Active object recognition by view integration and reinforcement learning ⋆

Lucas Paletta [1] Axel Pinz

Institute for Computer Graphics and Vision
Technical University Graz
Münzgrabenstrasse 11, 8010 Graz, Austria

Abstract

A mobile agent with the task to classify its sensor pattern has to cope with ambiguous information. Active recognition involves the observer in a search for discriminative evidence, e.g. by change of its viewpoint. This paper defines the recognition process as a sequential decision problem with the objective to disambiguate initial object hypotheses. Reinforcement learning provides then an efficient method to autonomously develop near-optimal decision strategies in terms of sensorimotor mappings.

The proposed system learns object models from visual appearance and uses a radial basis function (RBF) network for a probabilistic interpretation of the two-dimensional views. The information gain in fusing successive object hypotheses provides a utility measure to reinforce actions leading to discriminative viewpoints. The system is verified in experiments with 16 objects and 2 degrees of freedom in sensor motion. Crucial improvements in performance are gained using the learned in contrast to random camera placements.

Key words: Active recognition; Reinforcement learning; Information fusion; Viewpoint planning.

⋆ We gratefully acknowledge support by the Austrian 'Fonds zur Förderung der wissenschaftlichen Forschung' under grant S7003, and the Austrian Ministry of Science (BMWV Gz. 601.574/2-IV/B/9/96). We wish to thank Manfred Prantl and Hermann Borotschnig for useful discussions.

[1] E-mail: `lpaletta@icg.tu-graz.ac.at`

1 Introduction

Autonomous agents operating in unknown environments are challenged to benefit from the rich complexity in visual sensor information. The recognition of a three-dimensional object in its perceptual context provides important cues in control tasks, e.g. in robot navigation and visual inspection. While object identification by analysis of a single 2-D pattern undergoes permanent improvement, the results always depend on the ambiguity of a particular view, and on the uncertainty in the imaging process.

Active recognition [24,8] provides the framework for collecting evidence until the confidence in a particular hypothesis attains sufficient support. The recognition dynamics is actuated by controlled camera motion, while other effectors, e.g. illumination switches or visual modules, might contribute to discrimination as well. Within this context, object recognition can be formulated as an optimization task with the objective to find a sequence of observations to process the most informative features by minimum costs. In purposive and animate vision [1], task-dependent utility measures are considered to evaluate different solution behaviors. Most applications perform exhaustive search over future states, or restricted analysis of immediate consequences [12,23]. *Optimal* sequential recognition has been outlined with respect to specific 3-D shape models [8] and 2-D objects [3], whereas evidences from the temporal order of multiple 2-D views has been integrated for classification in [27,5,22]. A comprehensive analysis on the information content in successive recognition states is given in [26].

The contribution of this paper is to introduce *reinforcement learning* methods [13] for *active* discrimination of arbitrary 3-D objects, with computational costs being orders of magnitude lower than for exhaustive search. The presented framework complements current methods on active object recognition that use explicit reasoning for viewpoint planning [20,4]. The general concept of learning efficient perception-action cycles was introduced by Whitehead and Ballard [30]. Draper associated the decision task to the selection of visual modules [9]. Recent research on reinforcement learning in computer vision has focused on classifying single views. Peng and Bhanu [18] introduced a closed-loop system that adaptively determines the optimal image segmentation parameters in a recognition task. Bandera et al. [3] applied learning to optimize saccade sequences in a foveal task. In contrast, we provide a working system to select and interprete *multiple views* for 3-D object recognition.

A decision making *agent* (Figure 1) learns directly from visual interaction with its probabilistic environment. In a sensorimotor feedback loop, the agent learns a sensor planning strategy by reinforcing behavior that leads to informative viewpoints. Based on a probabilistic interpretation of the visual pattern, each

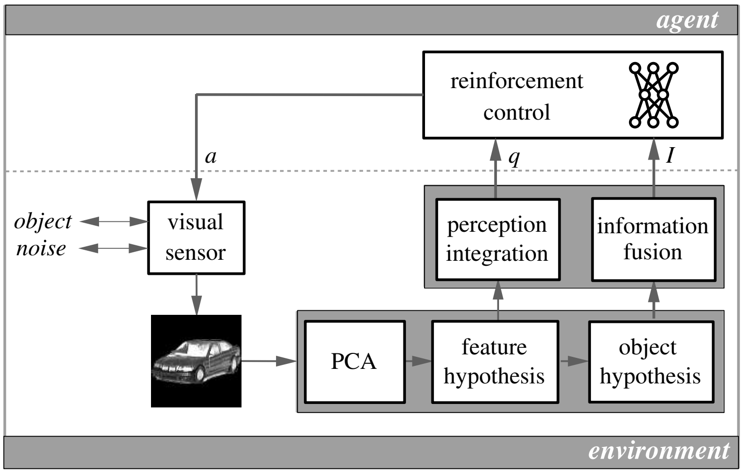


Fig. 1. Closed-loop recognition model: the agent recursively adjusts its discrimination behavior from visual feedback. Actions a and recognition states q are associated with information gains I in the fusion module. This mapping is estimated by a connectionist architecture to recover efficient action selections.

action is evaluated by the degree of disambiguating the current object hypotheses. The proposed recognition system operates *adaptively* and *autonomously*: it automatically acquires the appearance based object model, develops a mapping from 2-D views to object hypotheses, and learns to selectively fuse the view based information. The fusion strategy is stored by the parameters of a connectionist architecture, which enables the agent, in contrast to exhaustive planning methods, to *reactively* apply decisions, i.e. in response to perceptive inputs.

The paper is laid out as follows: section 2 describes the generation of object hypotheses in the appearance based context, section 3 outlines the fusion of sequential beliefs, reinforcement learning for optimizing recognition control is described in section 4, and section 5 presents experimental results. In section 6 we conclude with a discussion on application relevant issues.

2 Object hypotheses from appearance

Appearance based models enable objects of arbitrary shape and texture being automatically learned while they correspond to psychophysical findings about the human visual system [7]. In contrast, geometric models suffer from matching complexity and fail to work for complex shapes [10]. The view based model consists of a collection of raw sensor footprints combining effects of shape and reflectance. Invariances are encoded by neural mappings [19], aspect graphs [27], or parametric manifolds in subspace [16]. The latter method is extended considering the observation statistics in noisy environments.

