[[1]](#footnote-1)

A Self-Taught Vision System for Automatic Learning and Recognizing 3D Objects

Ren C. Luo∗, and Po-Yu Chuang†,

† Electrical Engineering Department, National Taiwan University, Taipei, Taiwan 106

*Abstract—Vision systems for 3D object recognition are widely applied on industrial robot arm recent years. Conventionally, vision, learning approach, and robot arm are separated into three different systems, so learning approach can only learn the distributions of input data, but cannot refine qualities of input and output* *without manual interference. Therefore, it may cause misfiring performance by poor input or sustainable growth database. In this paper, we propose an intelligent vision system for 3D object recognition which aims to construct and refine the relational models between input 2D images and output rotation angles of robot arm without any manual interference. Although 2D image and rotation angle are different domain knowledge, we model 3D objects by multiple 2D images and rotation angles. Therefore, information of different domains can be integrated by the relational model. The relational models are formulated based on proposed Hierarchical-Deep model and probabilistic descriptor to make the relational model can be constructed and refined without manual interference. The experimental results show the feasibility of proposed structure that can transfer knowledge in different domains, and complete assigned task by only modeling the relation between input and output.*

*Index Terms—industrial automation, machine vision, object recognition, Machine learning*

# INTRODUCTION

Robot arm with vision have been widely applied in automatic industrial production line in recent years [1-4]. New hurdles of conventional vision system arise due to rise of consumer electronic market. The components in assembling production line become small volume but large variety. The types of components are also changed rapidly because of short product life cycle. These conditions are tough for conventional vision system.

For traditional systems, model-based recognition methods are commonly used in industrial applications. The performance is mainly related to the manually labeled data. Hence, the system is hard to adapt with adding new kinds of work pieces. For 2D object recognition, users need to capture numerous raw data of target objects in different poses. These cumbersome works lift to much more complex level while the methods are expanded to 3D object recognition. These manual works not only increase the labor cost, but also point out the dilemma of present vision-based robot which is unable to automatically adapt to various assignments, and growing database.

In general industrial purposes, we do not really concern about all details of entire 3D object. Instead, the relation between input and target is the key point for completing the tasks like pick and place, inserting components, etc. To solve these issues, we propose an intelligent task-oriented vision system which acquires ability in automatic learning relational model of input and target. In task-oriented view, system tends to learn the relation between input and target rather than delicate models for target 3D objects. Therefore, the state problem is that we only provide target face of 3D objects which are intended to be placed on top by robot arm, but the other faces of 3D objects are unknown. The labelled data is target faces, and inputs are arbitrary objects with random faces on top. The input of system is 2D image data, and output is rotation angle of robot arm in Cartesian space which is different feature domains. Therefore, the traditional single layer models [5, 6] which end in a linear or kernel classifier is not enough. We introduced a ***Hierarchical-Deep(HD)*** model [7] to tackle the problems.

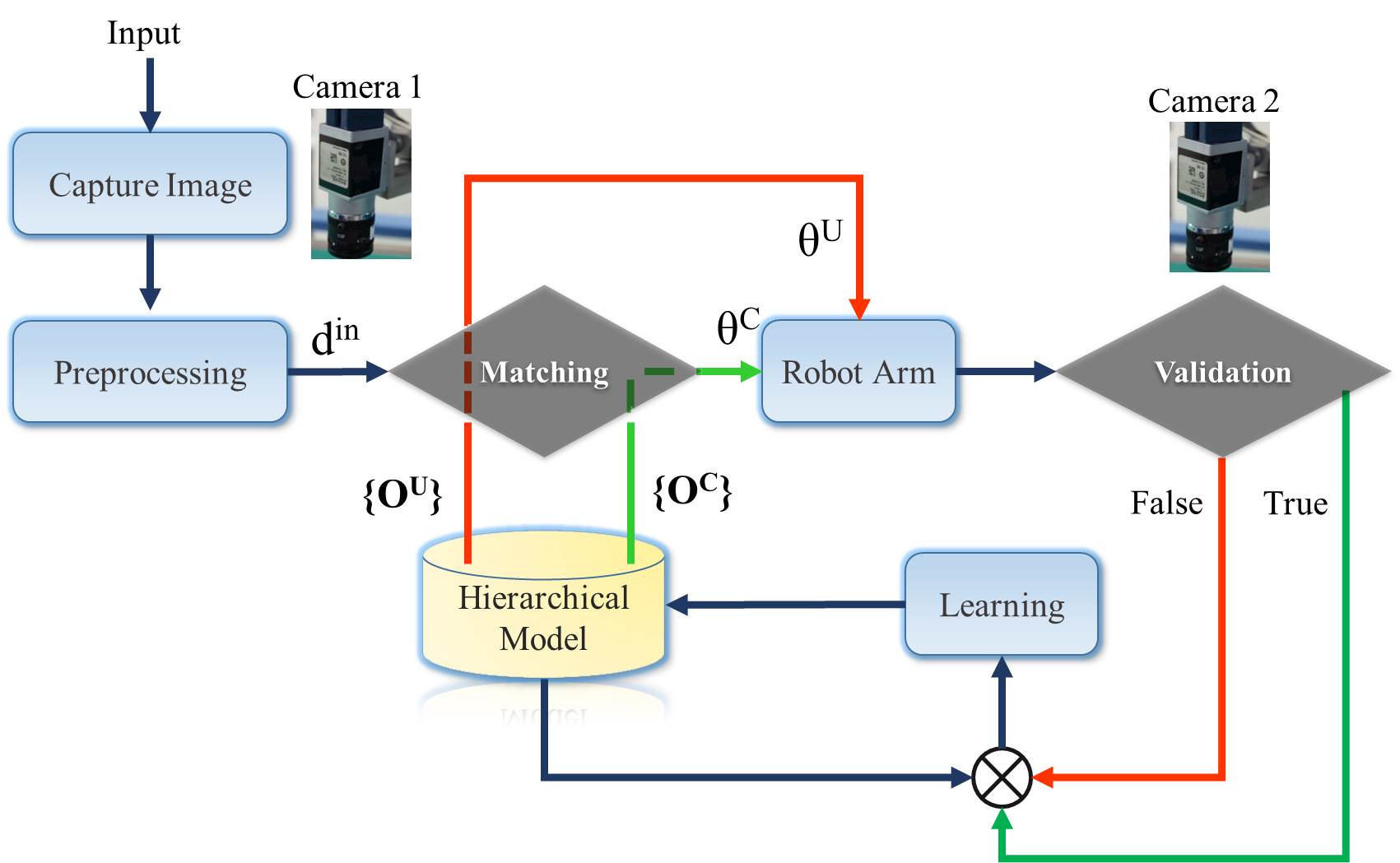
The learning of HD model achieves dramatically success recent years. Hinton et al. [7] proposed deep structure learning which hidden layers are formed by lower level feature to higher level hierarchically, and had been successfully applied on different research fields [8-11]. Enlightening by deep learning methods, hierarchical structure is applied to our model which is constructed by four layers: **Feature**, **Descriptor**, **Object** and **RotationAngle**

Through this model, the feature in different domains could be correlated through hierarchical structure, but the system still cannot automatically learn the relation between input images and output rotation angle. Being an automatic system, the ability which could "infer" latent edges between labeled and unlabeled data is needed. Latent edge means two variables in different layers exist an edge in graph model if prior data is sufficient, but, in our case, system only have small amount of prior data. Hence, there are many latent edges which are waiting to be revealed through learning process.

