Decision Theoretic Modeling of Human Facial Displays

Jesse Hoey and James J. Little

Department of Computer Science, University of British Columbia
2366 Main Mall, Vancouver, BC, CANADA V6T 1Z4
{jhoey,little}@cs.ubc.ca

Abstract. We present a vision based, adaptive, decision theoretic model of human facial displays in interactions. The model is a partially observable Markov decision process, or POMDP. A POMDP is a stochastic planner used by an agent to relate its actions and utility function to its observations and to other context. Video observations are integrated into the POMDP using a dynamic Bayesian network that creates spatial and temporal abstractions of the input sequences. The parameters of the model are learned from training data using an *a-posteriori* constrained optimization technique based on the expectation-maximization algorithm. The training does not require facial display labels on the training data. The learning process *discovers* clusters of facial display sequences and their relationship to the context automatically. This avoids the need for human intervention in training data collection, and allows the models to be used without modification for facial display learning in any context without prior knowledge of the type of behaviors to be used. We present an experimental paradigm in which we record two humans playing a game, and learn the POMDP model of their behaviours. The learned model correctly predicts human actions during a simple cooperative card game based, in part, on their facial displays.

1 Introduction

There has been a growing body of work in the past decade on the communicative function of the face [1]. This psychological research has drawn three major conclusions. First, facial displays are often purposeful communicative signals. Second, the purpose is not defined by the display alone, but is dependent on both the display and the context in which the display was emitted. Third, the signals are not universal, but vary widely between individuals in their physical appearance, their contextual relationships, and their purpose. We believe that these three considerations should be used as critical constraints in the design of communicative agents able to learn, recognise, and use human facial signals. They imply that a rational communicative agent must learn the relationships between facial displays, the context in which they are shown, and its own utility function: it must be able to compute the utility of taking actions in situations involving purposeful facial displays. The agent will then be able to make

T. Pajdla and J. Matas (Eds.): ECCV 2004, LNCS 3023, pp. 26–38, 2004.

value-directed decisions based, in part, upon the “meaning” of facial displays as contained in these learned connections between displays, context, and utility. The agent must also be able to adapt to new interactants and new situations, by learning new relationships between facial displays and other context.

This paper presents a vision-based, adaptive, Bayesian model of human facial displays. The model is, in fact, a partially observable Markov decision process, or POMDP [2], with spatially and temporally abstract, continuous observations over the space of video sequences. The POMDP model integrates the recognition of facial signals with their interpretation and use in a utility-maximization framework. This is in contrast to other approaches, such as hidden Markov models, which consider that the goal is simply to categorize a facial display. POMDPs allow an agent to make decisions based upon facial displays, and, in doing so, define facial displays by their use in decision-making. Thus, the POMDP training is freed from the curse of labeling training data which expresses the bias of the labeler, not necessarily the structure of the task. The model can be acquired from data, such that an agent can learn to act based on the facial signals of a human through observation. To ease the burden on decision-making, the model builds temporal and spatial abstractions of input video data. For example, one such abstraction may correspond with the wink of an eye, whereas another may correspond to a smile. These abstractions are also learned from data, and allow decision making to occur over a small set of states which are accurate temporal and spatial summarizations of the continuous sensory signals.

Our work is distinguished from other work on recognising facial communications primarily because the facial displays are not defined prior to learning the model. We do not train classifiers for different facial motions and then base decisions upon the classifier outputs. Instead, the training process *discovers* categories of facial displays in the data and their relationships with context. The advantage of learning without pre-defined labels is threefold. First, we do not need labeled training data, nor expert knowledge about which facial motions are important. Second, since the system learns categories of motions, it will adapt to novel displays without modification. Third, resources can be focused on useful tasks for the agent. It is wasteful to train complex classifiers for the recognition of fine facial motion if only simple displays are being used in the agent’s context.

The POMDPs we learn have observations which are video sequences, modeled with mixtures of coupled hidden Markov models (CHMMs) [3]. The CHMM is used to couple the images and their derivatives, as described in Section 3.1. While it is usual in a hierarchical model to commit to a most likely value at a certain level [4,5], our models propagate noisy evidence from video at the lowest level to actions at the highest, and the choice of actions can be probabilistically based upon all available evidence.

2 Previous Work

There are many examples of work in computer vision analysing facial displays [6], and human motion in general [7,4]. However, this work is usually supervised, in that models of particular classes of human motion are learned from labeled training data. There has been some recent research in unsupervised learning of motion models [8,5], but few have attempted to explicitly include the modeling of actions and utility, and none have looked at facial displays. Action-Reaction Learning [9] is a system for analysing and synthesising human behaviours. It is primarily reactive, however, and does not learn models conducive for high level reasoning about the long term effects of actions.