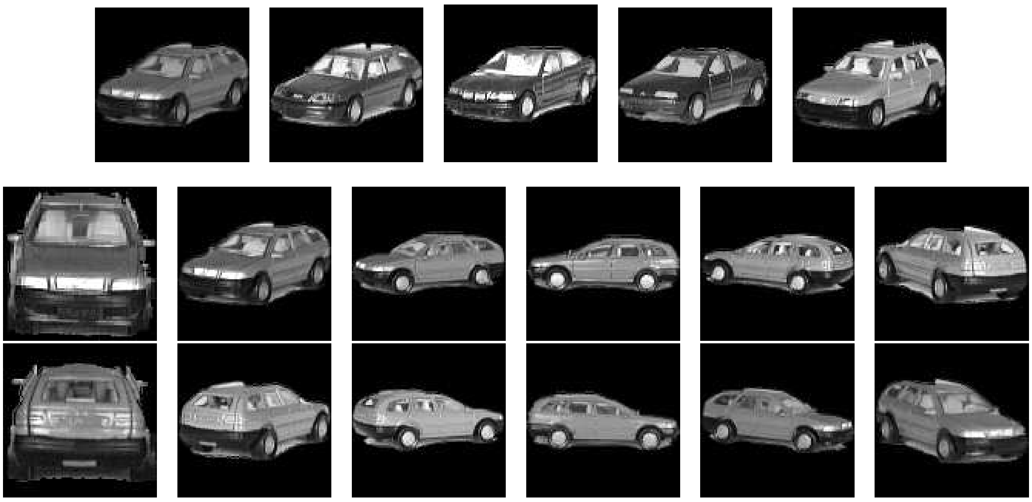


Fig. 2. Illustration of appearance based object representation with 5 objects and 1 degree of freedom. Each object (top row) is modeled by a set of 2-D views (below, for the object top left). This database is a subset of the one depicted in Figure 7.

2.1 Object models in eigenspace

Following the method of Murase and Nayar [16], the highdimensional sensor pattern is projected onto a discriminative subspace of low dimension. Recognition is supported by the property that close points in the subspace correspond to similar appearances of the object.

In a preprocessing step the relevant region is segmented from the background and the image is normalized in brightness and scale to account for small changes in distance and illumination. Each image is then read in a raster scan manner to a vector of pixel brightness values, $\mathbf{x} = (x_1, \ldots, x_P)$, P being the number of pixels. Variation of pose of all objects (Figures 2, 7) results in the collection of images $\mathbf{X} = (\mathbf{x}_{1,1}, \mathbf{x}_{2,1}, \ldots, \mathbf{x}_{\Omega,\Lambda})$, where Ω denotes the number of objects and Λ the number of sampled poses.

Principal component analysis (PCA) or Karhunen-Loéve expansion [11] transforms the observation space to a target subspace, called eigenspace, spanned by the orthogonal basis of eigenvectors of the covariance matrix of all observations $\mathbf{X}$. Only the $k \ll \Omega\Lambda$ most prominent eigenvectors $\mathbf{e}_i$ associated to the largest eigenvalues suffice to represent each image $\mathbf{x} \in \mathbf{X}$ by the approximation $\tilde{\mathbf{x}} = \sum_{s=1}^{k} g_s \mathbf{e}_s$, with $\mathbf{g} = (g_1, \ldots, g_k)^T$, and

$$\mathbf{g} = (\mathbf{e}_1, \ldots, \mathbf{e}_k)^T \mathbf{x} \qquad (1)$$

is the projection of image $\mathbf{x}$ into the eigenspace, where $(\mathbf{e}_1, \ldots, \mathbf{e}_k)^T$ is a $k \times P$ matrix. Under variation of a free parameter φ, the object model consists then of a set of discrete points $\mathbf{g}_\varphi$, that are interpolated to a smoothly varying

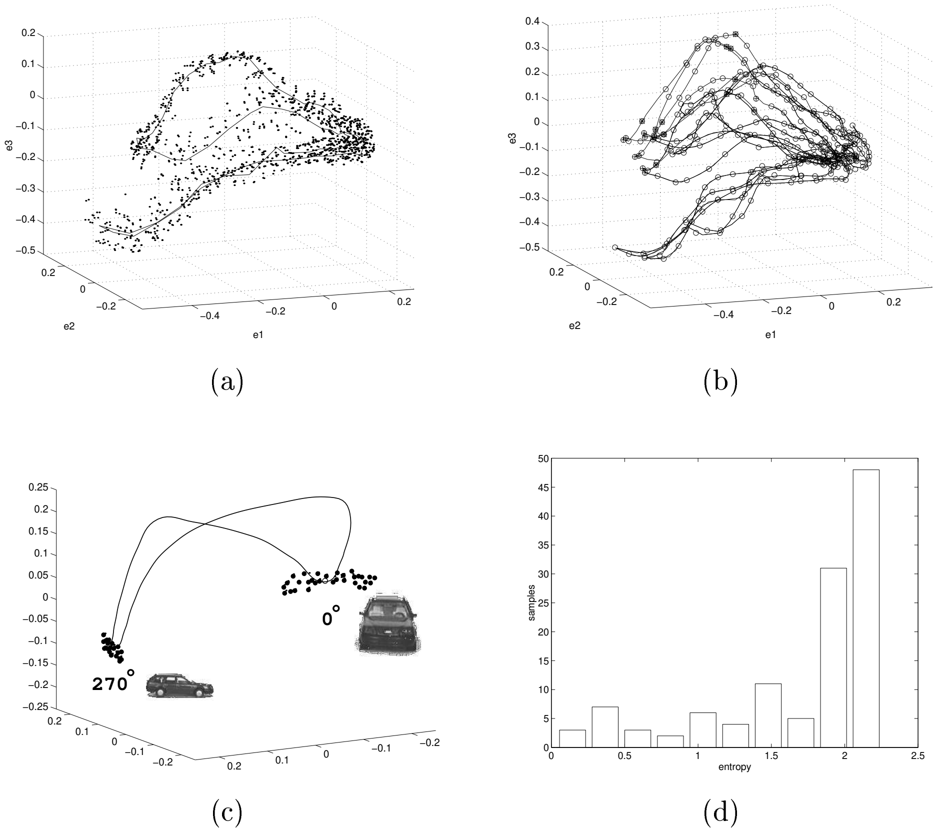


Fig. 3. Ambiguity in the eigenspace representation. (a) depicts the distribution of samples of a particular object. (b) For items carrying similar appearance, the manifolds frequently overlap. (c) Sample variances of individual views differ ($\sigma_{0^\circ} = 0.07, \sigma_{270^\circ} = 0.01$). (d) A histogram over all training samples entropies in the posteriors quantifies the ambiguity in the object views.

manifold.

