
Behavioural modeling of dynamic facial expression recognition

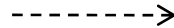
Thomas Robin, Michel Bierlaire, Javier Cruz

28th august 2008

The context



human



Face's video sequence



**Face expression
decision**

- Applications:**
- Driver's attention state;
 - Smart meeting rooms;
 - Human-Machine interfaces.

Objectives

- Model the facial expression recognition made by a person looking at a face video sequence
- Model explicitly the **dynamic process**
- Estimate the model on **behavioural** data (not classification)

Outline

- **Introduction**
- **Features extraction**
- **Data:**
 - Video data bases
 - Internet survey
- **Model:**
 - State transition process
 - Measurement equation
 - Likelihood function
- **Conclusion and Perspectives**

Introduction

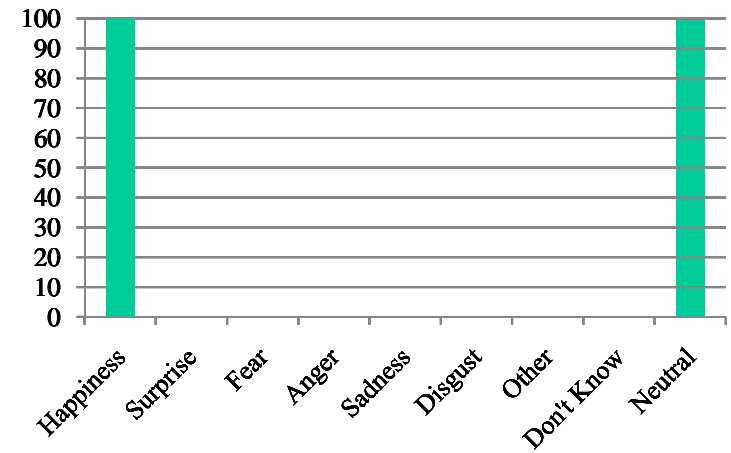
Input:



Model



Output:



Introduction

Input:

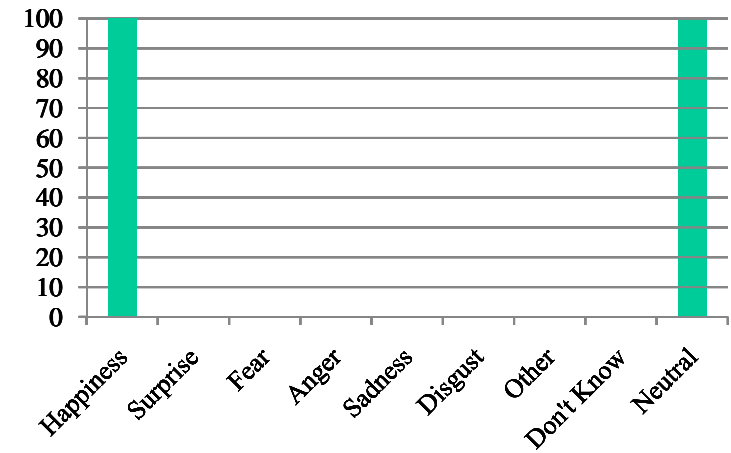
Video



Model



Output:

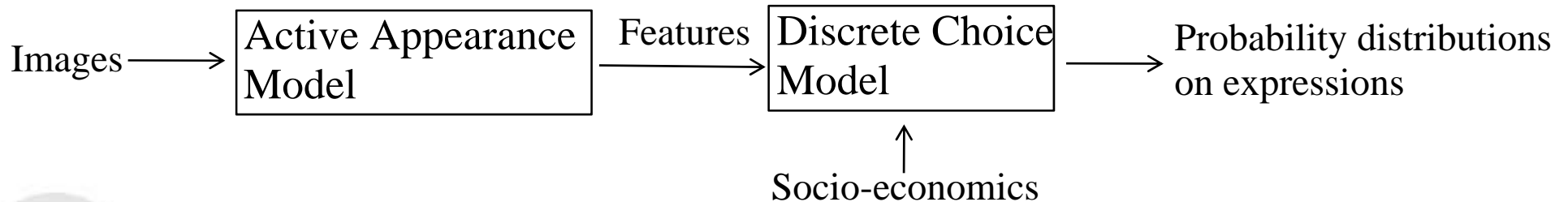


Introduction

- Static version of the work:

M.Sorci, M.Bierlaire, J-P.Thiran, J.Cruz, Th.Robin and G.Antonini (2008).
Modeling human perception of static facial expressions, *8th IEEE Int'l Conference on Automatic Face and Gesture Recognition*.