Our previous work on this topic has led to the development of many parts of the system described in this paper. In particular, the low-level computer vision system for instantaneous action recognition was described in [10], while the simultaneous learning of the high-level parameters was explored in [11]. This paper combines this previous work, explicitly incorporates actions and utilities, and demonstrates how the model is a POMDP, from which policies of action can be extracted. Complete details can be found in [12].

POMDPs have become the semantic model of choice for decision theoretic planning in the artificial intelligence (AI) community. While solving POMDPs optimally is intractable for most real-world problems, the use of approximation methods have recently enabled their application to substantial planning problems involving uncertainty, for example, card games [13] and robot control [14]. POMDPs were applied to the problem of active gesture recognition in [15], in which the goal is to model unobservable, non-foveated regions. This work models some of the basic mechanics underlying dialogue, such as turn taking, channel control, and signal detection. Work creating embodied agents has led to much progress in creating agents that interact using verbal and non-verbal communication [16]. These agents typically only use a small subset of manually specified facial expressions or gestures. They focus instead primarily on dialogue management and multi-modal inputs, and have not used POMDPs.

3 POMDPs for Facial Display Understanding

A POMDP is a probabilistic temporal model of an agent interacting with the environment [2], shown as a Bayesian network in Figure 1(a). A POMDP is similar to a hidden Markov model in that it describes observations as arising from hidden states, which are linked through a Markovian chain. However, the POMDP adds actions and rewards, allowing for decision theoretic planning. A POMDP is a tuple $\langle S, A, T, R, O, B \rangle$, where S is a finite set of (possible unobservable) states of the environment, A is a finite set of agent actions, $T : S \times A \rightarrow S$ is a transition function which describes the effects of agent actions upon the world states, O is a set of observations, $B : S \times A \rightarrow O$ is an observation function which gives the probability of observations in each state-action pair, and $R : S \rightarrow \mathcal{R}$ is a real-valued reward function, associating with each state s its immediate utility $R(S)$. A POMDP model allows an agent to predict the long term effects of its actions upon his environment, and to choose actions based on these predictions. Factored POMDPs [18] represent the state, S, using a set of variables, such that the state space is the product of the spaces of each variable. Factored POMDPs

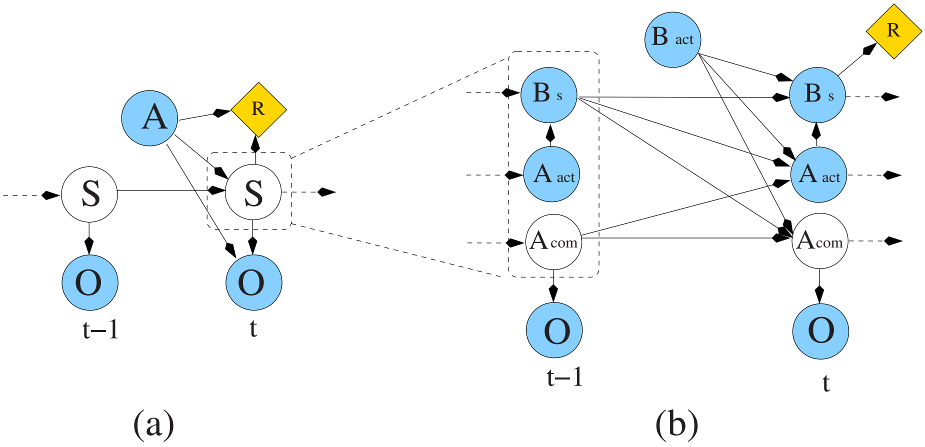


Fig. 1. (a) Two time slices of general POMDP. (b) Two time slices of factored POMDP for facial display understanding. The state, S, has been factored into $\{Bs, Aact, Acom\}$, and conditional independencies have been introduced: Ann's actions do not depend on her previous actions and Ann's display is independent of her previous action given the state and her previous display. These independencies are not strictly necessary, but simplify our discussion, and are applicable in the simple game we analyse.

allow conditional independencies in the transition function, T, to be leveraged. Further, T is written as a set of smaller, more intuitive functions.

Purposeful facial display understanding implies a multi-agent setting, such that each agent will need to model all other agent's decision strategies as part of its internal state [1]. In the following, we will refer to the two agents we are modeling as "Bob" and "Ann", and we will discuss the model from Bob's perspective. Figure 1(b) shows a factored POMDP model for facial display understanding in simple interactions. The state of Bob's POMDP is factored into Bob's private internal state, Bs, Ann's action, $Aact$, and Ann's facial display, $Acom$, such that $S_t = \{Bs_t, Aact_t, Acom_t\}$. While Bs and $Aact$ are observable, $Acom$ is not, and must be inferred from video sequence observations, $\mathbf{O}$. We wish to focus on learning models of facial displays, $Acom$, and so we will use games in which $Aact$ and Bs are fully observable, which they are not in general. For example, in a real game of cards, a player must model the suit of any played card as an unobservable variable, which must be inferred from observations of the card. In our case, games will be played through a computer interface, and so these kinds of actions are fully observable.