The most challenge part of state problem is that the appearance of different faces of a single object might be quite different, so we design three modules to tackle self-taught problem. Firstly, we design a probabilistic based image descriptor. Although many methods [12-16] provide high quality performance by extracting sparse features, the sparse feature is not compact on inferring the relational model. The sparse feature only model strong features of observed face, but most of faces is unknown in our case. We need a descriptor which can provide sufficient information for inferring latent edges, but still retain scale- and rotation-invariant. Proposed probabilistic based descriptor is established based on the ***Markov Logic Network (MLN***) [17-20]. MLN is an approach combines first-order logic and probabilistic graphical model. First-order logic enables compactly representing the neighborhood of feature points. Probabilistic graphical model can reveal latent edges by proper inference method, and also handle the uncertainty.

Secondly, transfer information module is proposed for constructing latent edges. Transfer information module is realized by Self-taught Clustering algorithm [21]. Self-taught Clustering algorithm is a transfer learning method [22-25] which is built for enhancing model through large amount of auxiliary unlabeled data. The input face can be considered auxiliary unlabeled data, and find co-cluster between priors face in the dataset. Hereafter, we further utilize the distribution of co-cluster to infer the possible rotation angle for robot arm, and robot arm will rotate target object from input face to output face. Finally, the validation module is an eye-to-hand camera which used to validate the error between the output face and desired target. Then, the validation module returns the error to the model in order to refine the existed model. Through these three modules, proposed system can automatically learn the relations between input images and corresponding rotation angles with only labelled the target face of each object.

In this paper, we start with briefly overview of system design and structure in section 2. The MLN-based descriptor for recognizing object is described in section 3. Section 4 introduces how to model the proposed hierarchical networks, and learn by self-taught learning. Then, we compare the performance of proposed system with several states of development in section 5. Finally, reviewing performance and conclusions are presented in final section.

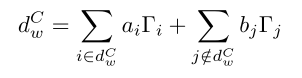
 Fig. 1 System architecture

# SYSTEM ARCHITECTURE

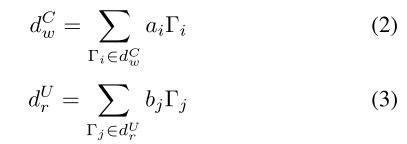
The main purpose of this system is to automatically derive the relationship between input face and target face of 3D assigned objects. The only prior knowledge are the target face. Input is arbitrary assigned object with random face on top, so input is very likely an unknown face of assigned object showed-up rather than prior target face. Therefore, system has to infer the correlation between input and existed priors. Proposed system is shown in Fig. 1. Camera 1 captures images of all input objects with random faces on top, and constructs MLN-based descriptor for each input. Then, system matches the input with data in database and output rotation angle for robot arm. After robot arm placing an object, camera 2 will validate result, and feedback error for refining existed model. The system architecture in Fig. 1 is realized by a hierarchical-deep model in Fig. 2.

The variables in the same layer are independent, and vertical adjacent two layers are full connected. Variables in **Feature**(**ΓC**) layer are extracted image features, and variables in both **Classiﬁed Descriptor**(**DC** ) and **unclassiﬁed Descriptor**(**DU**)are MLN-based descriptor. Variables in **Rotation angle(ΘC)** and **Inferred Rotation angle**(**ΘU** ) are set of rotation { Row (α), Pitch(β) , Yaw(γ) } angles respect to target faces. Finally, variables in **Object**(**OC** ) are combinations of rotation angles.

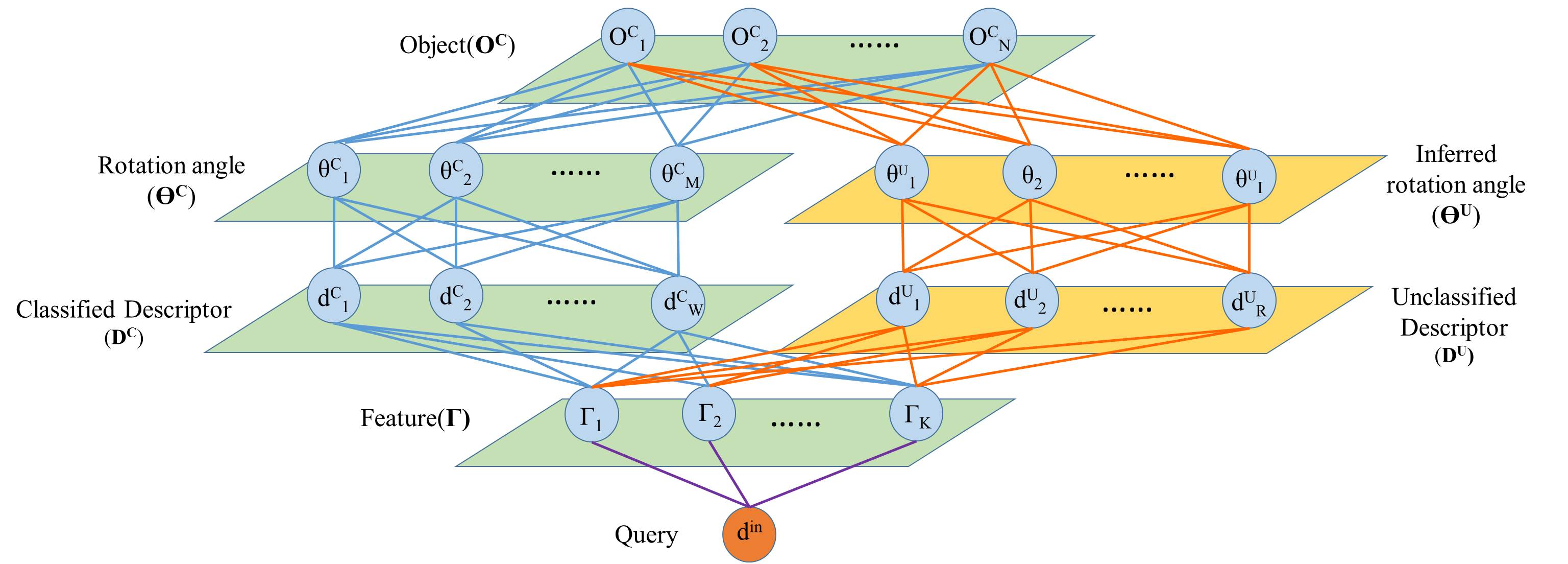
The difference between classic ***Deep Belief Networks(DBN*)** is that proposed model exist two parallel parts in Fig. 2. **DC** -**ΘC** and **DU** -**ΘU** have no connection between each other, but both have full connection with deepest layer **OC** and first layer **ΓC**. To handle tons of unknown data, the structure of connection will dynamically change with observed evidences. Sparse coding method is used to constructs edge in the model, most of connection is zero which is called latent edge in this paper. Latent edge might become non-zero while some new evidences have been discovered. For variable dCw  in layer **DC**, the sparse coding result should be formulated as:

 (1)

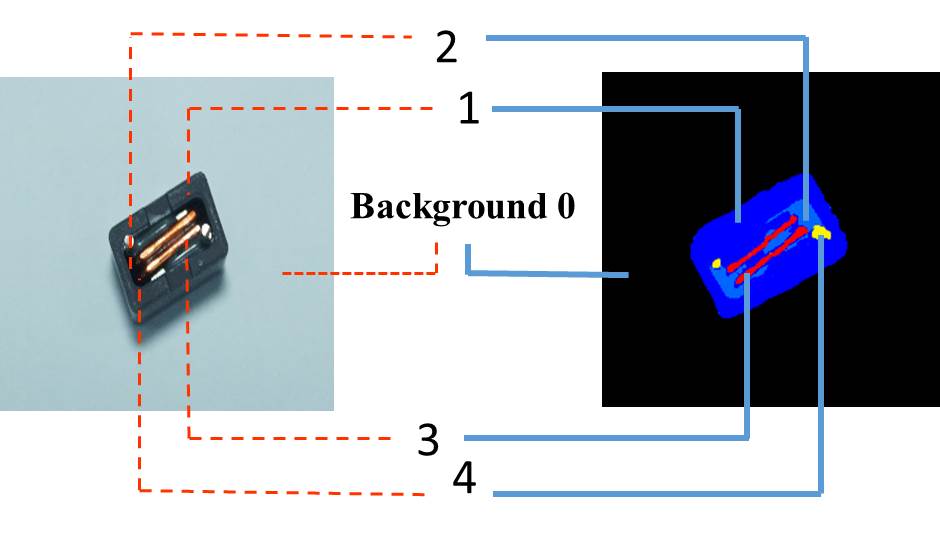
Although Eq.(1) can handle the problem of latent edge, it's impractical to sample all possible conditions whenever new evidence showing up. Therefore, proposed model separate descriptor layer into two parallel parts as:



dCw  is considered a prior descriptor which the edge between **ΘC** and **OC** had been established. Therefore, the left part of parallel layers can be considered as static model until there is a query classified to the **DU**. **DU** layer is the set of descriptors which we haven't known that these descriptors are correspondent to which object. Therefore, we propose an inference method to infer the possible rotation angle, and camera 2 will check inferred results. If inference is success, variable dUr and θU are used to re-estimate correlation between layers through hierarchical structure. Therefore, latent edges can be revealed though more success inferences.



**Fig. 2. Hierarchical-deep model for self-taught system**



(a) Result of background subtracting and clustering



(b) Serial captured frames

**Fig. 3. Preprocessing of input objects**

# MLN-based Descriptor

## The concept of constructing MLN-based descriptor

Being a self-taught system, deriving more valuable information from raw data helps system deriving more reliable results with scarce prior knowledge. Most of present image descriptors [12-16] are constructed based on strong sparse feature point, because these points are consistent even in different environment. These kinds of descriptor can efficiently and precisely match given image. Nevertheless, most of observed face is not in prior data, so we need a descriptor which can infer the relation between observations and priors. To avoid losing information, we choose normal distributed feature instead of sparse feature. Since different faces of an object may exist different strong features, normal distributed feature is more suitable for our case. Reference regard as a stable and normal exist in geometric space.

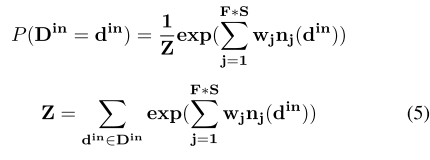
**Table I. Example of predicates and ﬁrst-order logic formulas**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Key Atom | 1 | 2 | 3 | 3 |
| Predicates | ne(1,2) | ne(2,1) | ne(3,1) | ne(4,1) |
| ne(1,3) | ne(2,3) | ne(3,2) | ne(4,2) |
| ne(1,4) | ne(2,4) |  |  |
| ne(1,0) |  |  |  |

For prepossessing of input images, each channel of RGB domain is classified into 5 parts, and get 125 classes in RGB domain. An input image will be segmented by these classes. In Fig. 3(a), an observed face of input object is segmented into 4 classes, and class 0 is background. Hereafter, predicates for MLN networks are constructed by segmented results. We have only two kinds of predicate ***ne(a,v)*** and ***des(x)***for MLN model. Variable ***a*** is an atom cluster, and variable ***v*** is a neighbor of atom cluster, so predicate ***ne(a,v)*** represent adjacency of atom cluster. Variable ***x*** in ***des(x)*** is a MLN-based descriptor. The variables of feature layer in Fig. 2 are predicates ***ne(a,v)***. Since every classes can be the atom cluster, we have C1252 binary variables in feature layer. Taking Fig. 3(a) as an example, the predicates of preprocessed image are shown in Table Ⅰ, and first order logic is formulated as:

∀a∀v *ne(a,v) ⇒ des(x)* (4)

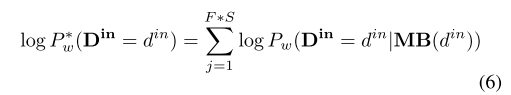
Each image will further be down sampled, and derived several images with different scales. For each image, we derive ***F\*S*** formulas where ***F*** is number of serial captured images and ***S*** is number of images with different scales. Through these formulas, a MLN model can be constructed. The probability distribution over possible world din specified by MLN is given by:



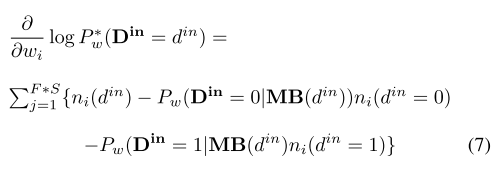
Where din is the descriptor of input image. nj (din ) is the number of true grounding of formula j in din, and wj is weight of formula j .