For the purpose of object recognition, a sample image is projected onto eigenspace, and then assigned the label of the closest manifold [16]. We identify three major sources of uncertainty that make straight classification based on a single image unreliable:

Noise in the visual pathway: Noise in the sensor or the experimental conditions results in eigenperceptions with significant deviations from the model templates (Figure 3a). A remedy is to weight distances according to statistical considerations (section 2.2).

Ambiguity in appearance: Ambiguous information in 2-D views results in overlapping manifolds (Figure 3b). An active step to gain additional evidence from the environment is required (section 3).

Imprecision of object model: To force efficient resource managing, most of the eigenvectors are omitted for projection, inducing a more generalized and potentially ambiguous representation.

Furthermore, the variances of individual views are differing (Figure 3c), rendering generalized models with unified spread [28] unreliable. Based on a probabilistic interpretation of the samples, the entropy in the resulting distributions over object hypotheses (Figure 3d) illustrates the extent of ambiguity in the sensor interpretation.

2.2 Robust object representation

The statistical approach is based upon a 'closed world assumption' that allows to assign to each point in eigenspace a probability distribution over all object hypotheses O_k, $k \in \{1, \ldots, \Omega\}$. A maximum likelihood formulation was outlined by Moghaddam and Pentland [15] with respect to distributions in original sensor space. We discuss an estimator situating classification directly in the target description space.

As an approximation to the continuous object manifold we sample at discrete positions in eigenspace. For each parameter value φ, the distribution of observations is modelled by a Gaussian function,

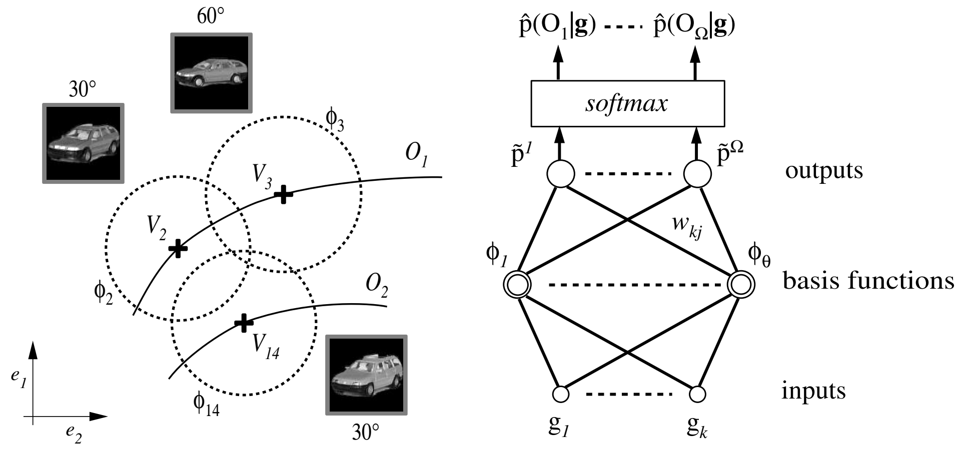
$$p(\mathbf{g}|V_\varphi) = \frac{1}{(2\pi\sigma_\varphi^2)^{d/2}} \exp\left\{-\frac{\|\mathbf{g}-\mu_\varphi\|^2}{2\sigma_\varphi^2}\right\}, \tag{2}$$

with mean vector μ_φ and variance σ_φ^2 estimated from the measurements $\mathbf{g}_\varphi$ (Figure 4a) in eigenspace of dimension d. The local visual error might be alternatively described by elliptical or multimodal Gaussian models, whereas the spherical Gaussian enhances generalization and simplifies matters. A Bayesian estimator of the posterior probabilities is then obtained by

$$p(O_k|\mathbf{g}) = \frac{p(\mathbf{g}|O_k)p(O_k)}{p(\mathbf{g})} = \frac{\sum_{\varphi=1}^{\Lambda} p(\mathbf{g}|V_\varphi)p(V_\varphi|O_k)p(O_k)}{p(\mathbf{g})} \tag{3}$$

where $p(V_\varphi|O_k)$ and $p(O_k)$ are assumed uniformly distributed and the index φ runs over views V_φ belonging only to object O_k.

The distribution model is refined using a mixture of local estimators, as represented by a radial basis function (RBF) architecture [14] (Figure 4b). Within the context of classification, its network weights $\mathbf{W}$ and the $\Theta = \Omega\Lambda$ basis functions ϕ_j, $j = 1 \ldots \Theta$, are interpreted as quantities of the Bayesian estimator,



(a) View features (b) RBF architecture

Fig. 4. (a) Gaussian basis functions ϕ_j model the distribution in each pose, i.e. the view feature V_j. (b) Structural sketch of the RBF mapping from eigenperceptions $\mathbf{g}$ to posterior probabilities $\hat{p}(O_k|\mathbf{g})$.

$$\phi_j \approx \frac{p(\mathbf{g}|V_j)p(V_j)}{p(\mathbf{g})} = p(V_j|\mathbf{g}), \tag{4}$$

$$w_j \approx \frac{p(V_j|O_k)p(O_k)}{p(V_j)} = p(O_k|V_j), \tag{5}$$

$$\tilde{p}^k(\mathbf{g}) \equiv \sum_{j=1}^{\Theta} w_{kj}\phi_j, \tag{6}$$

$$p(O_k|\mathbf{g}) \approx \hat{p}(O_k|\mathbf{g}) = \frac{\exp \tilde{p}^k(\mathbf{g})}{\sum_j \exp \tilde{p}^j(\mathbf{g})}, \tag{7}$$

where the network outputs $\tilde{p}^k$ are normalized by exponential transformation known as *softmax* to ensure a probabilistic interpretation [6]. Note that the posterior probabilities are estimated over *all* views V_j. The corresponding basis functions ϕ_j operate as local *receptive fields* in eigenspace (Figure 4a), with their responses being interpeted as characteristic features extracted from the sensor measurements.

