Knowledge transfer in learning to recognize visual objects classes

Li Fei-Fei
Electrical and Computer Engineering Dept. & Beckman Institute
University of Illinois Urbana-Champaign (UIUC)
feifeili@uiuc.edu

***Abstract*— Learning to recognize of object classes is one of the most important functionalities of vision. It is estimated that humans are able to learn tens of thousands of visual categories in their life. Given the photometric and geometric variabilities displayed by objects as well as the high degree of intra-class variabilities, we hypothesize that humans achieve such a feat by using knowledge and information cumulated throughout the learning process. In recent years, a handful of pioneering papers have applied various forms of knowledge transfer algorithms to the problem of learning object classes. We first review some of these papers by loosely grouping them into three categories: transfer through prior parameters, transfer through shared features or parts, and transfer through contextual information. In the second half of the paper, we detail a recent algorithm proposed by the author. This incremental learning scheme uses information from object classes previously learned in the form of prior models to train a new object class model. Training images can be presented in an incremental way. We present experimental results tested with this model on a large number of object categories.**

***Index Terms*— visual recognition, object classification, Bayesian learning, incremental learning, one-shot learning, knowledge transfer, priors**

I. Introduction

Humans interact with each other and the external world through various sensory and motor systems. Roughly one third of the brain cortex is devoted for functions related to vision. In computer vision, a primary goal is to replicate the visual functionalities that are important to humans. Among them, visual object recognition is a leading area of research in recent years. But it is only till recently that we start looking at the problems of learning object classes under different viewing conditions, and in large numbers (∼hundreds). In this paper, we will mainly discuss methods related to object classification. We define the task of identifying, with the possibility of localizing and/or segmenting a member of a defined object class as *object classification*. In order to learn to recognize object classes, an algorithm is destined to overcome a number of challenges.

Objects could be viewed under different *poses*. We sometimes call it viewpoint transformation such as translation, scaling, affine and projective transformations. Photometrically, different *illumination* conditions would render very different images of the same objects. A large number of variations occur due to the *deformations* of articulated parts in objects like human body and animals. Objects often do not exist by themselves. Embedded in a background, *clutters* pose a big challenge to recognition algorithms, especially in which objects are *occluded* due to the clutter. Beyond these transformations of single objects, object classification is further complicated by the large intra-class variabilities exhibited within a single class of objects. This is especially troubling when the intra-class distance between members of the objects are often larger than the inter-class distances between different classes.

While the problem of object classification seems daunting to the machines, it is well appreciated in psychology that humans possess a superb ability in classifying different objects at fleeting speed and without much attention [16], [20], [22]. Biederman et al. have estimated that there are about $10 \sim 30$ thousands object classes in the world [3]. It is known that a child at age six have learnt roughly the same number of object categories as adults. This suggests that on average, a child can learn about 4 to 5 object classes per day.

Our own experiences also tell us that learning a new object class is not often not a difficult task. Most of us would agree that past experiences with other objects and object classes have already taught us much of what objects are. We have come to know how to handle lighting changes, assume various geometric transformations, deal with clutters and occlusions. To learn a new class, we rely on these previous knowledge, and only simply register the incremental new knowledge of the new object class. This process, formally speaking, is called *knowledge transfer* in machine learning terms. In stark contrast to the superb ability of learning to classify object classes humans possess, most of today's object classification algorithms require a large number of training examples to learn just every single class of objects. Depending on the dimensionality of the image representations and the specific algorithm, the number of training examples would range from hundreds to thousands [21], [24], [27].

The rest of this paper is organized in the following way. In Section II, we give an overview of the existing literature of object classification algorithms that have exploited the idea of knowledge transfer in various forms. We then detail a recent incremental algorithm we developed to learn object classes in Section III. This method uses prior information gathered from previously learned object classes. Finally we conclude this paper in Section IV.

II. KNOWLEDGE TRANSFER IN OBJECT CLASSIFICATION

Knowledge transfer bears many possible forms in the task of learning to classify objects. There is yet to date a unifying definition or framework that encompasses the variety of works proposed by researchers. In fact, while this line of research is certainly gaining momentum in recent years, so far a relatively small number of pioneering works have explored this concept. It is worth noting that the problem of incremental learning through knowledge transfer is intimately related to the problem of one-shot learning. In the case of one-shot learning, a single exemplar of an object class is presented to the algorithm. As we have seen that image data is notoriously complex and high dimensional. Learning from a small number of training examples is often unfeasible due to overfitting effects. Previous knowledge related to the new object class is, therefore, highly important to assist learning in such conditions and to facilitate to overcome such problems. We will see in the sections below that in many of the proposed knowledge transfer frameworks, the authors have in their mind the problem of one-shot learning of object classes. Roughly speaking, we could categorize the different forms of knowledge transfer in the following three ways: *by model parameters*, *by feature or part sharing*, *by contextual information*.