## Inference and Weight learning of MLN-based descriptor

The weights of MLN-based descriptor are learned by maximizing the pseudo-log-likelihood. Since each descriptor can be consider as a closed world, we only need to consider the atoms which derive from captured serial frames. Comparing with uniform sampling approach, maximizing pseudo-log-likelihood is more efficient, because pseudo-log likelihood only need to consider relational data. The pseudo-log -likelihood of Eq.(5) can be written as:



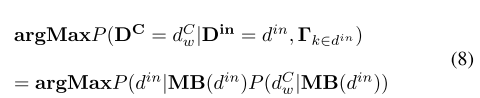
Where MB(din) is Markov blanket while din is observed. The MLN weights are learned generatively by maximizing the pseudo-log-likelihood of Markov blanket. The gradient of the pseudo-log-likelihood with respect to the weight is:



Where ni (din = 0) is the number of true grounding of j th formula while force din = 0, and similar for ni (din = 0\1). The learning of pseudo-log-likelihood in our approach are further boosted by ***Limited-memory Broyden-Fletcher-Goldfarb-Shanno(L-BFGS)*** optimizer [20] to make entire process become more efficiency.

## Matching of MLN-based descriptors

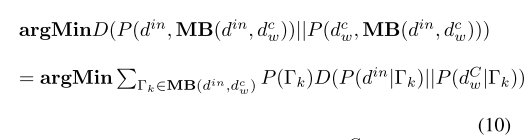
For each constructed input descriptor din, system would search for the matched descriptor in the database, and further arrange it to the proper layer of **DC** or **DU** as shown in Fig.2. Since input is possible to be assigned to one of parallel layers, matching step is separated into two parts. One is using pseudo-log-likelihood for deciding observation should be assigned to which layer. The pseudo-log-likelihood of descriptors matching could be formulated as:



If input descriptor doesn’t match any descriptor in **DC** layer,the descriptor becomes a variable of **DU** layer. For a variablein **DU** , we would like to infer rotation angle to make inputobject can be placed on corresponding target face. Since the rotation angles for descriptors in **DU** had been identiﬁed, the second part for matching is trying to ﬁnd a descriptor in **DC** which have max co-cluster with input descriptor. Finding max co-cluster can be alternately considered as minimizing loss of information as:



The common feature Γk∈din ∩dWC is further represented by co-Markov Blanket of din and dC , and the loss of mutual information can be further formulated by KL divergence **[28]** as:

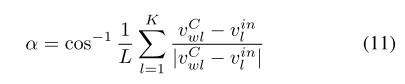


By Eq.(10), classiﬁed descriptor dCw with min DL-divergence is considered acquired max co-cluster with dCw.The relation between the co-cluster become the evidence forinferring rotation angle of din. Through Eq.(8) and Eq.(10),the input descriptors are classiﬁed to corresponding layer, andbecome inputs **ΘC** or **ΘU** layer.

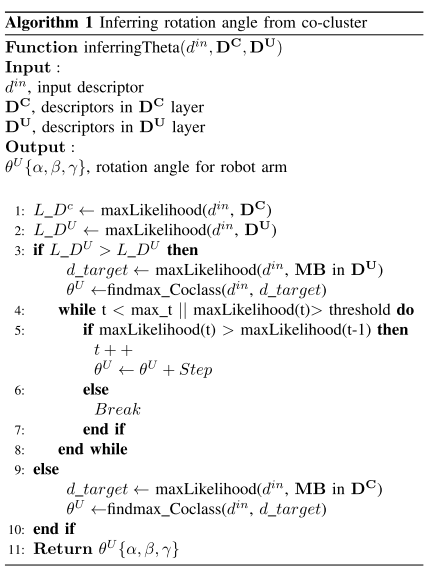
# Hierarchical Model

## Inference of rotation angle in **ΘU** layer

Inference rotation angle θiU is based on max co-cluster between din and dC. A set of co-cluster {Cw1 , Cw2 , ..., CwL } can be derived by minimizing KL divergence. The center of co-cluster with respect to center of camera in Cartesian space can be derived into two sets: **Vin** ={v1 , v2 , ..., vL } and **Vw** ={vCw1 , vCw2 , ..., vCwL }. The roll angle α of robot arm is calculated by:



Where roll angle α is the mean angle of co-cluster in two descriptors. As for pitch angle β and yaw angle γ, the pitch and yaw angle are hard to be estimated by 2D descriptor directly. We make random sample these two angles in value π/2 , and −π/2 initially, and approximate to actual angles by algorithm 1.

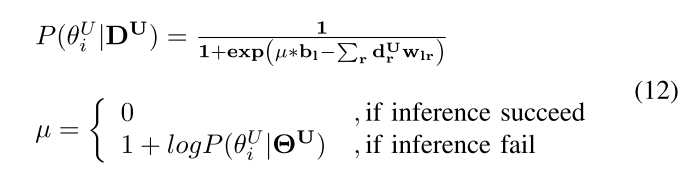


## Inference and learning of hierarchical-deep model

Proposed hierarchical model is a generative model of **Deep**

**Belief Network (DBN)**. Structure between each layer is shown in Fig. 4. Each layer is considered as an **Restricted Boltzmann** **Machine (RBM)[8]** except **Γ** , **DC** , and **DU** . The MLN is trained by pseudo-log-likelihood as mentioned previously, and RBM is trained by greedy layer-wise training [30].