The transition function is factored into four terms. The first involves only fully observable variables, and is the conditional probability of the state at time t under the effect of both player's actions: $\Theta_S = P(Bs_t|Aact_t, Bact, Bs_{t-1})$.

[1] This is known as the *decision analytic* approach to games, in which each agent decides upon a strategy based upon his subjective probability distribution over the strategies employed by other players.

The second is over Ann's actions given Bob's action, the previous state, and her previous display: $\Theta_A = P(Aact_t|Bact, Acom_{t-1}, Bs_{t-1})$. The third describes Bob's expectation about Ann's displays given his action, the previous state and her previous display: $\Theta_D = P(Acom_t|Bact, Bs_{t-1}, Acom_{t-1})$. The fourth describes what Bob expects to see in the video of Ann's face, $\mathbf{O}$, given his high-level descriptor, *Acom*: $\Theta_O = P(\mathbf{O}_t|Acom_t)$. For example, for some state of *Acom*, this function may assign high likelihood to sequences in which Ann smiles. This value of *Acom* is only assigned meaning through its relationship with the context and Bob's action and utility function. We can, however, look at this observation function, and interpret it as an $Acom =$ 'smile' state. Writing $C_t = \{Bact_t, Bs_{t-1}\}$, $A_t = Aact_t$, and $D_t = Acom_t$, the likelihood of a sequence of data, $\underset{1,T}{\{\mathbf{OCA}\}} = \{O_1 \ldots O_T, C_1 \ldots C_T, A_1 \ldots A_T\}$, is

$$P(\underset{1,T}{\{\mathbf{OCA}\}}|\Theta) = \sum_k P(\mathbf{O}_T|D_{T,k}) \sum_l \Theta_A \Theta_D P(D_{T-1,l}, \underset{1,T-1}{\{\mathbf{OCA}\}}|\Theta) \tag{1}$$

where $D_{t,k}$ is the k^{th} value of the mixture state, D, at time t. The observations, $\mathbf{O}$, are temporal sequences of finite extent. We assume that the boundaries of these temporal sequences will be given by the changes in the fully observable context state, C and A. There are many approaches to this problem, ranging from the complete Bayesian solution in which the temporal segmentation is parametrised and integrated out, to specification of a fixed segmentation time [4].

3.1 Observations

We now must compute $P(\mathbf{O}|Acom)$, where $\mathbf{O}$ is a sequence of video frames. We have developed a method for generating temporally and spatially abstract descriptions of sequences of facial displays from video [10,12]. We give a brief outline of the method here. Figure 2 shows the model as a Bayesian network being used to assess a sequence in which a person smiles.

We consider that spatially abstracting a video frame during a human facial display involves modeling both the current configuration and dynamics of the face. Our observations consist of the video images, I, and the temporal derivatives, f_t, between pairs of images. The task is first to spatially summarise both of these quantities, and then to temporally compress the entire sequence to a distribution over high level descriptors, *Acom*. We assume that the face region is tracked through the sequence by a separate tracking process, such that the observations arise from the facial region in the images only. We use a flow-based tracker, described in more detail in [12].

The spatial abstraction of the derivative fields involves a projection of the associated optical flow field, v, over the facial region to a set of pre-determined basis functions. The basis functions are a complete and orthogonal set of 2D polynomials which are effective for describing flow fields [12]. The resulting feature vector, Z_x, is then conditioned on a set of discrete states, X, parametrised by normal distributions. The projection is accomplished by analytically integrating the observation likelihood, $P(f_t|X)$, over the space of optical flow fields and