A. transfer by model parameters

Success of an object classification systems relies critically on the models one use for learning and recognition the objects. By models, we mean the characterization that define the task of object classification. For instance, in part-based models [5], [9], [12], [27], a class of objects is defined by parameters that characterize the distributions of the positions and appearances of the object parts.

In [17], the authors propose a model for object classes characterized by deformation matrices. In their paper, the algorithm is to learn, say a digit "4" from a single image example. But given that the algorithm has already learnt letters, say the letter "A" from many training examples, the task is to find an appropriate way to transfer the information in the "A" models to the learning of digit "4" [17]. Here an object class (e.g. "A" class) is modeled by a set of image transformation matrices that act upon each image samples in order to bring them in correspondence with other members of the same class. Suppose a set of transformations T_i is learnt for the "A" class. Given a single example of digit "4", the algorithm first uses T_i to artificially transfer the example image "4" into many different variations. Given the enlarged set of training examples, the new class, digit "4" could be learnt as normal. In this work, the main idea is to use the knowledge of transformations of old classes and transfer it to the new training class.

In another work by [11], the authors propose a way of knowledge transfer from a different angle. The task here is similar to [17], in which the algorithm is to learn a new class of objects, say character "e" given only one training example. In order to overcome the destined overfitting problem when the training number is extremely small, the authors utilize information gained from learning other classes of objects. In the learning process of other characters, a set of "relevant dimensions" are discovered by a nonlinear, kernel based metric learning algorithm [11]. Given this learnt distance function and the single example of a new object class, the algorithm learns a nearest neighbor classifier [11].

Both of the works above tackle the problem of one-shot learning. The essence of these ideas is to obtain model related parameters through training of relevant classes. Under the assumption that the new class of objects is relatively close to the old classes (digits and characters), these model related information could be used to apply to the new class, either through geometric transformations or through discriminative dimensions.

Related to transferring through model parameters, another line of work, inspired by [19], takes advantage of Bayesian learning concepts [7]. The authors combine two learning scenario under one unified framework: learning with few examples by borrowing information from other classes; and learning incrementally by updating models as training images trickles in. We will detail this work in Section III.

B. transfer by sharing features

Another group of works emphasizes the use of sharing features in order to learn new object classes.

One intuitive idea is to recognize the fact that for many different classes of object, there are decomposable parts or features that are shared across these classes. So in a setting of learning more and more new classes, an efficient way to do it is to maximize the repeated usage of these parts. Krempp et al. have formalized this idea in their paper [15].

In this work, the authors focus on the problem of sequentially learning new classes of objects through knowledge transfer, rather than the scenario of limited training numbers. The key idea is to learn new object class from reusable parts obtained from object classes that have already been learned [15].

Recently, another work using transferred information through features and parts has been applied to one shot learning. In their paper, Bart et al. base the learning of diagnostic features for new object class on similar features learnt from other related classes [2]. The authors hypothesize that diagnostic information exists in discriminative patches that distinguish objects in one class from the other. They propose an algorithm that automatically extract such patches by maximizing the mutual information of such features with the objects' class identity. Suppose horses and cows are learnt with ample examples, and useful features found. For for a single image of a dog, diagnostic patches can be derived from visual similar patches in horses and cows. So while there is limited information given in the new class, knowledge from the already learned classes helps the algorithm to smartly select the useful features for the new class recognition.

We have shown that knowledge transfer is important in learning conditions that are either limited by the number of training examples, or set up in such a way that incremental learning is required. But the idea of sharing knowledge is not limited to such condition. Torralba et. al. have shown in that by sharing visual features, one could learn multiple classes of objects more efficiently than otherwise [23].

C. transfer by contextual information

So far we have been mostly looking at works that are aimed to borrow information from one object class to the other. Many of these works [2], [11], [17] rely on the fact that the object classes are relatively similar to each other, such as digits to letters, and dogs to horses. [7] and [23] have taken broader steps and have used relatively different objects classes to transfer useful information.