Initially, left part of model (**Γ-DC-ƟC-****OC**) are trained with prior target face of objects, and number of variables **N** in **OC** equal to the number of prior target faces. The right part of model is activated only when a new observation is classified into **OU**. The activation probability of θUi is a sigmoid activation function:

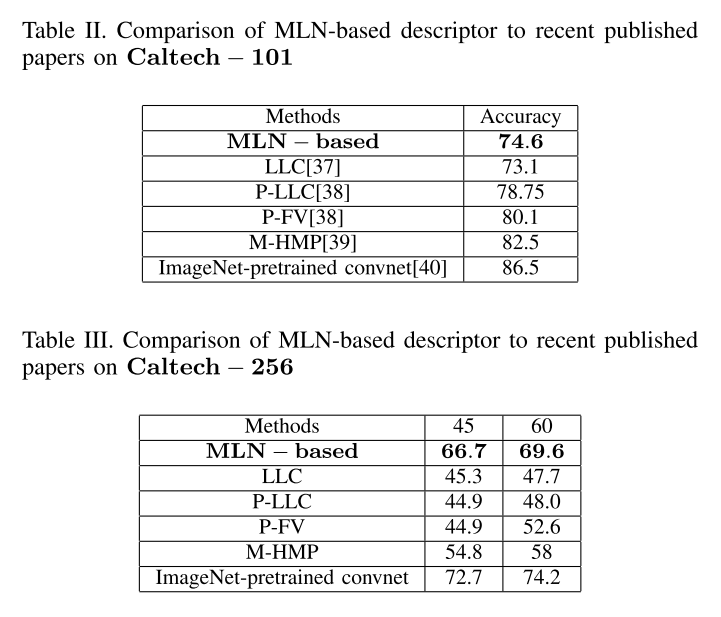


where $\mu$ is penalty factor which decreases the probability while the inference is failed. $\mu$ is depended on log-likelihood of θUi which can lead to lower activation probability if inference result failed several times, and avoid system derives wrong results over again.

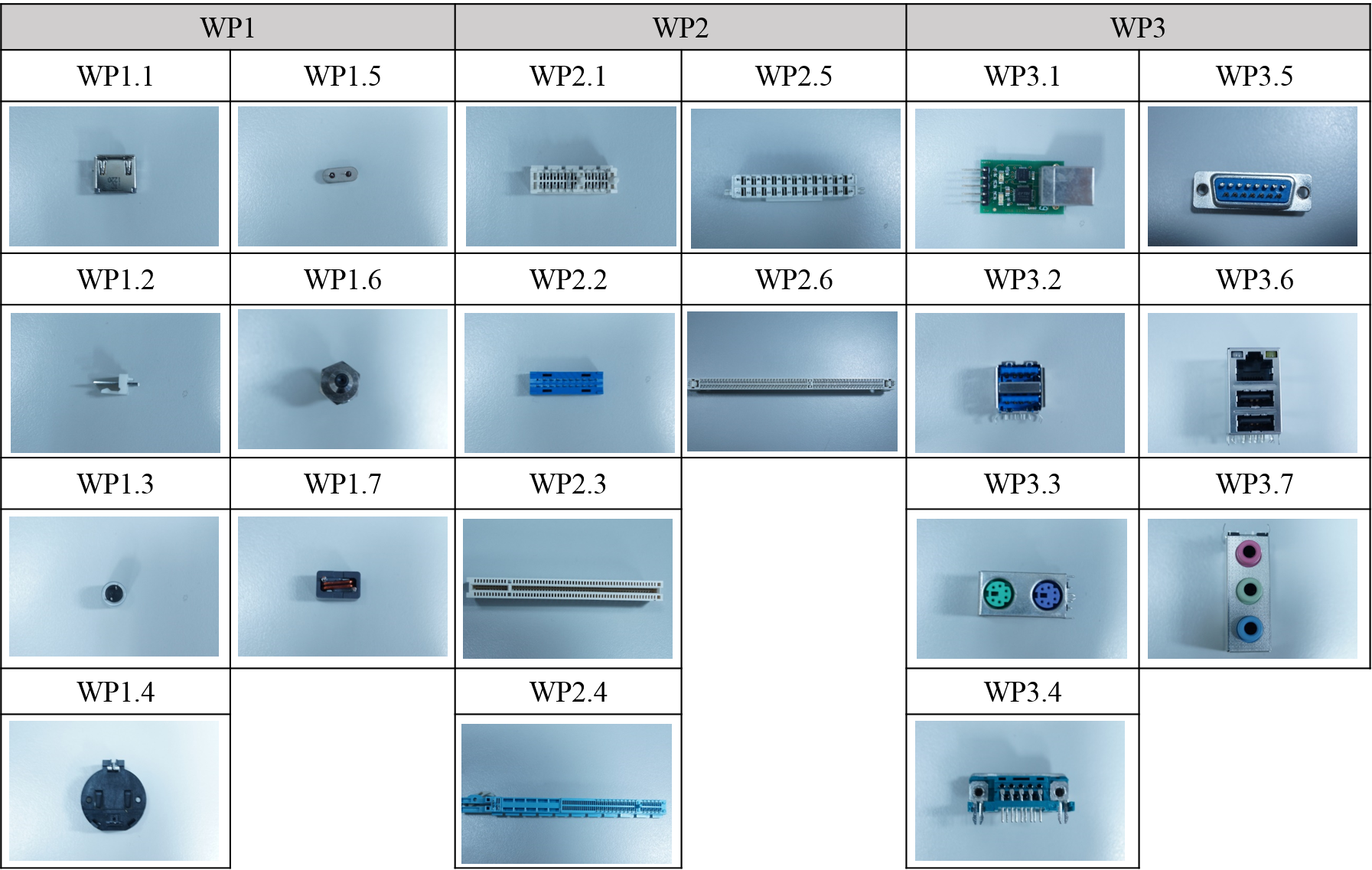
On the other hand, for both **ƟU** and **ƟC** layer, if results are correct, the model will be retrained by greedy layer-wise training. If validated result is derived from left part of Fig. 4, the generative model is defined by the joint distribution of top layers ***P***(**OC**,**ƟC**). If the result is derived from right part, the generative model is defined by ***P***(**OC**, **ƟU)**. By doing so, the relation between prior and observations can be self-taught from numerous random unlabeled inputs, and self-inferred with possible relational model while new assigned objects involved with proposed model.