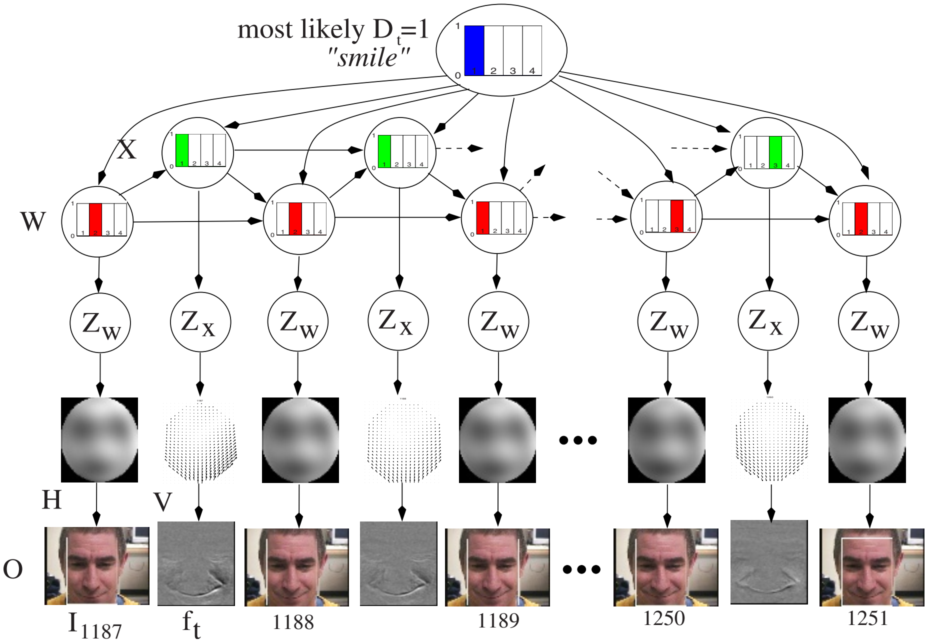


Fig. 2. A person smiling is analysed by the mixture of CHMMs. Observations, O, are sequences of images, I, and image temporal derivatives, f_t, both of which are projected over the facial region to a set of basis functions, yielding feature vectors, Z_x and Z_w. The image regions, H, are projected directly, while it is actually the optical flow fields, V, related to the image derivatives which are projected to the basis functions [10]. Z_x and Z_w are both modeled using mixtures of Gaussians, X and W, respectively. The class distributions, X and W, are temporally modeled as mixture, D, of coupled Markov chains. The probability distribution over D is at the top. The most likely state, $D = 1$, can be associated with the concept "smile". Probability distributions over X and W are shown for each time step. All other nodes in the network show their expected value given all evidence. Thus, the flow field, v, is actually $\langle v \rangle = \int_v vP(v|O)$.

over the feature vector space. This method ensures that all observation noise is propagated to the high level [10]. The abstraction of the images also uses projections of the raw (grayscale) images to the same set of basis functions, resulting in a feature vector, Z_w, which is also modeled using a mixture of normal distributions with mixture coefficients W.

The basis functions are a complete and orthogonal set, but only a small number may be necessary for modeling any particular motion. We use a feature weighting technique that places priors on the normal means and covariances, so that choosing a set of basis functions is handled automatically by the model [10].

At each time frame, we have a discrete dynamics state, X, and a discrete configuration state, W, which are abstract descriptions of the instantaneous dynamics and configuration of the face, respectively. These are temporally abstracted using a mixture of coupled hidden Markov models (CHMM), in which the dynamics and configuration states are interacting Markovian processes. The conditional dependencies between the X and W chains are chosen to reflect the relationship between the dynamics and configuration. This mixture model can be used to compute the likelihood of a video sequence given the facial display descriptor, $P(\mathbf{O}|Acom)$:

$$P(\underset{1,T}{\{\mathbf{O}\}}|D_T) = \sum_{ij} P(f_t|X_{T,i})P(I_t|W_{T,j}) \sum_{kl} \Theta_{Xijk}\Theta_{Wjkl}P(X_{T-1,k}, W_{T-1,l} \underset{1,T-1}{\{\mathbf{O}\}} |D_T) \tag{2}$$

where Θ_X, Θ_W are the transition matrices in the coupled X and W chains, and $P(f_t|X_{T,i}), P(I_t|W_{T,j})$ are the associated observation functions [12]. The mixture components, D, are a set of discrete abstractions of facial behavior. It is important to remember that there are no labels associated with these states at any time during the training. Labels can be assigned after training, as is done in Figure 2, but these are only to ease exposition.

3.2 Learning POMDPs

We use the expectation-maximization (EM) algorithm [17] to learn the parameters of the POMDP. It is important to stress that the learning takes place over the *entire* model simultaneously: both the output distributions, including the mixtures of coupled HMMs, and the high-level POMDP transition functions are all learned from data during the process. The learning classifies the input video sequences into a spatially and temporally abstract finite set, $Acom$, and learns the relationship between these high-level descriptors, the observable context, and the action. We only present some salient results of the derivation here. We seek the set of parameters, Θ^*, which maximize

$$\Theta^* = \arg\max_{\Theta} \left[\sum_{\mathbf{D}} P(\mathbf{D}|\mathbf{O}, \mathbf{C}, \mathbf{A}, \theta') \log P(\mathbf{D}, \mathbf{O}, \mathbf{C}, \mathbf{A}|\Theta) + \log P(\Theta) \right] \tag{3}$$

subject to constraints on the parameters, Θ^*, that they describe probability distributions (they sum to 1). The "E" step of the EM algorithm is to compute the expectation over the hidden state, $P(\mathbf{D}|\mathbf{O}, \mathbf{C}, \mathbf{A}, \theta')$, given θ', a current guess of the parameter values. The "M" step is then to perform the maximization which, in this case, can be computed analytically by taking derivatives with respect to each parameter, setting to zero and solving for the parameter.