Objects do not often exist by themselves in the visual world. They are likely to be embedded in a cluttered scene in which different objects interact together. This observation prompts Murphy et al to propose an algorithm that utilizes ambient information from the environ to help learning and recognizing objects [18]. The idea is that certain global information of the scenes, such as frequency distributions, can act as reliable cues for object recognition. They, therefore, propose a conditional random field algorithm to perform object classification by borrowing such contextual information [18].

Another recent work by Hoeim et al. also exploits contextual information to assist object recognition. They observe that object detection can be made easier in cluttered images if one knows something about the geometry of the scenes, such as the camera height, horizon position as well as the support surfaces of the objects ([13] and personal communication). They show in the experiments of pedestrian and car recognition that many false positive detections of these objects can be pruned out by using local and global geometry knowledge.

III. INCREMENTAL LEARNING WITH PRIOR KNOWLEDGE

We present here in more details an algorithm proposed recently to learn new object classes using knowledge from other classes [7]. This work is most similar in spirit to Section II-A. It unifies incremental learning scenario with knowledge transfer from prior models. Due to space limitation, we only highlight here ideas and important points from papers [6]–[8]. The overall system is shown in Fig.1.

A. Overall approach

We first formalize the task of object classification. We start with a learnt object class model and its corresponding model

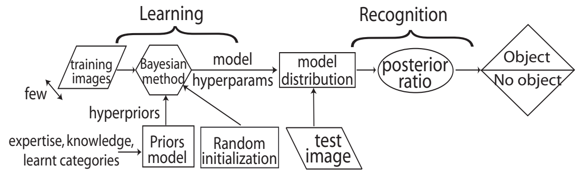
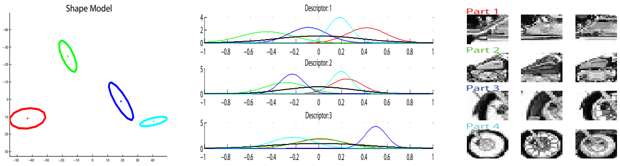


Fig. 1. Schematic illustration of the Bayesian learning algorithm [6], [7].



(a) the shape model (b) the appear. model (c) sample parts

Fig. 2. Illustration of the constellation model

distribution $p(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a set of model parameters for the distribution. Give a new image and we want to decide if it contains an instance of our object class or not. In this query image we have identified N interesting features with locations $\mathcal{X}$, and appearances $\mathcal{A}$. We now make a Bayesian decision, R. For clarity, we express training images through the detected feature locations $\mathcal{X}_t$ and appearances $\mathcal{A}_t$.

$$R = \frac{p(\text{Object}|\mathcal{X}, \mathcal{A}, \mathcal{X}_t, \mathcal{A}_t)}{p(\text{No Object}|\mathcal{X}, \mathcal{A}, \mathcal{X}_t, \mathcal{A}_t)} \quad (1)$$

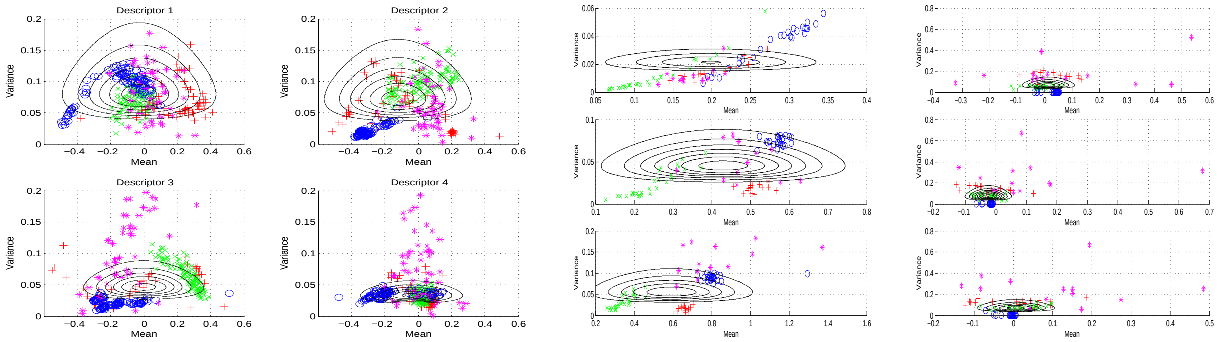
$$= \frac{p(\mathcal{X}, \mathcal{A}|\mathcal{X}_t, \mathcal{A}_t, \text{Object})\ p(\text{Object})}{p(\mathcal{X}, \mathcal{A}|\mathcal{X}_t, \mathcal{A}_t, \text{No object})\ p(\text{No object})} \quad (2)$$