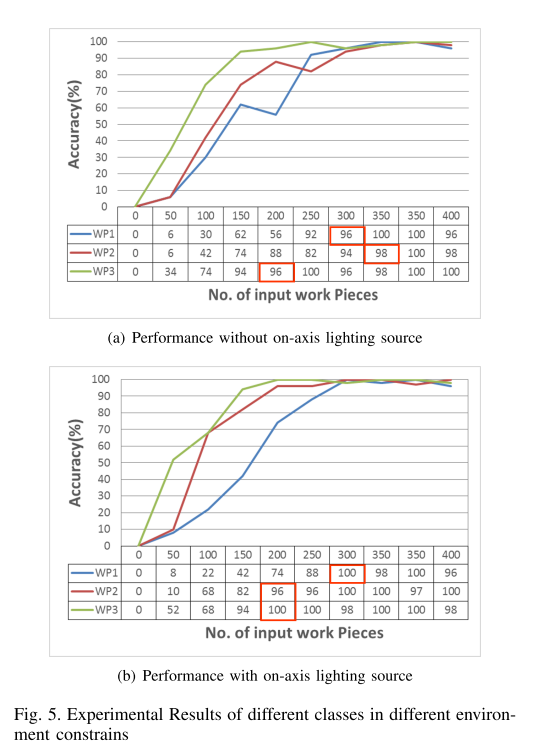
# Experiments

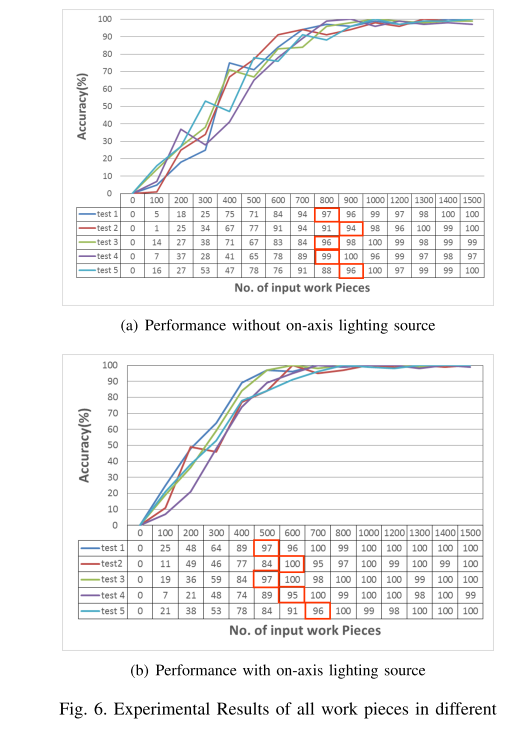
The experiments for proposed model are separated into two parts. Firstly, we evaluate the performance of MLN-based descriptor by standard object recognition datasets:

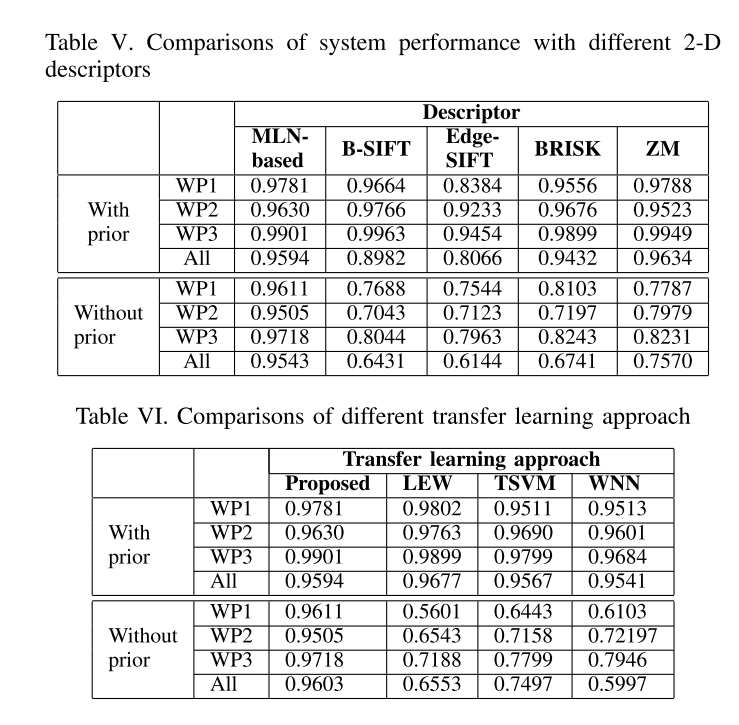


Caltech − 101 [31] and Caltech − 256 [32]. Results shown in Table II are comparisons with recently published papers. The images in the datasets are rescaled into ﬁve different scales for training proposed MLN-based descriptor. For Caltech − 101, we follow general procedure and randomly selecting 30 images for each class. For Caltech − 256, select 45 and 60 images for each class, and trained by pseudo-log likelihood. Although, for the proposed model, the result does not outperform in Caltech − 101, but the accuracy in Caltech − 256 is slightly behind ImageNet-pretrained model. In the other scopes, the result shows that MLN-based descriptor keep well performance even increasing categories. Most of recently published methods get dramatically performance decreasing while categories increase from 101 to 256. Therefore, MLN-based descriptor is compatible to be a descriptor in large amounts of unlabeled data.

Table IV. Three classes of testing work pieces for experiments





 For the second part of experiment, we implement the proposed system in real industrial application. The prior knowledge is target faces of assigned objects, and there are twenty kinds of assigned object in our experiment. Table IV shows twenty target faces for each assigned object. The experiment is implemented based on several assumptions: The input objects are not occluded, and not adjacent with each other. Hereafter, the inputs of self-taught system are random choose the assigned objects with random face on top. The testing objects are classiﬁed into three classes in Table IV. For class WP1, the work pieces are featureless and relative small, so it’s hard to construct robustness descriptor even building relational model for entire model. For classWP2, all work pieces acquire similar shapes and size, so this kind of object is easily mismatched in the matching process. The work pieces in WP3 are matched group of this experiment. The work pieces acquire sufﬁcient features for descriptor, and have plenty of information for identifying and constructing relational model. In the ﬁrst stage of experiment, we compare the performance of proposed system between different classes in different lighting conditions. The results of different classes are shown in Fig.5. In Fig.5(b), the environment lighting is controlled by on-axis lighting source, so the information of object are more complete and distinct than images without lighting control in Fig.5(a). The accuracy is average of 100 times repeatedly testing.

The system is considered convergence while accuracy is over 95%, and stop learning approach. If the accuracy is under 95% again, the learning approach would be re-excuted. Comparing the results, in both cases, class WP3 could be convergent with least input sample, and convergent time of class WP2 is slowest. The results show that the efﬁciency of learning could be slightly improved by environment constrain, but the accuracy is not effected, and always keep over 95% after learning approach stopped.

Fig. 6 shows the result while all twenty kinds of assigned object are involved in the same time. The result shows that system need more inputs to converge while more kinds of objects are involved, but the system still slightly converge, and accuracy is all keeping over 95% for both conditions. In brief, these two experiments verify proposed system is competent to learn the relational model automatically. Although the learning rate is dragged by the number of assigned objects, the learning rate still can be convergent by reasonable number of inputs. The experiment in Fig. 5 and 6 testiﬁed the performance of proposed system can meet our requirements. We compare the performance of proposed system with other advanced approaches. Since none of similar systems could handle this issue in our literature survey results, so the comparisons are done by dividing our system into two parts. One is 2-D descriptors for each face of objects, and the other is machine learning approach for learning relational model.