The update for the D transition parameter, $\Theta_{Dijk} = P(D_{t,i}|D_{t-1,j}C_{t,k})$, is then

$$\Theta_{Dijk} = \frac{\alpha_{Dijk} + \sum_{t\in\{1\ldots N_t\}|C_t=k} P(D_{t,i}D_{t-1,j}|\mathbf{O}, \mathbf{A}, \mathbf{C}\theta')}{\sum_i \left[\alpha_{Dijk} + \sum_{t\in\{1\ldots N_t\}|C_t=k} P(D_{t,i}D_{t-1,j}|\mathbf{O}, \mathbf{A}, \mathbf{C}\theta')\right]}$$

where the sum over the temporal sequence is only over time steps in which $C_t = k$, and α_{Dijk} is the parameter of the Dirichlet smoothing prior. The summand can be factored as

$$P(D_{t,i}D_{t-1,j}|\mathbf{O}, \mathbf{A}, \mathbf{C}\theta') = \beta_{t,i}\Theta_{A*i*}P(\mathbf{O}_t|D_{t,i})\Theta_{Dijk}\alpha_{t-1,j}$$

where $\alpha_{t,j} = P(D_{t,j}\underset{1,t}{\{\mathbf{OAC}\}})$ and $\beta_{t,i} = P(\underset{t+1,T}{\{\mathbf{OAC}\}}|D_{t,i})$ are the usual forwards and backwards variables, for which we can derive recursive updates

$$\alpha_{t,j} = \sum_k P(\mathbf{O}_t|D_{t,j})\Theta_{A*j*}\Theta_{Djk*}\alpha_{t-1,k} \qquad \beta_{t-1,i} = \sum_k \beta_{t,k}\Theta_{A*k*}P(\mathbf{O}_t|D_{t,k})\Theta_{Dki*}$$

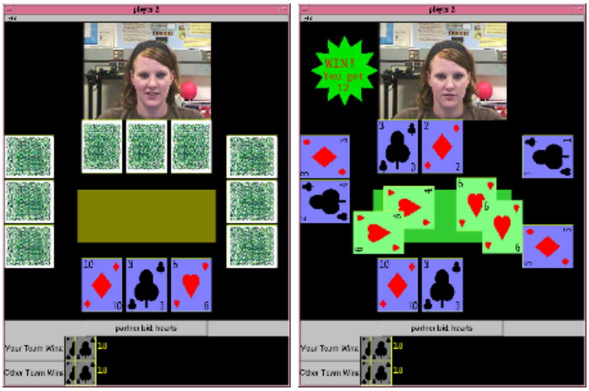
where we write $\Theta_{A*j*} = P(A_t = *|D_{t,j}C_t = *)$ and $P(\mathbf{O}_t|D_{t,i})$ is the likelihood of the data given a state of the mixture of CHMMs (Equation 2). The updates to $\Theta_{Aijk} = P(A_{t,i}|D_{t,j}C_{t,k})$ are $\Theta_{Aijk} = \sum_{t\in\{1...N_t\}|A_t=i\vee C_t=k} \xi_j$, where $\xi_j = P(D_{t,j}|\mathbf{OAC}) = \beta_{t,j}\alpha_{t,j}$. The updates to the j^{th} component of the mixture of CHMMs are weighted by ξ_j, but otherwise is the same as for a normal CHMM [3]. The complete derivation, along with the updates to the output distributions of the CHMMs, including to the feature weights, can be found in [12].

3.3 Solving POMDPs

If observations are drawn from a finite set, then an optimal policy of action can be computed for a POMDP [2] using dynamic programming over the space of the agent's belief about the state, $b(s)$. However, if the observation space is continuous, as in our case, the problem becomes much more difficult. In fact, there are no known algorithms for computing optimal policies for such problems. Nevertheless, approximation techniques have been developed, and yield satisfactory results [14]. Since our focus in this paper is to learn POMDP models, we use the simplest possible approximation technique, and simply consider the POMDP as a fully observable MDP: the state, S, is assigned its most likely value in the belief state, $S = \arg\max_s b(s)$. Dynamic programming updates then consist of computing value functions, V^n, where $V^n(s)$ gives the expected value of being in state s with a future of n stages to go (horizon of n), assuming the optimal actions are taken at each step. The actions that maximize V^n are the n stage-to-go policy (the policy looking forward to a horizon 3 stages in the future). We use the SPUDD solver to compute these policies [18].

4 Experiments

In order to study the relationships between display recognition and action we constrain the structure of an interaction between two humans using rules in a computer game. We then observe the humans playing the game and learn models of the relationships between their facial motions and the states and actions in the game. Subsequent analysis of the learned models reveals how the humans were using their faces for achieving value in the game. Our learning method allows such games to be analysed without any prior knowledge about what facial displays will be used during game play. The model automatically "discovers" what display classes are present. We can also compute policies of action from the models. In the following, we describe our experiments with a simple card game. Results on two other simple games, along with further details on the game here described, can be found in [12].



stage 1 stage 2

Fig. 3. Bob's game interfaces during a typical round. His cards are face up below the "table", while Ann's cards are above it. The current bid is shown below Bob's cards, and the winnings are shown along the bottom. The cards along the sides belong to another team, which is introduced only for motivation. A bid of hearts in stage 1 is accepted by Ann, and both players commit their heart in stage 2.