$$\approx \frac{\int p(\mathcal{X}, \mathcal{A}|\boldsymbol{\theta}, \text{Object}) p(\boldsymbol{\theta}|\mathcal{X}_t, \mathcal{A}_t, \text{Object})\ d\boldsymbol{\theta}}{\int p(\mathcal{X}, \mathcal{A}|\boldsymbol{\theta}_{bg}, \text{No Object}) p(\boldsymbol{\theta}_{bg}|\mathcal{X}_t, \mathcal{A}_t, \text{No Object})\ d\boldsymbol{\theta}_{bg}} \quad (3)$$

Note the ratio of $\frac{p(\text{Object})}{p(\text{No Object})}$ in Eq.2 is usually set manually to 1, hence omitted in Eq.3.

The goal of learning in this formulation is to estimate the density of the object models $p(\boldsymbol{\theta}|\mathcal{X}_t, \mathcal{A}_t, \text{Object})$. In other words, in the high dimensional space that characterize the objects, we would like to find the appropriate distribution that defines the extent of where and how the models occupy this space. We do this through the usage of prior knowledge.

At this point, we make a necessary diversion give a more detailed description of what the object class model is. That is to show what these $\boldsymbol{\theta}$ look like. Our chosen representation is based on the *constellation model* [4], [10], [26], [27]. A constellation model consists of a number of parts, each encoding information on both the shape and appearance. The appearance of each part is modeled and the shape of the object is represented by the mutual position of the parts [10]. The entire model is generative and probabilistic, so appearance and shape are all modeled by probability density functions, which are Gaussians. The model is best explained by first considering recognition. We have learned a generative object model, with P parts and a posterior distribution on the parameters $\boldsymbol{\theta}$: $p(\boldsymbol{\theta}|\mathcal{X}_t, \mathcal{A}_t)$ where $\mathcal{X}_t$ and $\mathcal{A}_t$ are the location



(a) (b) (c)

Fig. 3. A visualization of the prior parameter density, estimated from ML models of spotted cats (green $\times$'s), faces (red $+$'s) and airplanes (blue $\circ$'s). Models trained on background data are shown as magenta $*$'s but are not used in estimating the prior density. In all figures the mean is plotted on the x-axis and the variance on the y-axis. **(a)** Appearance parameter space for the first 4 descriptors. **(b)** X component of the shape term for each of the non-landmark model parts. **(c)** Y component of shape. This figure is best viewed in color with magnification.

and appearances of interesting features found in the training data. Recall the Bayesian decision rule in Eq.1 to 3. We assume that all non-object images can also be modeled by a background with a single set of parameters $\boldsymbol{\theta}_{bg}$ which are fixed. The ratio of the priors may be estimated from the training set or set manually (usually to 1). Our decision then requires the calculation of the ratio of the two likelihood functions. In order to do this, the likelihoods may be factored as follows:

$$p(\mathcal{X},\mathcal{A}|\,\boldsymbol{\theta}) = \sum_{\mathbf{h}\in H} p(\mathcal{X},\mathcal{A},\mathbf{h}|\,\boldsymbol{\theta}) = \sum_{\mathbf{h}\in H} \underbrace{p(\mathcal{A}|\mathbf{h},\boldsymbol{\theta})}_{Appearance}\ \underbrace{p(\mathcal{X}|\mathbf{h},\boldsymbol{\theta})}_{Shape} \quad (4)$$

Since our model only has P (typically 3-7) parts but there are N (up to 100) features in the image, we introduce an indexing variable $\mathbf{h}$ which we call a *hypothesis* which allocates each image feature either to an object or to the background. We show in Fig.2 an example of the constellation model learned for the motorbike class.

Appearance. Each feature's appearance is represented as a point in some appearance space. Each part p has a Gaussian density (denoted by $\mathcal{G}$) within this space, with mean and precision parameters $\boldsymbol{\theta}^{\mathcal{A}}_p = \{\mu^{\mathcal{A}}_p, \Gamma^{\mathcal{A}}_p\}$ which is independent of other parts' densities.

Shape. The shape is represented by a joint Gaussian density of the locations of features within a hypothesis. For each hypothesis, the coordinates of all parts are subtracted off from the left most part coordinates. Additionally, it is scale is used to normalize the constellation. This enables our model to achieve scale and translational invariance. The density has parameters $\boldsymbol{\theta}^{\mathcal{X}} = \{\mu^{\mathcal{X}}, \Gamma^{\mathcal{X}}\}$.