During training, the configuration of basis functions is kept constant whereas the output weights $[\mathbf{W}]_{kj}$ are optimized by gradient descent on the concave error function

$$E(\mathbf{W}) = -\frac{1}{N}\log P(D|\mathbf{W}) = -\frac{1}{N}\log\prod_{\nu=1}^{N}\hat{p}(O^\nu|\mathbf{g}^\nu), \tag{8}$$

where $P(D|\mathbf{W})$ denotes the likelihood of the training set $D = \{(\mathbf{g}^\nu, O^\nu)|\nu = 1,\ldots,N\}$, N is the sample size. The RBF mapping (eq. 7) outperforms the simple Bayes estimator (eq. 3, section 5.2) while its structure provides a framework to minimize complexity with respect to the global error function.

3 Active recognition

Each new measurement promises additional evidence to identify the object in question. The interpretation relying solely on the information in the current image pattern can be updated using probabilistic inference on the complete sequence of observations registered so far. The model associated with the maximum a posteriori (MAP) probability in the fused belief vector denotes then the current classification. Collecting new information about a three-dimensional object is done, e.g., by sampling from multiple viewpoints. The problem arises then which view to select to achieve a maximum of discrimination.

3.1 Information fusion

In a sequence of T recognition states, evidence about the object in question is collected by interpretation of the encountered image patterns, i.e. the posterior distributions $\hat{p}_t \equiv \hat{p}(O_k|\mathbf{g}_t)$ (eq. 7), $1 \leq t \leq T$. The agent updates the confidence in its classification by *fusion* of the current $\hat{p}_T$ with the integrated hypotheses $\hat{p}(O_k|\mathbf{g}_1, \ldots, \mathbf{g}_{T-1})$. Statistical inference is accomplished using recursive Bayesian updating of the posterior probabilities [17]. The overall belief in hypothesis O_k is calculated by Bayesian inversion,

$$p(O_k|\mathbf{g}_1, \ldots, \mathbf{g}_T) = \alpha p(\mathbf{g}_1, \ldots, \mathbf{g}_T|O_k)p(O_k), \tag{9}$$

where α is a normalizing constant. Assuming conditional independence of the observations with respect to O_k implies

$$p(O_k|\mathbf{g}_1, \ldots, \mathbf{g}_T) = \alpha p(O_k) \prod_{t=1}^{T} p(\mathbf{g}_t|O_k) \equiv \bar{p}_T^k. \tag{10}$$

The conditional probabilities $p(\mathbf{g}_t|O_k)$ are recovered via inversion from the posteriors $p(O_k|\mathbf{g}_t)$, which are estimated by the RBF architecture (eq. 7).

The Shannon entropy $H(\bar{p}_T) = -\sum_{k=1}^{\Omega} \bar{p}_T^k \log \bar{p}_T^k$ appears an appropriate measure to quantify the ambiguity in the current state of belief, or recognition state s_T. Consequently, the entropy loss $\mathcal{I}_T$ between two subsequent distributions $\bar{p}$,

$$\mathcal{I}_T = H(\bar{p}_{T-1}) - H(\bar{p}_T), \tag{11}$$

reflects the improvement in the recognition process. $\mathcal{I}_T$ is used to define the utility $\mathcal{U}(s_T, a)$ in taking action a, and to evaluate its future consequences $\mathcal{I}_t$, $t > T$. A predefined threshold on the entropy level determines the stopping criterion for the sequence of actions within a single recognition trial.

3.2 Viewpoint planning

In the context of object recognition, viewpoint planning consists of selecting the observer motion, i.e. the action that promises a maximum of discrimination, with respect to the current recognition state. Crucial improvements in performance are gained when using planned in contrast to random camera motion (section 5).

From a certain recognition state s_T, the utilities $\mathcal{U}$ of all possible actions a are compared to select the action of maximum evaluation,

$$a = \arg\max_{a'} \mathcal{E}(\mathcal{U}(s_T, a')), \tag{12}$$

$\mathcal{E}(\cdot)$ is the expectation operator. Explicit evaluation [4] of the expected utility is based upon the confidence in the current viewpoint $p(V_j|\mathbf{g}_1, \ldots, \mathbf{g}_T)$ and investigates the entropy loss $\mathcal{E}(\mathcal{I}_T)$ in taking action a from that location by

$$\mathcal{E}(\mathcal{U}(s_T, a)) = \mathcal{U}_1 = \sum_{j=1}^{\Omega\Lambda} p(V_j|\mathbf{g}_1, \ldots, \mathbf{g}_T)\mathcal{E}(\mathcal{I}_T(V_j, a, \mathbf{g}_1, \ldots, \mathbf{g}_T)), \tag{13}$$

where the utility $\mathcal{U}_1$ is defined on the consequence of the next action only. Estimation of $\mathcal{E}(\mathcal{I}_T)$ requires a provisional fusion evaluation over the entire target receptive field V_j,

$$\mathcal{E}(\mathcal{I}_T) = H(o_i|\mathbf{g}_1, \ldots, \mathbf{g}_T) - \int_{s\in\mathcal{S}} p(\mathbf{g}_s|V_j)H(o_i|\mathbf{g}_1, \ldots, \mathbf{g}_T, \mathbf{g}_s)d\mathbf{g}_s, \tag{14}$$

$\mathcal{S}$ is the sample space. The computational complexity for each planning step, assuming precomputation of $p(O_k|\mathbf{g}_s)$ using training samples, is in general $\mathcal{O}(\Omega^2\Lambda\Sigma)$ [4], Σ is the number of training samples per view. However, look ahead the consequences in entire sequence length requires $\mathcal{O}(\Omega^2\Lambda!\Lambda\Sigma)$, if each view is visited only once.

While explicit reasoning on each planning step investigates responses from tentative fusion steps, an efficient strategy can be *learned* from experience in terms of a sensorimotor mapping. Learning a mapping from images to actions moves the load of reasoning about actions to an initial training phase.

The resulting decision function, with the current observation as input, is then evaluated with lower computational costs (section 4).

4 Learning recognition control

In each state of the recognition process, a decision making agent is asked to select an action to drive its classifier towards a reliable decision. E.g., a view planning system provides choices among a set of camera motions that determine a specific sequence of fusion steps. *Learning* to recognize objects means then to *explore* different viewpoints, to quantify consequences in terms of a utility measure, and to *adjust* the control strategy thereafter.