4.1 Card Matching Game

Two players play the card matching game. At the start of a round, each player is dealt three cards from a deck with three suits (♡,♢,♣), with values from 1 to 10. Each player can only see his own set of cards. The players play a single card simultaneously, and both players win the sum of the cards if the suits match. Otherwise, they win nothing. On alternate rounds (*bidding rounds*), a player has an opportunity to send a confidential *bid* to his partner, indicating a card suit. The *bids* are non-binding and do not directly affect the payoffs in the game. During the other rounds (*displaying rounds*), the player can only see her partner's bids, and then play one of her cards. There is no time limit for playing a card, but the decision to play a card is final once made. Finally, each player can see (but not hear) their teammate through a real-time video link. There are no game rules concerning the video link, so there are no restrictions placed on communication strategies the players can use. The card matching game was played by two students in our laboratory, "Bob" and "Ann" through a computer interface. A picture of Bob's game interface during a typical interaction is shown in Figure 3. Each player viewed their partner through a direct link from their workstation to a Sony EVI S-video camera mounted about their partner's screen. The average frame rate at 320 × 240 resolution was over 28fps. The rules of the game were explained to the subjects, and they played four games of five rounds each. The players had no chance to discuss potential strategies before the game, but were given time to practice.

We will use data from Bob's bidding rounds in the first three games to train the POMDP model. Observations are three or four variable length video sequences for each round, and the actions and the values of the cards of both players, as shown in Table 1. The learned model's performance will then be tested on the data from Bob's bidding rounds in the last game. It is possible to implement a combined POMDP for both bidding and displaying rounds [12].

There are nine variables which describe the state of the game when a player has the bid. The suit of each the three cards can be one of $\heartsuit$, $\diamondsuit$, $\clubsuit$. Bob's actions, *Bact*, can be *null* (no action), or sending a confidential bid ($bid\heartsuit, bid\diamondsuit, bid\clubsuit$) or committing a card ($cmt\heartsuit, cmt\diamondsuit, cmt\clubsuit$). Ann's observed actions, *Aact*, can be *null*, or committing a card. The *Acom* variable describes Ann's communication through the video link. It is one of N_d high-level states, $D = d_1 \ldots d_{N_d}$, of the mixture of CHMMs model described previously. Although these states have no meaning in isolation, they will obtain meaning through their interactions with the other variables in the POMDP. The number of states, (N_d), must be manually specified, but can be chosen as large as possible based on the amount of training data available. The other six, observable, variables in the game are more functional for the POMDP, including the values of the cards, and whether a match occurred or not. The reward function is only based upon fully observable variables, and is simply the sum of the played card values, if the suits match.

4.2 Results

The model was trained with four display states. We inspected the model after training, and found that two of the states (d_1, d_3) corresponded to "nodding" the head, one (d_4) to "shaking" the head, and the last (d_2) to a null display with little motion. Training with only three clusters merges the two nodding clusters together. Figures 4 and 5 show example frames and flows from sequences recognized as d_4 (shake) and as d_1 (nod), respectively. The sequences correspond to the last two rows in Table 1, in which Ann initially refuses a bid of $\diamondsuit$ from Bob, then accepts a bid of $\clubsuit$.

Table 2(a) shows a part of the learned conditional probability distribution over Ann's action, *Aact*, given the current *bid* and Ann's display, *Acom*. We see that, if the bid is null, we expect Ann to do nothing in response. If the bid is $\heartsuit$, and Ann's display (*Acom*) is one of the "nodding" displays d_1 or d_3, then we expect Ann to commit her $\heartsuit$. On the other hand, if Ann's display is "shaking", d_4, then we expect her to do nothing (and wait for another bid from Bob).

The learned conditional probability distribution of Ann's display, *Acom*, at time t, given the previous and current bids, bid_{t-1}, and bid_t, carried two important pieces of information for Bob: First, at the beginning of a round, any bid is likely to elicit a non-null display d_1, d_3 or d_4. Second, a "nodding" display is more likely after a "shaking" display if the bid is changed.