B. *Learning with prior*

The task in learning is to estimate the density $p(\boldsymbol{\theta}|\mathcal{X}_t,\mathcal{A}_t)$. This is done using the Variational Bayes procedure [1], [14], [25]. It approximates the posterior distribution $p(\boldsymbol{\theta}|\mathcal{X}_t,\mathcal{A}_t)$ by $q(\boldsymbol{\theta},\boldsymbol{\omega},\boldsymbol{h})$. ω is the mixture component label and $\boldsymbol{h}$ is the hypothesis. Using Bayes' rule: $q(\boldsymbol{\theta},\boldsymbol{\omega},\boldsymbol{h}) \approx p(\boldsymbol{\theta}|\mathcal{X}_t,\mathcal{A}_t) \propto p(\mathcal{X}_t,\mathcal{A}_t|\boldsymbol{\theta})p(\boldsymbol{\theta})$. The likelihood terms use Gaussian densities and by assuming priors of a conjugate form, in this case a Normal-Wishart, our posterior q-function is also a Normal-Wishart density. The variational Bayes procedure is a variant of EM which iteratively updates the hyper-parameters and latent variables to monotonically reduces the Kullback-Liebler distance between $p(\boldsymbol{\theta}|\mathcal{X}_t,\mathcal{A}_t)$ and $q(\boldsymbol{\theta},\boldsymbol{\omega},\boldsymbol{h})$. Using this approach allows us to incorporate prior information in a systematic way and is far more powerful that a maximum-likelihood approach used in [10]. We first briefly give an overview of the algorithm [6], based on [1], which is a batch learning algorithm. Then we introduce the new incremental version of the algorithm. In order to provide some further intuition of the prior models, we show in Fig.3 a visualization of the prior parameter densities obtained from previously learned object classes (in this case spotted cats, faces and airplanes) for learning a new object class (motorbike).

There are two stages to learning: an E-step where the responsibilities of the hidden variables are calculated and an M-step where we update the hyperparameters of $q(\boldsymbol{\theta},\boldsymbol{\omega},\boldsymbol{h})$, $\boldsymbol{\Theta} = \{\boldsymbol{\lambda},\boldsymbol{m},\boldsymbol{\beta},\boldsymbol{a},\boldsymbol{B}\}$. The responsibilities for each image n is:

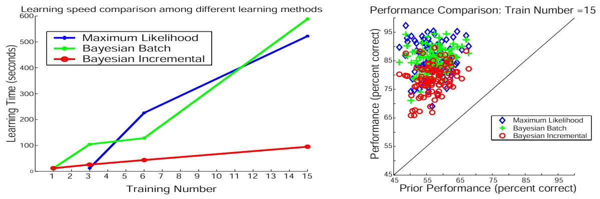
$$\tilde{\gamma}^n_{\omega,h} = \tilde{\pi}_\omega\ \tilde{\gamma}_\omega(\mathcal{X}^n_h)\ \tilde{\gamma}_\omega(\mathcal{A}^n_h) \quad (5)$$

using the update rules given in [1]. The hyperparameters are updated from these responsibilities. This is done by computing the sufficient statistics. While the update rules for the shape components are shown, they are of the same form for the appearance terms. The sufficient statistics, for mixture component ω are calculated as follows::

$$\bar{\pi}_\omega = \frac{1}{N}\sum_{n=1}^{N}\sum_{h=1}^{|H^n|}\gamma^n_{\omega,h} \quad \text{and} \quad \bar{N}_\omega = N\bar{\pi}_\omega \quad (6)$$

$$\bar{\boldsymbol{\mu}}^{\mathcal{X}}_\omega = \frac{1}{\bar{N}_\omega}\sum_{n=1}^{N}\sum_{h=1}^{|H^n|}\gamma^n_{\omega,h}\boldsymbol{\mathcal{X}}^n_h \quad \text{and} \quad (7)$$

$$\bar{\boldsymbol{\Sigma}}^{\mathcal{X}}_\omega = \frac{1}{\bar{N}_\omega}\sum_{n=1}^{N}\sum_{h=1}^{|H^n|}\gamma^n_{\omega,h}(\boldsymbol{\mathcal{X}}^n_h - \bar{\boldsymbol{\mu}}^{\mathcal{X}}_\omega)(\boldsymbol{\mathcal{X}}^n_h - \bar{\boldsymbol{\mu}}^{\mathcal{X}}_\omega)^T \quad (8)$$