4.1 Reinforcement learning

The Markov decision process (MDP [21]) provides the general framework to outline active object recognition in a multistep decision task with respect to the discrimination dynamics. A MDP is then defined by a tuple $(\mathcal{S}, \mathcal{A}, \delta, \mathcal{I})$ with state recognition set $\mathcal{S}$, action set $\mathcal{A}$, probabilistic transition function δ and reward function $\mathcal{I}$. $\delta : \mathcal{S} \times \mathcal{A} \rightarrow \Pi(\mathcal{S})$ describes a probability distribution over subsequent states, given the action $a \in \mathcal{A}$ executed in state $s \in \mathcal{S}$. In each transition, the agent receives reward according to $\mathcal{I} : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{J}$, $\mathcal{I}_t \in \mathcal{J}$. The agent must act to maximize the utility Q, i.e. the expected discounted reward

$$Q = \mathcal{U}_\infty = \mathcal{E}(\sum_{t=0}^{\infty} \gamma^t \mathcal{I}(s_t, a_t)), \tag{15}$$

where $\gamma \in [0, 1]$ is a constant controlling contributions of delayed reward. Without prior knowledge of δ and Π, reinforcement learning [13] is concerned with obtaining an optimal policy by exploring the utility Q by trial and error. In a complex task, extensive search by explicit evaluation of all possible state sequences appears costly, whereas reinforcement learning focuses on visiting those states that have already proved of relevance. Furthermore, the use of parametric function approximators to estimate the decision function enables implicit generalization over entire regions in state space. After training, the utilities are stored by use of a parameter set $\mathbf{W}$. Thereafter, viewpoint planning consists just in selecting the action a with largest utility estimate $\hat{Q}$ (section 4.2).

The perception of a mobile agent encountering ambiguous, i.e. incomplete information about its environment is naturally defined by a partially observable state. The approaches reported to solve the corresponding decision tasks (POMDPs [13]) work either for small state sets or short windows of observations. Conversely, a recognition state in a MDP framework should hold the Markovian property and thus contain any information about the history of the process needed for future inferences.

The *recognition state* s of the active system is now represented by a vector $\mathbf{s}_T$ that captures the information from successive observations, by the encountered feature responses. Each measurement in eigenspace generates a vector of confidences in the *feature hypotheses*, $\hat{p}(V_j|\mathbf{g}_t)$, which are integrated over the sequence. Consequently the recognition state is dependent on the vector $\mathbf{f}_T = (f_{T,1}, \ldots, f_{T,\Theta})$, each entry denoting the maximum confidence in a view feature V_j, $j \in \{1, \ldots, \Theta\}$, registered so far, $f_{T,j} = \max_{t \leq T} \hat{p}(V_j|\mathbf{g}_t)$. Since the same set of features can be collected over different state sequences, the effect of perceptual aliasing must be prevented, and thus the recognition state vector $\mathbf{s}_T$ is supplemented by the current eigenperception $\mathbf{g}_T$, giving $\mathbf{s}_T = (\mathbf{f}_T, \mathbf{g}_T)$.

4.2 Optimal view selection strategies

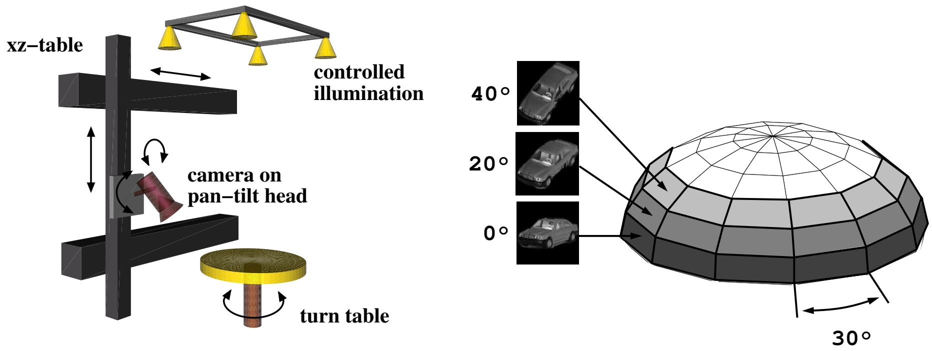
In Q-learning [29], a derivate of the reinforcement learning philosophy, each state-action pair $(\mathbf{s}_t, a_t)$ is associated with the expected utility Q, and the action with the largest Q-value is selected for execution. The estimates $\hat{Q}(\mathbf{s}_t, a_t)$ for the maximum *expected cumulative utility* Q_t received in subsequent steps are recursively updated for residuals of the consistency condition

$$\hat{Q}(\mathbf{s}_t, a_t) \equiv \mathcal{I}_t + \gamma \max_a \hat{Q}(\mathbf{s}_{t+1}, a). \tag{16}$$

The probabilistic, continuous definition of the recognition state (section 4.1) requires the Q-values to be approximated by a universal function approximator Ψ, e.g. by a neural architecture $\hat{Q} = \Psi(\mathbf{s}_t, a_t, \mathbf{W})$. In *residual* Q-learning [2], the free parameters $w = [\mathbf{W}]_i$ of the estimator Ψ are updated by

$$\begin{aligned}\Delta w = -\alpha &\left(\mathcal{I}_t + \gamma \max_a \hat{Q}(\mathbf{s}_{t+1}, a) - \hat{Q}(\mathbf{s}_t, a)\right) \times \\ &\left(\phi\gamma \frac{\partial}{\partial w} \max_a \hat{Q}(\mathbf{s}_{t+1}, a) - \frac{\partial}{\partial w} \hat{Q}(\mathbf{s}_t, a)\right),\end{aligned} \tag{17}$$

where α denotes the learning rate. ϕ is the weighting factor between an algorithm performing gradient descent on the mean squared residual of the consistency condition, i.e. the *Bellman residual*, and a direct method similar to temporal difference learning (TD(0) [2,13]). The algorithm guarantees convergence in the limit to the accurate estimates Q [2].



(a) Actuators (b) View sphere

Fig. 5. (a) Sketch of the actuators in the *Active Vision Lab* setup with 7 degrees of freedom. (b) View sphere by two-dimensional rotation about the object (at sphere center). Images are captured at three levels of vertical elevation and tilt (0°, 20°, 40°) and at 30° azimuth intervals.