4.3 Computing and Using a Policy

A 3 stage-to-go policy was computed by assuming that the facial display states are observable. There are ten possible values for each card, which expands the state space and makes it more difficult to learn accurate models from limited training data. To reduce this complexity, we approximate these ten values with

Table 1. Log for the first two bidding rounds of one of the training games. A blank means the card values were the same as the previous sequence. Ann's display, *Acom*, is the most likely as classified by the final model.

round	frames	Bob's cards			Ann's cards			*bid*	Bob's act	Ann's act	Ann's display
		♡	♢	♣	♡	♢	♣		*Bact*	*Aact*	*Acom*
1	40-150	3	4	7	2	10	7	-	bid♣	-	d_3
1	151-295							♣	cmt♣	cmt♣	d_1
2	725-827	2	5	2	7	3	8	-	bid♢	-	d_4
2	828-976							♢	bid♣	-	d_4
2	977-1048							♣	cmt♣	cmt♣	d_1

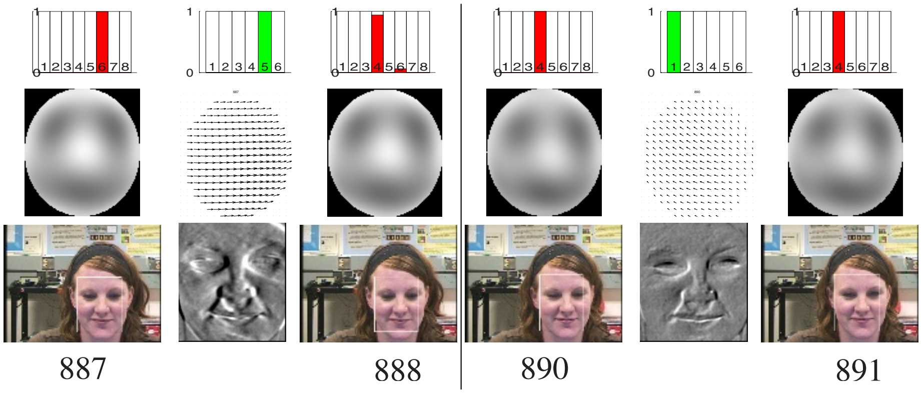


Fig. 4. Frames from the second-to-last row in Table 1. This sequence occurred after Bob had bid ♢, and was recognized as *Acom* = d_4: a head shake. The bottom row shows the original images, I, with tracked face region, and the temporal derivative fields, f_t. The middle row shows the expected configuration, H, and flow field, V (scaled by a factor of 4.0 for visibility). The top row shows distributions over W and X.

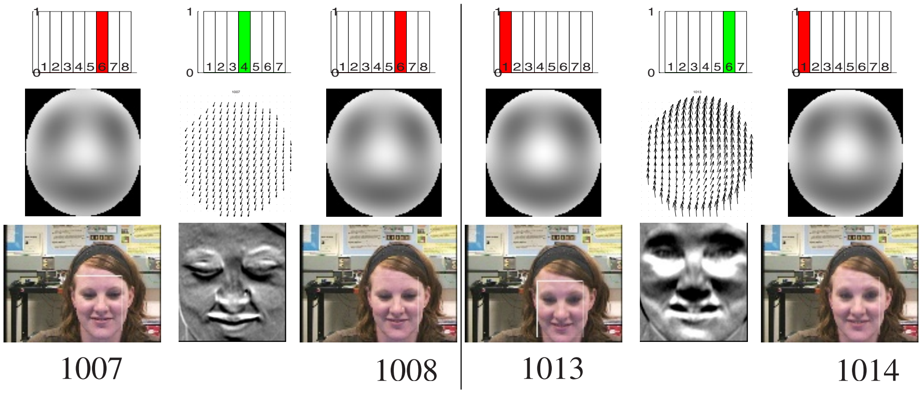


Fig. 5. Frames from the last row in Table 1. This sequence occurred after Bob had made his second bid of ♣ after Ann's negative response to his first bid, and was recognized as *Acom* = d_1: a nod. See Figure 4 for more details.

Table 2. (a) Selected parts of the learned conditional probability distribution over Ann's action, *Aact*, given the current *bid* and Ann's display, *Acom*. Even distributions are because of lack of training data. (b) Selected parts of policy of action in the card matching game for the situation in which $B\heartsuit v = v3, B\diamondsuit v = v3$ and $B\clubsuit v = v1$.

(a)

bid	*Acom*	*Aact*			
		null	*cmt*♡	*cmt*♢	*cmt*♣
null	-	0.40	0.20	0.20	0.20
♡	$d1, d3$	0.20	0.40	0.20	0.20
♡	$d2$	0.25	0.25	0.25	0.25
♡	$d4$	0.40	0.20	0.20	0.20

(b)

bid	*Acom*	policy *Bact*
null	$d1$	*bid*♡
"	$d2, d3$	*bid*♢
"	$d4$	*cmt*♡
♡	$d1, d2, d3$	*cmt*♡
"	$d4$	*bid*♢

three values, $v1, v2, v3$, where cards valued 1-4 are labeled $v1$, 5-7 are $v2$ and 8-10 are labeled $v3$. More training data would obviate the need for this approximation. We then classified the test data with the Viterbi algorithm given the trained model to obtain a fully observable state vector for each time step in the game. The computed policy was consulted, and the recommended actions were compared to Bob's actual actions taken in the game. The model correctly predicted 6/7 actions in the testing data, and 19/20 in the training data. The error in the testing data was due to the subject glancing at something to the side of the screen, leading to a classification as d_4. This error demonstrates the need for dialogue management, such as monitoring of the subject's attention [14].

