w¶6›æO}¶6±ïÓç›t x|* 0Åú.›Ú™ws8àtHlo|° w¯S]qt s8^dhhþ›;M”}\\p|$ 5tÔb‘Ot|-£ô ø q`o¤ú.wŸN›)Q”}\•t‘“|ŸNwümQt £è`h¶6Uæ˜•”}f`o|¶6`h DCNN ›;Mo›Ã¨Z›æO}$ 6tfwrg q›Ôb}¶6`h DCNN t* 0Åú.w¶6±ïÓç›6|Ö—`|¤ ÚwZ—››Ã”q`o¨Zb”}\w‘Ot¨Z`h›Ã”t0`|Íåݍæ¿«{ í Oq‰7t4 rg›æsM| 7.›Ï™b”}

3. ŸN›-£ô ø q`h DCNN w¶6

ŸNwümQUôM›Ã›¨Zb”hŠt|ŸN›-£ô ø q`o DCNN ›¶6b”}fwM|-£ô øw)QMwŸs”Ž<w 3mwÞÃç› Šb”}

- € PoseNetC•ú.ŸN›-£ô ø q`hü¨ÞÃç}
- € PoseNetR•ú.ŸN›-£ô ø q`hs<ÞÃç}
- € TriNetR •ŸNw3¯ :‹›-£ô ø q`hs<ÞÃç}

3.1 PoseNetC

PoseNetC qxú.ŸN›-£ô ø q`hü¨ÞÃçpK”}\wÞÃçpx|ŸN›m„=`|ŸN* ðJ›«åµü¨ðJq`o Ü=b”}

$ 7tÞÃçw“A›Ôb}¶owú.pfw ¨t‘’c|ŸN›-£ô ø q`o¶6›æsO}m‡“|ŸN ›m„=`|0 b”«åµwˆ› 1|f•ŽŽ› 0q`h¶6›æsO}fwhŠ| Output layer wUnit :x¶6t;MhŸNw¨¡iZ Ob”}¡) :tx¦)¤ïÄéÐ”›;M”}\w‘Ot¶6›æsO\qp|›Ãí Íp‰°«åµtb”ŸNUf•g•Ÿs” 1:tB¤b”‘Os¶6Uæs

[illegible] 3 [illegible]

	[illegible]
Pixel	1.16 [illegible]
PCA	1.39 [illegible]
ObjNetC	1.70 [illegible]
PoseNetC	1.09 [illegible]
PoseNetR	1.59 [illegible]
TriNetR	1.72 [illegible]
OverFeat [13]	1.89 [illegible]

TriNetR [illegible] PoseNetR [illegible] PoseNetC [illegible]

$$\sin\theta^2 + \cos\theta^2 = 1 \quad (1)$$

[illegible] PoseNetC [illegible] 5 [illegible]

4.2 DCNN を用いた多様体構築による姿勢推定

[illegible] PoseNetC, PoseNetR, TriNetR [illegible] COIL-20 [illegible] ObjNetC [illegible] ImageNet 2012 training set [12] [illegible] OverFeat [illegible] [13] [illegible]

[illegible] Convolution layer [illegible] Fully-connected layer [illegible] COIL-20 [illegible] Fully-connect 3 [illegible] OverFeat [illegible] Fully-connect 8 [illegible] 512 [illegible] 4,096 [illegible] 80% [illegible] 10 [illegible] 16,384 [illegible] 2 [illegible] DCNN [illegible]

[illegible] 3 [illegible] PoseNetC [illegible] ObjNetC [illegible] OverFeat [illegible]

[illegible] 4 DCNN [illegible] Unit [illegible]

	Units: 256	Units: 512	Units: 1024
PoseNetC	12.36 [illegible]	7.92 [illegible]	7.33 [illegible]
PoseNetR	32.42 [illegible]	28.32 [illegible]	27.00 [illegible]
TriNetR	12.76 [illegible]	9.29 [illegible]	9.75 [illegible]

[illegible] 5 [illegible] Unit [illegible]

	Units: 256	Units: 512	Units: 1024
ObjNetC	1.76 [illegible]	1.70 [illegible]	1.91 [illegible]
PoseNetC	1.17 [illegible]	1.09 [illegible]	1.16 [illegible]
PoseNetR	1.59 [illegible]	1.59 [illegible]	1.75 [illegible]
TriNetR	1.62 [illegible]	1.73 [illegible]	2.11 [illegible]