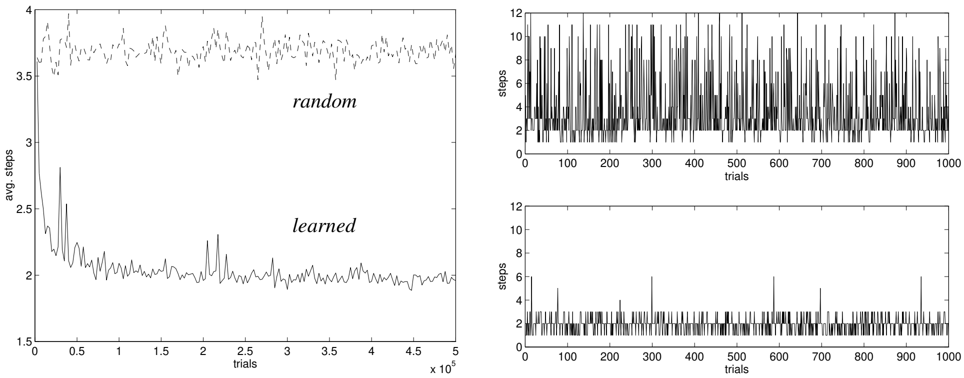
After convergence, the reinforcing controller is assumed to perform a near-optimal strategy with respect to the defined utility measure $\mathcal{U}_\infty$ (eq. 15). In response to the current visual observation, the projection to eigenspace $\mathbf{g}_T$ is fused to a new recognition state $\mathbf{s}_T$ and the agent selects the action $a_T \in \mathcal{A}$ with largest $\hat{Q}(\mathbf{s}_T, a)$,

$$a_T = \arg\max_{a'} \hat{Q}(\mathbf{s}_T, a', \mathbf{W}), \tag{18}$$

i.e. the action which promises maximum reduction of entropy. The planning step consists of a single feed-forward processing of the neural architecture. The computational cost for a network with Θ hidden and $\Omega\Lambda$ input units amounts then to $\mathcal{O}(\Omega\Lambda\Theta)$. Note that the complexity does not increase for arbitrary look-ahead strategies, instead the load is transfered to an extended training phase.

5 Experimental results

The experiments described below were performed in an active vision setup that enables controlled manipulation by multiple degrees of freedom (Figure 5a). Using coordinated sensor translation, tilt, and rotation of a turn-table, it is possible to investigate observer motion on the view sphere (Figure 5b).



(a) Learning curve (b) Performance

Fig. 6. Performance of the learned recognition strategy. (a) Stochastic learning curve comparing test results of random (dotted) and learned strategy (line). (b) depicts a comparison of individual trial lengths between a random (top) and a learned (bottom) policy.

5.1 Illustrative environment

In a first experiment, the proposed recognition system was tested with 5 objects (Figure 2), presented on a computer controlled turn-table by $30°$ intervals, with constant illumination and fixed distance to the camera. The object relevant region was first segmented from the background, rescaled to 100×100 pixels and projected to an eigenspace of dimension 10. For each view, additional samples were collected, emulating segmentation errors by shifting the object region in a randomly selected direction by 3% of the image dimension.

The posterior probabilities are approximated by a RBF network of 60 basis functions, one for each object pose (section 2.2). The estimated ambiguity in the training views is illustrated in Figure 3d. In the experiments, transitions between states were initiated by table rotations a^k of $k \times 30°$, $k \in \{1, \ldots, 11\}$, to enable direct access to the next promising viewpoint. The Q function for each action a^k was estimated by a multilayer perceptron [25], each network with a hidden layer of 5 sigmoidal units. For the stopping criterion the threshold on the entropy level in the posterior distribution on object membership was chosen 0.5.

The stochastic on-line learning process (Figure 6a, with discount factor $\gamma = 0.3$, learning rate $\alpha = 0.1$, and residual coefficient $\phi = 0.8$) converges to a recognition strategy that significantly outperforms random action selection (Figure 6b). In the training phase, the average number of steps per trial converged to ≈ 1.9 in contrast to ≈ 3.6 using a random strategy.

An extended experiment with 16 objects (Figure 7) and 2 degrees of freedom ($j \times 30°$ azimuth, $k \times 20°$ vertical elevation and tilt, $j \in \{1, \ldots, 11\}$, $k \in \{0, \ldots, 2\}$, Figure 5b) in camera motion demonstrates the behavior of a complex recognition system. Objects were chosen of different appearance to increase variances in the eigenspace, and consequently the ambiguity in the representation of similar objects. Object o_8 and o_9 are identical except for a white marker. For comparison, the data were interpreted by a simple Bayes estimator (B, eq. 3) and a RBF net (R, eq. 3). Evaluation of the cross entropy (eq. 8, B: 1.12, R: 0.82) proved superior RBF performance. All data were collected from a real scene, with view specific samples obtained by random deviation in azimuth rotation of $\pm 7°$ amplitude. The fusion strategies were evaluated on-line in the lab environment.

The described system was tested using an alternative input representation that implicitly records features and actions. The sequence of eigenperceptions $\mathbf{g}_t$ is scanned and compiled to a history vector. The resulting vector $\Gamma = (\mathbf{g}_1, \ldots, \mathbf{g}_T)$ was fed to a RBF network with 12 output nodes representing the utility of azimuth actions, and 5 additional units for relative vertical elevation and tilt actions ($-40°$,$-20°$,$0°$,$+20°$,$+40°$). In the experiments, observations in 5-dimensional eigenspace are compiled to a sequence of 10 vectors, giving an input vector of dimension 50. The basis layer of the RBF network consisted of 100 spherical Gaussian basis functions. Using this architecture, the complexity of the network forward pass is $\mathcal{O}(\Theta\Phi)$, using Θ basis units and Φ actions ($\mathcal{O}(\Theta\Lambda)$ for 1 degree of freedom). The entropy stopping criterion on a single trial was chosen 0.1.

Persistent learning again outperforms random strategies, while performance improvement is merely achieved for ambiguous objects (Figure 8a). Policies evaluating multistep look-ahead are observed outperforming strategies concerning payoff collected only at the immediate next step (Figure 8b). Learned selections needed ≈ 1.62 steps per trial, in contrast to ≈ 2.42 steps by random decisions. The learned strategy not only attains a substantial level of discrimination more rapidly (Figure 9a), but also associates the true hypotheses to the captured evidences, as recorded by the recognition rate (Figure 9b). Sample sequences (Figure 10) are displayed for comparison between the two action selection policies. We observe that the learned strategy (a) immediately turns its focus to the discriminating marker, whereas the random selection (b) has to integrate multiple ambiguous views until the fusion operator extracts a reliable classification.