(a) (b)

Fig. 4. (a) Average learning time for ML, Bayesian Batch and Bayesian Incremental methods over all 101 categories.(a) Performance comparison between ML, Bayesian Batch and Bayesian Incremental methods for all 101 object categories given 15 training images per class. A category has three markers: Red-Circle represents Bayesian Incremental method, Green-Plus Bayesian Batch method and Blue-Diamond Maximum Likelihood method. The x-axis indicates Bayesian method categorization performance with only the prior model. The y-axis indicates categorization performance for each of the three methods.

Note that to compute these, we need the responsibilities from across all images. From these we can update the hyper-parameters (update rules are in [1]).

1) Extension to incremental learning: We now give an incremental version of the update rules, based on Neal and Hinton's adaptation of conventional EM [19]. Let us assume that we have a model with hyper-parameters $\boldsymbol{\Theta} = \{\boldsymbol{\lambda}, \boldsymbol{m}, \boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{B}\}$, estimated using M previous images ($M \geq 0$) and we have N new images ($N \geq 1$) with which we wish to update the model. From the M previous images, we have retained sufficient statistics $\pi^e_\omega, \boldsymbol{\mu}^e_\omega, \boldsymbol{\Sigma}^e_\omega$ for each mixture component ω. We then compute the responsibilities for the new images, i.e. $\gamma^n_{\omega,h}$ for $n = 1 \ldots N$ and from them, the sufficient statistics, $\bar{\pi}_\omega, \bar{\boldsymbol{\mu}}_\omega, \bar{\boldsymbol{\Sigma}}_\omega$ using eqn.'s 6 and 8. In the Incremental M-step we then combine the sufficient statistics from these new images with the existing set of sufficient statistics from the previous M images. Then the overall sufficient statistics, $\hat{\pi}_\omega, \hat{\boldsymbol{\mu}}_\omega, \hat{\boldsymbol{\Sigma}}_\omega$ are computed:

$$\hat{\pi}_\omega = \frac{M\pi^e_\omega + N\bar{\pi}_\omega}{M+N} \tag{9}$$

$$\hat{\boldsymbol{\mu}}_\omega = \frac{M\boldsymbol{\mu}^e_\omega + N\bar{\boldsymbol{\mu}}_\omega}{M+N} \tag{10}$$

$$\hat{\boldsymbol{\Sigma}}_\omega = \frac{M\boldsymbol{\Sigma}^e_\omega + N\bar{\boldsymbol{\Sigma}}_\omega}{M+N} \tag{11}$$

From these we can then update the model hyper-parameters. Note the existing sufficient statistics are not updated within the update loop. When the model converges, the final value of the sufficient statistics from the new images are combined with the existing set, ready for the next update: $\pi^e_\omega = \hat{\pi}_\omega, \boldsymbol{\mu}^e_\omega = \hat{\boldsymbol{\mu}}_\omega, \boldsymbol{\Sigma}^e_\omega = \hat{\boldsymbol{\Sigma}}_\omega$. Initially $M = 0$, so $\pi^e_\omega, \boldsymbol{\mu}^e_\omega, \boldsymbol{\Sigma}^e_\omega$ drop from our equations and our model hyper-parameters are set randomly (within some sensible range).

C. Experiments and results

Due to space limitation, we will not elaborate here the details of the experimental methods as well as some of the implementation technicalities. Information of experiments performed here can be found in [7]. Fig.4 is a summary of overall classification results as well as the learning time tested on the Caltech 101 dataset [7].

Fig.5 shows in details the results from learning the grand-piano object class using our algorithm. As the number of training examples increases, we observe that the shape model more defined and structured with reducing variance. This is expected sinc ehte algorithm should be more and more confident of what is to be learned. Fig.5(c) shows examples of the part appearance that are closest to the mean distribution of the appearance model. Notice that critical features such as keyboards for the piano are successfully learned by the algorithm. Three learning methods' performances are compared in Fig.5(d). The Bayesian methods clearly show a big advantage over the ML method when training number is small. Bayesian Incremental method, however, shows greater performance fluctuations as compared to the Bayesian Bath method. Finally, we show some classified test images, using an incremental model trained from a single image.