For the descriptor part, four kinds of other descriptors are chosen to compare with proposed system. B-SIFT[35]and Edge-SIFT[36] are modiﬁed versions of SIFT approach which enhanced the accuracy of feature point registration. ***Binary Robust Invariant Scalable Keypoints(BRISK)***[15] descriptor is constructed based on binary robust invariant scalable key points, and ***Zernike Moment (ZM)***[13] phase-based descriptor is a moment-based descriptor which use the phase information of signal. All of these descriptors are representative methods in relative ﬁeld recent years, and had been testiﬁed by many researchers. To compare the robustness and accuracy, the performance is testiﬁed by two conditions as shown in Table V. One is relationship of each face is prior of system, and the descriptors only provide information for object matching. The experiments are implemented by the same learning approach which is proposed in the previous section. The other is no prior for learning approach that information of descriptor need to be used for inferring the relational model. The ZM descriptor has the best performance in the condition with prior, but accuracies of descriptors are close. In condition, without prior, the MLN- based descriptor acquires best performance which testiﬁed MLN-based descriptor is suited for self-taught system.

Hereafter, the performances of different learning methods are further discussed. The other three kinds of transfer learning approaches: ***Locally Weighted Ensemble approach (LWE)***[25], ***Transductive SVM(TSVM***)[26], and ***Weighted Neural Net- work(WNN)***[27] are chosen to compare with proposed method. Similarly, the experiments are divided into two parts as shown in Table VI. The result shows LEW acquires the best accuracy in the condition with priors, and proposed learning approach acquire greatest performance in condition without prior. Although the performances between different methods are quite close when priors are provided, the accuracy of the other methods goes down in no prior condition. It seems that the results are not only affected by descriptor, but also learning approach.

# Comparisons of related works

The proposed system is a HDmodel with parallel branches. Conventionally, the descriptors [13, 15, 35-39] and learning model [27-28] are separated, so the learning approaches only learn the distribution of descriptors but cannot adjust the distribution of descriptors. According to our surveys, no such a descriptor constructor can fit every case without manual interference. Therefore, the proposed approach takes an alternative way to make learning approach can adjust the distribution of descriptors to make the MLN-based descriptor can be refined during learning stage. This thought is proved by experimental results show in Table V and VI. The descriptors have to be constructed without manual interference in condition of without prior knowledge, and the result shows the performances of other than proposed method are dramatic decreased.

Moreover, the other feature of proposed system is that learn knowledge in both image and rotation angle by just one model. Most of HDmodels [8-11] are focus on learning knowledge in the same domains, e.g. handwriting [8], text categorization[9], Speech Recognition [10], images[11]. To handle the knowledge in different domains, we involve the technique of self-taught clustering and parallel structure model, and the experimental result also shows the feasibility of learning knowledge in different domains by proposed HD model.

# Conclusion

An automatic learning system for vision system is an important part in assembling production line with small-volume, large-variety components. In this work, we reverse the concept of traditional vision system. The robustness of feature points and descriptor is not main concerns. Instead, the relational model between input and output is the most essential.

To learn the relationship between input and output, we propose a HD model which combines the concept of deep learning, transfer learning, and Markov logic network. The model acquires self-taught ability which can infer relational model and self-supervised the performance of learning results. Being an automatic system, tackling large amount of unlabeled data and inferring relation with labeled data is necessary. The relation between features in different levels can be represented as a discriminative distribution by the model. Through the discriminative distributions, KL divergence is further involved to ﬁnd the max co-cluster between labeled and unlabeled data, and makes system stable under growing database.

Moreover, proposed system includes image features, descriptors and rotation angles for robot arm. These different domain features are impossible to be learned simultaneously by traditional single layer model, but the experimental results prove proposed HD model can transfer and learn different domain knowledge, and recognize 3D object by only learning relational model. We believe this system is practical in real industrial assembling production line, and save labor cost.

References

[1] Torgny Brogrdh, ”Present and future robot control developmentAn indus-

trial perspective”, Annual Reviews in Control, Volume 31, Issue 1, pp.

6979, 2007.

[2] Ebrahim Mattar, ”Robotics Arm Visual Servo: Estimation of Arm-Space

Kinematics Relations with Epipolar Geometry, Robotic Systems - Appli-

cations, Control and Programming”, Dr. Ashish Dutta (Ed.), ISBN:

978-953-307-941-7, InTech, DOI: 10.5772/25605.

[3] So-Youn Park ,Yeoun-Jae Kim ,Ju-Jang Lee ,Byung Soo Kim, and Khalid

A.Alsaif,”Controlling robot arm manipulator using image-based visual

servoing without pre-depth information”, 37th IEEE Interantional

Conferece on Industrial Electronics, pp.3157-3161,Nov. 2011.

[4] K. Deguchi, H. Sakurai, and S. Ushida,”A Goal Oriented just-in-time

visual servoing for ball catching robot arm, in Int. Conf. on Intelligent

Robots and Systems, Sept. 2008, pp. 30343039.

[5] J. Baker, L. Deng, J. Glass, S. Khudanpur, Chin hui Lee, N. Morgan, and

D. OShaughnessy,” Developments and directions in speech recognition

and understanding, part 1”, Signal Processing Magazine, IEEE, vol.

26, no. 3, pp. 7580, may 2009.

[6] S. Furui,” Digital Speech Processing, Synthesis” , Marcel Dekker, 2000.

[7] Tong Simon, and Daphne Koller,”Support vector machine active learning

with applications to text classiﬁcation”, The Journal of Machine

Learning Research 2 pp 45-66, 2002.

[8] Hinton, G. E., Osindero, S. and Teh, Y,” A fast learning algorithm for

deep belief nets” , Neural Computation, 18, pp 1527-1554,2006.

[9] Srivastava, N., Salakhutdinov, R. R. and Hinton, G. E.,” Modeling

Documents with a Deep Boltzmann Machine” , In IEEE International

Conference on Acoustic Speech and Signal Processing (ICASSP 2013)

Vancouver, 2013.

[10] Graves, A., Mohamed, A. and Hinton, G. E,” Speech Recognition

with Deep Recurrent Neural Networks” , In Uncertainty in Artiﬁcial

Intelligence(UAI 2013).

[11] Ranzato, M., Mnih, V., Susskind, J. and Hinton, G. E.,” Modeling

Natural Images Using Gated MRFs” , IEEE Trans. Pattern Analysis

and Machine Intelligence, 2013.

[12] H. Bay, A. Ess, T. Tuytelaars, and L. Gool,”SURF: Speeded up robust

features”, Comput. Vis. Image Understand., vol. 110, no. 3, pp. 346359,

Mar. 2008.