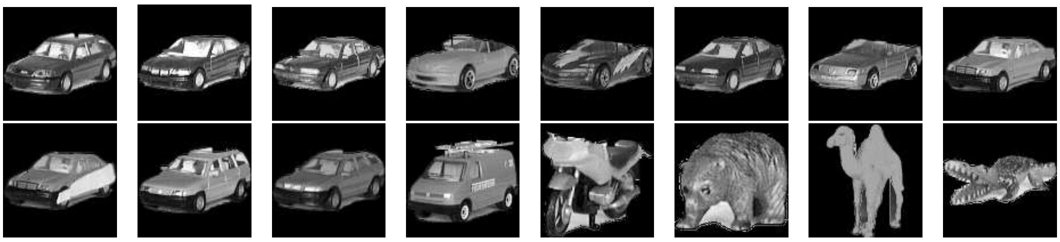
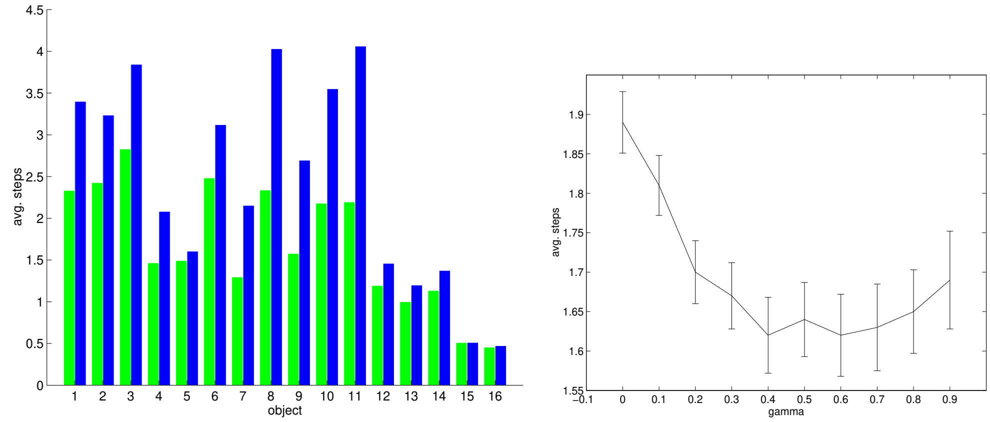


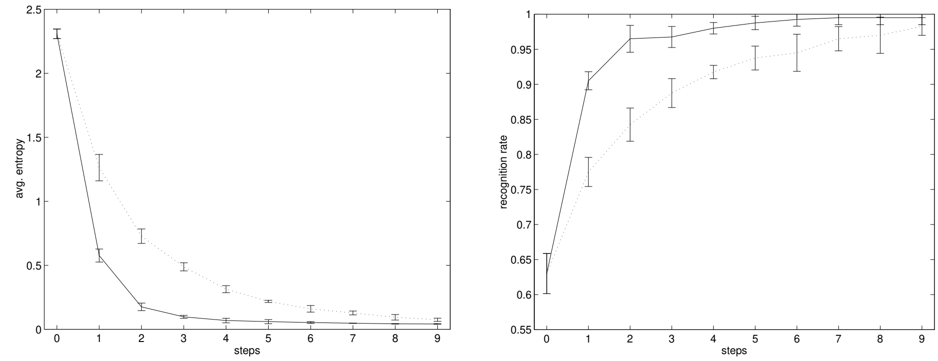
Fig. 7. Extended database consisting of 16 objects (cars, bike, animals). Top row (left to right): objects $o_1 - o_8$, bottom row: objects $o_9 - o_{16}$.



(a) Object statistics

(b) Look-ahead

Fig. 8. Performance statistics. (a) Object dependant performance (gray and black bars denote learned and random sequence lengths, respectively). (b) Multistep look-ahead policies ($\gamma > 0$) improve upon one-step strategies ($\gamma = 0$).



(a) Entropy

(b) Recognition rate

Fig. 9. Convergence rate improvement by learning. (a) demonstrates the speedup in reaching the entropy goal ($H = 0.1$) by a learned (line) versus a random (dotted) strategy (on-line evaluation for object o_{11}). (b) confirms a corresponding acceleration in increasing the recognition rate.

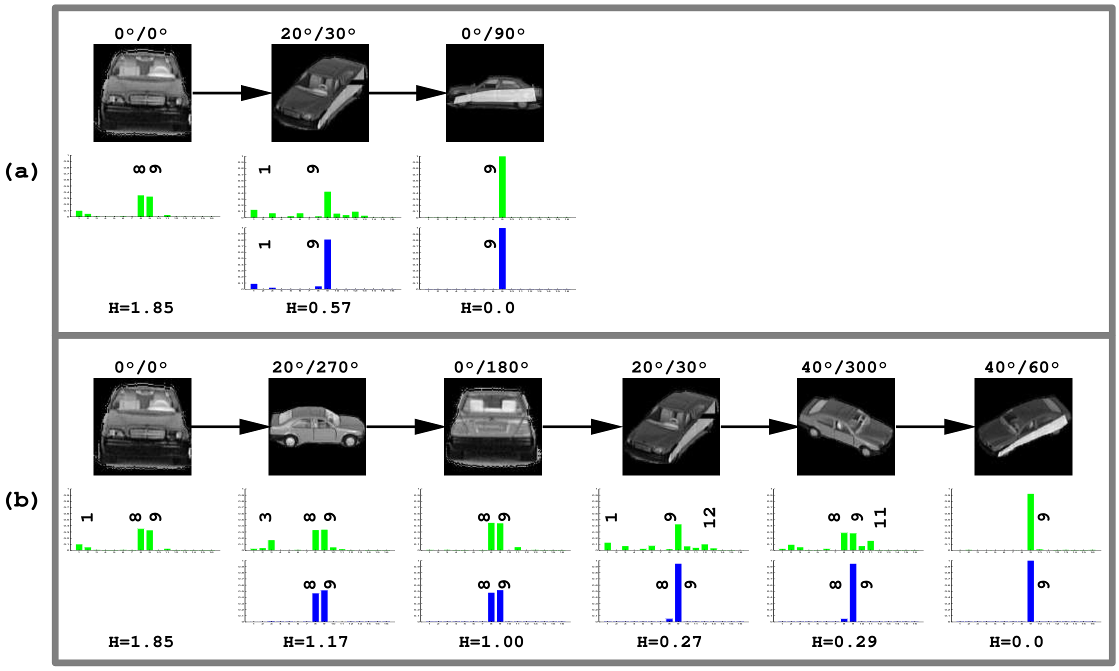


Fig. 10. Sample fusion sequences exhibited on object o_9 (Figure 7). The learned (a) and the random (b) strategy are compared starting from equal viewpoints. Top to bottom rows: poses (vert. elevation and tilt/azimuth), segmented images, *local* posterior distributions (with marked object hypotheses superimposed), *fused* posterior distributions (section 3.1), entropy values H of fused distributions.