Table 2(b) shows a part of the policy of action if the player's cards have values $B\heartsuit v = v3, B\diamondsuit v = v3$ and $B\clubsuit v = v1$. For example, if there is no bid on the table, then Bob should bid one of the high cards: hearts or diamonds. If the bid is hearts and Ann nodded or did nothing ($d1, d2$ or $d3$), then Bob should commit his heart. If Ann shook her head, though, Bob should bid the diamond.

Notice that, in Table 2(b), the policy is the same for $Acom = d2, d3$. These states hold similar value for the agent, and could be combined since their distinction is not important for decision making. It is believed that this type of learning, in which the state space is reduced for optimal decision making, will lead to solution techniques for very large POMDPs in the near future [12].

More complex games typically necessitate longer term memory than the Markov assumption we have used. However, POMDPs can accomodate longer dependencies by explicitly representing them in the state space. Further, current research in logical reasoning in first-order POMDPs will extend these models to be able to deal with more complex high-level situations.

5 Conclusion

We have presented an adaptive dynamic Bayesian model of human facial displays in interactions. The model is a partially observable Markov decision process, or POMDP. The model is trained directly on a set of video sequences, and does not need any prior knowledge about the expected types of displays. Without any behavior labels, the model discovers classes of video sequences and their relationship with actions, utilities and context. It is these relationships which define, or give meaning to, the discovered classes of displays. We demonstrate the method on videos of humans playing a computer game, and show how the model is conducive for intelligent decision making or for prediction.

Acknowledgments. Supported by the Institute for Robotics and Intelligent Systems (IRIS), and a Precarn scholarship. We thank our anonymous reviewers, Pascal Poupart, Nicole Arksey and Don Murray.

References

1. Russell, J.A., Fernández-Dols, J.M., eds.: The Psychology of Facial Expression. Cambridge University Press, Cambridge, UK (1997)
2. Kaelbling, L.P., Littman, M.L., Cassandra, A.R.: Planning and acting in partially observable stochastic domains. Artificial Intelligence **101** (1998) 99–134
3. Brand, M., Oliver, N., Pentland, A.: Coupled hidden Markov models for complex action recognition. In: Proc. CVPR (1997), Puerto Rico
4. Oliver, N., Horvitz, E., Garg, A.: Layered representations for human activity recognition. In: Proc. Intl. Conf. on Multimodal Interfaces, Pittsburgh, PA (2002)
5. Galata, A., Cohn, A.G., Magee, D., Hogg, D.: Modeling interaction using learnt qualitative spatio-temporal relations. In: Proc. ECAI. (2002)
6. Tian, Y., Kanade, T., Cohn, J.F.: Recognizing action units for facial expression analysis. IEEE Trans. PAMI **23** (2001)
7. Bregler, C.: Learning and recognising human dynamics in video sequences. In: Proc CVPR (1997), Puerto Rico, 568–574
8. Brand, M.: Structure learning in conditional probability models via an entropic prior and parameter extinction. Neural Computation **11** (1999) 1155–1182
9. Jebara, A., Pentland, A.: Action reaction learning: Analysis and synthesis of human behaviour. In: IEEE Workshop on The Interpretation of Visual Motion. (1998)
10. Hoey, J., Little, J.J.: Bayesian clustering of optical flow fields. In: Proc. ICCV 2003, Nice, France 1086–1093
11. Hoey, J.: Clustering contextual facial display sequences. In: Proceedings of IEEE Intl Conf. on Face and Gesture, Washington, DC (2002)
12. Hoey, J.: Decision Theoretic Learning of Human Facial Displays and Gestures. PhD thesis, University of British Columbia (2004)
13. Fujita, H., Matsuno, Y., Ishii, S.: A reinforcement learning scheme for a multi-agent card game. IEEE Trans. Syst., Man. & Cybern (2003) 4071–4078
14. Montemerlo, M., Pineau, J., Roy, N., Thrun, S., Verma, V.: Experiences with a mobile robotic guide for the elderly. In: Proc, AAAI 2002, Edmonton, Canada.
15. Darrell, T., Pentland, A.: Active gesture recognition using partially observable Markov decision processes. In: 13th IEEE ICPR, Austria (1996)
16. Cassell, J., Sullivan, J., Prevost, S., Churchill, E., eds.: Embodied Conversational Agents. MIT Press (2000)
17. Dempster, A., Laird, N.M., Rubin, D.: Maximum likelihood from incomplete data using the EM algorithm. Journal of the Royal Statistical Society **39** (1977) 1–38
18. Hoey, J., St-Aubin, R., Hu, A., Boutilier, C.: SPUDD: Stochastic planning using decision diagrams. In: Proc. UAI 1999, Stockholm, Sweden