IV. Summary

In this paper, we first present a brief summary of some of the methods used in object classification algorithms that exploit the idea of knowledge transfer to learn object classes either incrementally or with a very small training number. Several interesting themes have been explored to transfer knowledge from elsewhere to the targeted object classes. One idea is summarize knowledge from other classes into model parameters. By using Bayesian learning method or other techniques, such knowledge could be transferred into the new learning task. A second idea is to share some fundamental building blocks of the object classes. In the problem of visual recognition, local features or parts are the most natural choices of this method. A third way of transferring knowledge is to use contextual information to improve recognition or help learning. These methods are developed under the assumption that most objects do not exist by themselves in the visual world. Instead, they are often related to their environment either geometrically or semantically, or both. Useful knowledge of the surroundings could therefore help the learning and recognition of objects.

There is still much to be explored in how to utilize knowledge from different sources for learning object classes. For instance, object classes can be organized into a hierarchy of relationships. We often tend to implicitly learn an new class, say the dalmatian dog, by evoking information of its related classes, say dogs, German shepards and so. Information of taxonomy and ontology might play useful role in object classification. There is still much more to investigate in how to use contextual information to infer objects. Other modalities, such as information from texts, audio recordings and so on may also provide useful knowledge.

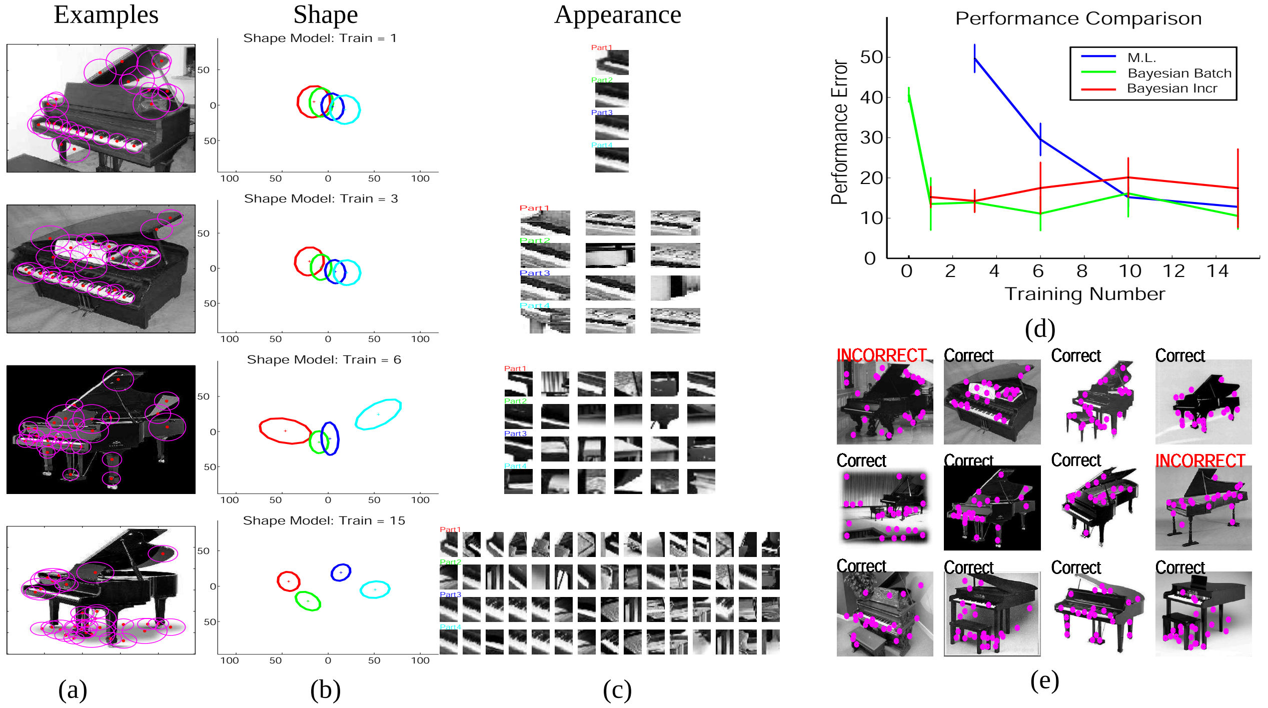


Fig. 5. Results for the "grand-piano" category. Panel (a) shows examples of feature detection. Panel (b) shows the shape models learned at Training Number $= (1, 3, 6, 15)$. Similarly to Fig.3(a), the x-axis represents the x position, measured by pixels, and the y-axis represents the y position, measured by pixels. Panel (c) shows the appearance patches for the model learned at Training Number $= (1, 3, 6, 15)$. Panel (d) shows the comparative results between ML, Bayesian Batch and Bayesian Incremental methods (the error bars show the variation over the 10 runs). Panel (e) shows recognition result for the incremental method at Training Number $= 1$. Pink dots indicate the center of detected interest points. This figure is best viewed in color with magnification.