[illegible] ObjNetC [illegible] OverFeat [illegible] PoseNetR [illegible] TriNetR [illegible] PoseNetC [illegible] PoseNetC [illegible] 2 [illegible] DCNN [illegible]

[illegible]

4.3 DCNN のネットワークアーキテクチャの違いが姿勢推定精度に与える影響

[illegible] DCNN [illegible] Fully-connected layer [illegible] Unit [illegible] Fully-connected layer [illegible] Unit [illegible] 256, 512, 1024 [illegible] 3 [illegible]

[illegible] 4, 5 [illegible] DCNN [illegible] Unit [illegible] 512 [illegible] DCNN [illegible] Unit [illegible] Unit [illegible]

5. むすび

[illegible]

ν Κ ρ Κ ᶅ ` Μ д ь ˒ Њ ς λ Κ ν ب ₮ Ɣ Ŋ ǫ ɰ Ï ο Ғ " مخي ╚ عجٍ
ȯ ’ ᛋ عج ς ؤ ΰ ɰ ǃ à Ƕ υ ب ₮ ثي Σ ả Ω ﾅ Ҷ ɰ ‘
謝辞 ᾅ ẓ υ ӿ لجي φ " ψ ᶅ ᾅ ẓ سد ˥ Ự ς Ѯ ɰ ‘

文 献

[1] N. Correll, K.E. Bekris, D. Berenson, O. Brock, A. Causo, K. Hauser, K. Okada, A. Rodriguez, J.M. Romano, and P.R. Wurman, “Lessons from the Amazon picking challenge,” ArXiv:1601.05484, Jan. 2016.

[2] J. Broekens, M. Heerink, and H. Rosendal, “Assistive social robots in elderly care: A review,” Gerontechnology, vol.8, no.2, pp.94–103, April 2009.

[3] M. Torki and A. Elgammal, “Regression from local features for viewpoint and pose estimation,” Proc. 2011 IEEE Computer Society Conf. on Computer Vision, pp.2603–2610, Nov. 2011.

[4] H. Murase and S.K. Nayar, “Visual learning and recognition of 3-D objects from appearance,” Int. J. Computer Vision, vol.14, no.1, pp.5–24, Jan. 1995.

[5] A. Krizhevsky, I. Sutskever, and G.E. Hinton, “ImageNet classification with deep convolutional neural networks,” Proc. 26th Annual Conf. on Neural Information Processing Systems, pp.1097–1105, June 2012.

[6] A.S. Razavian, H. Azizipour, J. Sullivan, and S. Carlsson, “CNN features off-the-shelf: An astounding baseline for recognition,” Proc. 2014 IEEE Conputer Society Conf. of Computer Vision and Pattern Recognition, pp.806–813, June 2014.

[7] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, “DeCAF: A deep convolutional activation feature for generic visual recognition,” arXiv preprint arXiv:1310.1531, Oct. 2013.

[8] M. Cimpoi, S. Maji, and A. Vedaldi, “Deep filter banks for texture recognition and segmentation,” Proc. 2015 IEEE Computer Society Conf. of Computer Vision and Pattern Recognition, pp.3828–3836, June 2015.

[9] S.A. Nene, S.K. Nayar and H. Murase, “Columbia Object Image Library (COIL-20),” Technical Report, Columbia Univ. Dept. of Computer Science, CUCS-005-96, Feb. 1996.

[10] V. Nair and G.E. Hinton, “Rectified linear units improve restricted Boltzmann machines,” Proc. 27th Int. Conf. on Machine Learning, pp.807–814, June 2010.

[11] G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R.R. Salakhutdinov, “Improving neural networks by preventing co-adaptation of feature detectors,” arXiv preprint arXiv:1207.0580, Oct. 2012.

[12] J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” Proc. 2009 IEEE Computer Society Conf. of Computer Vision and Pattern Recognition, pp.248–255, June 2009.

[13] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus and Y. LeCun, “Overfeat: Integrated recognition, localization and detection using convolutional networks,” arXiv preprint arXiv:1312.6229, Dec. 2013.