6 Conclusion

It has been demonstrated that reinforcement learning provides near-optimal strategies to recognize three-dimensional objects. The appearance based model is extended by probabilistic decision making which proposes a process to actively increase confidence in the classification, under fusion of the visual evidence. Gaussian modelling of view specific receptive fields in connection with a RBF optimization framework allows simple and efficient designs of probabilistic inference systems. Eventually, learning viewpoint planning enables to apply efficient actions in reaction to perceptive inputs, a considerable advantage to explicit reasoning methods in case of large data sets. Furthermore, the learning framework proposes autonomous observers that adjust their recognition behavior to varying environmental conditions.

There are many possible applications of this technique. Autonomous robots take advantage from 3-D object recognition to localize their position in the environment, e.g. for navigation purposes. Visual inspection bears the requirement of high confidence in the decision, e.g. in dangerous environments. The representation model permits objects of any appearance, arbitrary shape and texture, to be included in the database. An important prerequisite is an attention controlling mechanism to segment relevant information from the background, e.g. using optical flow information in image sequences. Limitations will be due to the complexity of the classifier estimating the posterior distribution, and the estimator of action utility. The RBF framework promises relief with currently available methods for global complexity reduction.

Research on learning view selection should develop methods to handle continuous actions, actions towards the perceptual system, and occlusion in a complex scene. The optimization of alternative utility measures, integrating costs on camera motion or processing time, is considered. The described method applies by the same concept to *incremental* sensor motion, i.e. to exploit the information while tracking the object in question. This will open a wide area of applications for autonomous systems in uncertain environments.

References

[1] Y. Aloimonos. Purposive and qualitative active vision. In *International Conference on Pattern Recognition*, pages 346–360, 1990.

[2] L. Baird. Residual algorithms: Reinforcement learning with function approximation. In *12th International Conference on Machine Learning*, pages 30–37, 1995.

[3] C. Bandera, F.J. Vico, J.M. Bravo, M.E. Harmon, and L.C. Baird III. Residual Q-learning applied to visual attention. In *13th International Conference on Machine Learning*, pages 20–27, 1996.

[4] H. Borotschnig, L. Paletta, M. Prantl, and A. Pinz. Active object recognition in parametric eigenspace. In *Proc. 9th British Machine Vision Conference, Southampton, UK*, pages 629–638, 1998.

[5] G. Bradski and S. Grossberg. Fast learning VIEWNET architectures for recognizing three-dimensional objects from multiple two-dimensional views. *Neural Networks*, 8(7/8):1053–1089, 1995.

[6] J. S. Bridle. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In F. Fogelman Soulie and J. Herault, editors, *Neurocomputing: Algorithms, Architectures and Applications*, pages 227–236. Springer Verlag, New York, 1990.

[7] H. H. Bülthoff and S. Edelman. Psychophysical support for a 2-D view interpolation theory of object recognition. In *Proc. of the National Academy of Science*, volume 89, pages 60–64, 1992.

[8] F. G. Callari and F. P. Ferrie. Autonomous recognition: driven by ambiguity. In *Conference on Computer Vision and Pattern Recognition*, pages 701–707, 1996.

[9] B. A. Draper. Learning control strategies for object recognition. In K. Ikeuchi and M. Veloso, editors, *Symbolic Visual Learning*, chapter 3, pages 49–76. Oxford University Press, New York, 1997.

[10] S. Edelman. Computational theories of object recognition. *Trends in Cognitive Sciences*, 1997.

[11] K. Fukunaga. *Introduction to Statistical Pattern Recognition.* Academic Press, New York, 1990.

[12] G. D. Hager. *Task-Directed Sensor Fusion and Planning: A Computational Approach.* Kluwer Academic Publishers, Boston, MA, 1990.

[13] L. P. Kaelbling and A. W. Moore. Reinforcement learning: a survey. *Journal of Artificial Intelligence Research*, 4:237–285, 1996.

[14] D. Lowe. Radial basis function networks. In M. A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 779–782. MIT Press, Cambridge, MA, 1995.

[15] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):696–710, 1997.

[16] H. Murase and S. K. Nayar. Visual learning and recognition of 3-D objects from appearance. *International Journal of Computer Vision*, 14(1):5–24, 1995.

[17] J. Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann, San Francisco, CA, 1988.

[18] J. Peng and B. Bhanu. Closed-loop object recognition using reinforcement learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(2):139–154, February 1998.

[19] T. Poggio and S. Edelman. A network that learns to recognize three-dimensional objects. *Nature*, 317:314–319, 1990.

[20] M. Prantl, H. Borotschnig, H. Ganster, D. Sinclair, and A. Pinz. Object recognition by active fusion. In *Intelligent Robots and Computer Vision, SPIE Conference Proc.*, volume 2904, pages 320–330, Boston, 1996.

[21] M. L. Puterman. *Markov Decision Processes.* John Wiley and Sons, New York, NY, 1994.

[22] R. P. N. Rao. Dynamic appearance-based recognition. In *Conference on Computer Vision and Pattern Recognition*, pages 540–546, 1997.

[23] R. D. Rimey and C. M. Brown. Control of selective perception using bayes nets and decision theory. *International Journal of Computer Vision*, 12(2/3):173–207, 1994.

[24] E. Rivlin, Y. Aloimonos, and A. Rosenfeld. Purposive recognition: an active and qualitative approach. In *SPIE 1611 Sensor Fusion IV*, volume 1611, pages 225–240, 1991.

[25] D. Rumelhart, G. Hinton, and R. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 9 October 1986.

[26] B. Schiele and J. L. Crowley. Transinformation of object recognition and its application to viewpoint planning. *Robotics and Autonomous Systems*, 21(1):95–106, 1997.

[27] M. Seibert and A. M. Waxman. Adaptive 3-D object recognition from multiple views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):107–124, 1992.

[28] M. A. Sipe, D. Casasent, and L. Neiberg. Feature space trajectory representation for active vision. In *Proc. SPIE*, pages 254–265. volume 3077, April 1997.

[29] C.J.C.H. Watkins and P. Dayan. Q-learning. *Machine Learning*, 8(3):279–292, 1992.

[30] S. D. Whitehead and D. H. Ballard. Learning to perceive and act by trial and error. *Machine Learning*, 7:45–83, 1991.