REFERENCES

[1] H. Attias. Inferring parameters and structure of latent variable models by variational Bayes. In *15th CUAI*, pp 21–30, 1999.

[2] Evgeniy Bart and Shimon Ullman. Cross-generalization: learning novel classes from a single example by feature replacement. In *CVPR*, 2005.

[3] I. Biederman. Recognition-by-Components: A theory of human image understanding. *Psychological Review*, 94:115–147, 1987.

[4] M. Burl and P. Perona. Recognition of planar object classes. In *CVPR*, pp 223–230, 1996.

[5] M. Burl, M. Weber, and P. Perona. A probabilistic approach to object recognition using local photometry and global geometry. In *ECCV*, pp 628–641, 1996.

[6] L. Fei-Fei, R. Fergus, and P. Perona. A Bayesian approach to unsupervised one-Shot learning of object categories. In *ICCV*, pp 1134–1141, October 2003.

[7] L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. In *Workshop on GMBV, CVPR*, 2004.

[8] L. Fei-Fei, R. Fergus, and P. Perona. One-Shot learning of object categories. *IEEE Trans. PAMI*, 2006.

[9] P. Felzenszwalb and D. Huttenlocher. Pictorial structures for object recognition. *IJCV*, 1:55–79, 2005.

[10] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *CVPR*, pp 264–271, 2003.

[11] M. Fink. Object classification from a single example utilizing class relevance pseudo-metrics. In *NIPS*, 2004.

[12] M.A. Fischler and R.A. Elschlager. The representation and matching of pictorial structures. *IEEE Trans. on Computer*, c-22(1):67–92, 1973.

[13] D. Hoiem, A.A. Efros, and M. Hebert. Cgeometric context from a single image. In *ICCV*, 2005.

[14] M. Jordan, Z. Ghahramani, T. Jaakkola, and L. Saul. An introduction to variational methods for graphical models. In *Machine Learning*, v(37), pp 183–233, 1999.

[15] S. Krempp, D. Geman, and Y. Amit. Sequential learning with reusable parts for object detection. Technical report, Johns Hopkins University, 2002.

[16] F.F. Li, R. VanRullen, C. Koch, and P. Perona. Rapid natural scene categorization in the near absence of attention. *PNAS*, 99(14):9596–9601, 2002.

[17] E. Miller, N. Matsakis, and P. Viola. Learning from one example through shared densities on transforms. In *CVPR*, v(1), pp 464–471, 2000.

[18] K. Murphy, A. Torralba, and W.T. Freeman. Using the forest to see the trees:a graphical model relating features, objects and scenes. In *NIPS* , 2004.

[19] R. Neal and G. Hinton. A view of the EM algorithm that justifies incremental, sparse and other variants. In M.I. Jordan, editor, *Learning in Graphical Models*, pages 355–368. Kluwer academic press, Norwell, 1998.

[20] M.C. Potter. Short-term conceptual memory for pictures. *J. of Exp. Psychol. Hum. Learning and Mem.*, 2:509–522, 1976.

[21] H. Schneiderman and T. Kanade. A statistical approach to 3D object detection applied to faces and cars. In *CVPR*, pp 746–751, 2000.

[22] S. Thorpe, D. Fize, and C. Marlot. Speed of processing in the human visual system. *Nature*, 381:520–522, 1996.

[23] A. Torralba, K. Murphy, and W.T. Freeman. Sharing features: efficient boosting procedures for multiclass object detection. In *CVPR.*, 2004.

[24] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *CVPR*, v(1), pp 511–518, 2001.

[25] S. Waterhouse, D. MacKay, and T. Robinson. Bayesian methods for mixtures of experts. In *NIPS*, pages 351–357, 1995.

[26] M. Weber, W. Einhaeuser, M. Welling, and P. Perona. Viewpoint-invariant learning and detection of human heads. In *ICAFGR*, pp 20–27, 2000.

[27] M. Weber, M. Welling, and P. Perona. Unsupervised learning of models for recognition. In *ECCV*, v(2), pages 101–108, 2000.