[13] Zen Chen and Shu-Kuo Sun,”A Zernike Moment Phase-Based Descrip-

tor for Local Image Representation and Matching”, IEEE Trans. Image

Process., vol. 19, no. 1,pp. 205219, Jan. 2010.

[14] A. Alahi, R. Ortiz, and P. Vandergheynst,”Freak: Fast retina keypoint”,

CVPR, 2012.

[15] S. Leutenegger, M. Chli, and R. Siegwart,”Brisk: Binary Robust Invari-

ant Scalable Keypoints”, International conference on Computer Vision,

2011.

[16] Vijay Chandrasekhar, Gabriel Takacs, David Chen, Sam S. Tsai, Jatinder

Singh, and Bernd Girod,”Transform coding of image feature descrip-

tors,”, SPIE 7257, Visual Communications and Image Processing, 2009.

[17] Matthew Richardson and Pedro Domingos,”Markov logic networks,”,

International Journal of Machine Learning, Volume 62, Issue 1-2,pp 107-

136, Feb. 2006.

[18] L. Mihalkova, T. Huynh, and R.J. Mooney,Mapping and Revising

Markov Logic Networks for Transfer Learning,, Proc. 22nd Assoc.

for the Advancement of Artiﬁcial Intelligence (AAAI) Conf. Artiﬁcial

Intelligence, pp 608-614, July 2007.

[19] Kok, Stanley and Domingos, Pedro,”Learning the Structure of Markov

Logic Networks”, Proceedings of the 22Nd International Conference on

Machine Learning, pp 441-448, Germany, 2005.

[20] Parag Singla and Pedro Domingos,”Discriminative training of Markov

logic networks”, Proceedings of the international Conf. on Artiﬁcial

Intelligence, 2005.

[21] Wenyuan Dai, Qiang Yang, Gui-Rong Xue and Yong Yu,”Self-taught

Clustering”, Proceedings of the 25th International Conference on

Machine Learning(ICML),2008.

[22] Saˇso Dˇzeroski”Multi-relational Data Mining: An Introduction”,

SIGKDD Explore Newsletter, Volume 5, Issue 1, July 2003, pp.1-16.

[23] Sinno Jialin Pan, and Qiang Yang,”A Survey on Transfer Learning”,

Knowledge and Data Engineering, IEEE Transactions on , vol.22, no.10,

pp.1345,1359, Oct. 2010

[24] T. Dietterich, L. Getoor, and K. Murphy,”Statistical Relational Learning

and its Connections to Other Fields”, ICML-2004 Workshop on

Statistical Relational Learning (SRL), Banff, Canada, July 2004.

[25] Jing Gao and Wei Fan and Jing Jiang and Jiawei Han, ”Knowledge

Transfer via Multiple Model Local Structure Mapping”, in the 14th

ACM SIGKDD International Conference on Knowledge Discovery and

Data Mining,pp. 283-291,New York, USA,2008

[26] T. Joachims, ”Making large-scale svm learning practical.”, advances

in kernel methods - support vector learning, MIT-Press, 1999. D. Lowe,

”Distinctive Image Features from Scale-Invariant Keypoints”, Interna-

tional Journal of Computer Vision 60(2), pp. 91110, 2004.

[27] A. J. Carlson, C. M. Cumby, J. L. R. Nicholas D. Rizzolo, and D.

Roth,”Snow learning architecture”, Technical report UIUCDCS, 1999.

[28] Cover, T. M. and Thomas, J. A.,”Elements of information theory”,

Wiley-Interscience.

[29] Yoshua Bengio, Pascal Lamblin, Popovici Dan, and Larochelle

Hugo,”Greedy Layer-Wise Training of Deep Networks”, Advances in

Neural Information Processing Systems, 2007.

[30] Fei-Fei L., R. Fergus , and P. Perona ,” Learning generative visual

models from few training examples: an incremental Bayesian approach

tested on 101 object categories”, IEEE. CVPR 2004, Workshop on

Generative-Model Based Vision, 2004.

[31] Grifﬁn, G. Holub, and P. Perona ,” The Caltech 256”, Caltech Technical

Report.

[32] Karthikeyan Vaiapury, Anil Aksay and Ebroul Izquierdo,”GrabcutD:

Improved Grabcut Using Depth Information”, Proceedings of the 2010

ACM Workshop on Surreal Media and Virtual Cloning, pp 57-62, New

York, USA, 2010.

[33] Z. Zivkovic,”Improved adaptive Gaussian mixture model for background

subtraction”, Pattern Recognition, 2004. ICPR 2004. Proceedings of the

17th International Conference on , vol.2, no., pp.28,31 Vol.2, 23-26 Aug.

2004.

[34] Fei Sha and Fernando Pereira,”Shallow parsing with conditional random

ﬁelds”, Proceedings of the 2003 Conference of the North American

Chapter of the Association for Computational Linguistics on Human

Language Technology, Volume 1, 2003.

[35] Yanning Zhang,Zhi-Hua Zhou, Changshui Zhang and Li, Ying,”B-

SIFT: A Highly Efﬁcient Binary SIFT Descriptor for Invariant Feature

Correspondence”, Intelligent Science and Intelligent Data Engineering,

pp 426-433, 2012

[36] S. Zhang, Q. Tian, K. Lu, Q. Huang and W. Gao,”Edge-SIFT: Discrim-

inative binary descriptor for scalable partial-duplicate mobile search”,

IEEE Trans. Image Process., vol. 22, no. 7, pp. 28892902, Jul. 2013.

[37] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Guo.,”Locality-

constrained Linear Coding for Image Classication”, Computer Vision

and Pattern Recognition (CVPR), 2010 IEEE Conference on , vol., no.,

pp.3360,3367, 13-18 June 2010.

[38] L. Seidenari, G. Serra, AD. Bagdanov, and A. Del Bimbo,”Local

Pyramidal Descriptors for Image Recognition”, Pattern Analysis and

Machine Intelligence, IEEE Transactions on , vol.36, no.5, pp.1033,1040,

May 2014.

[39] Liefeng Bo, Xiaofeng Ren, and D. Fox,”Multipath Sparse Coding

Using Hierarchical Matching Pursuit”, Computer Vision and Pattern

Recognition (CVPR), 2013 IEEE Conference on , vol., no., pp.660,667,

23-28 June 2013.

[40] Matthew D Zeiler, and Rob Fergus,”Visualizing and Understanding

Convolutional Networks”, Computer Vision and Pattern Recognition

(CVPR), Arxiv 1311.2901, Nov 28, 2013.

1. [↑](#footnote-ref-1)