- 
- Images: Cohn-Kanade database
 - Behavioral data: internet survey



Introduction

- Inspired from dynamic model:

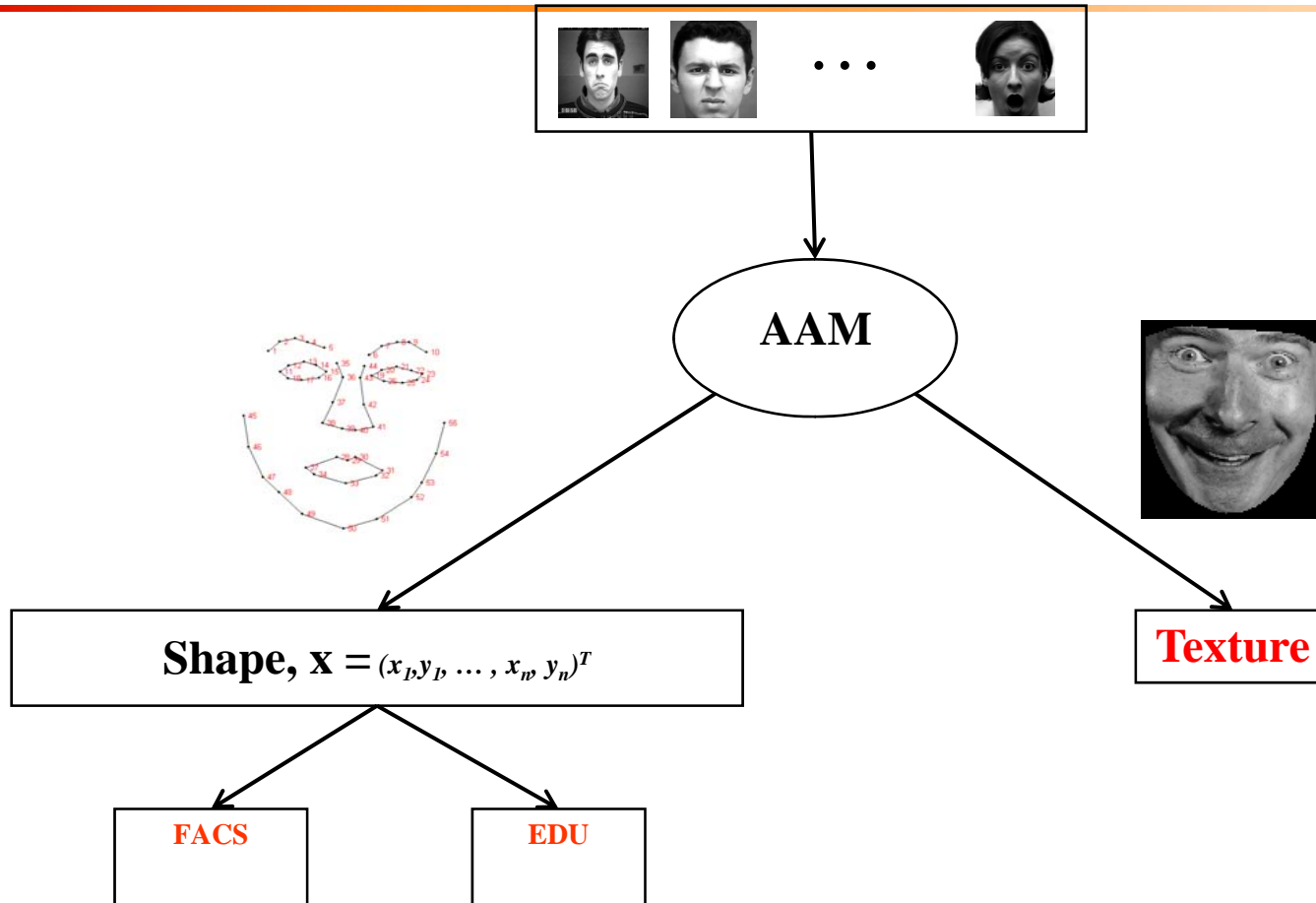
➔ Hidden Markov Model

- State transition process
- Measurement equation

➔ Choudhury, C. F. (2007). *Model Driving Decisions with Latent Plans*, PhD thesis, Massachusetts institute of technology.

- Latent decisions
- Estimation by likelihood maximization

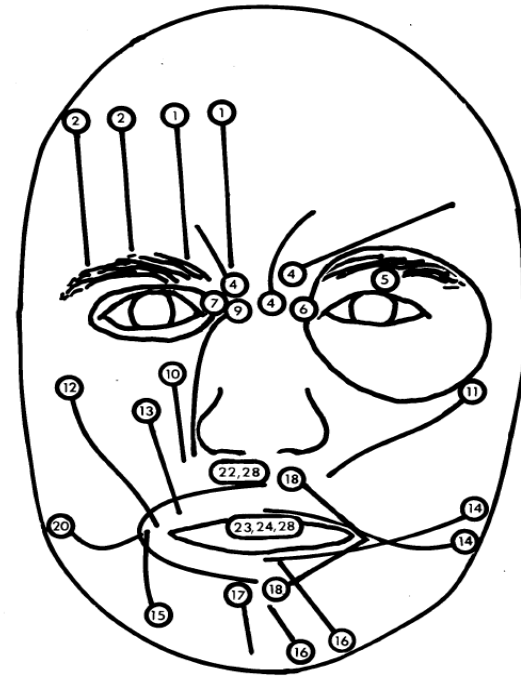
Features extraction: Active Appearance Model



Features extraction: Active Appearance Model

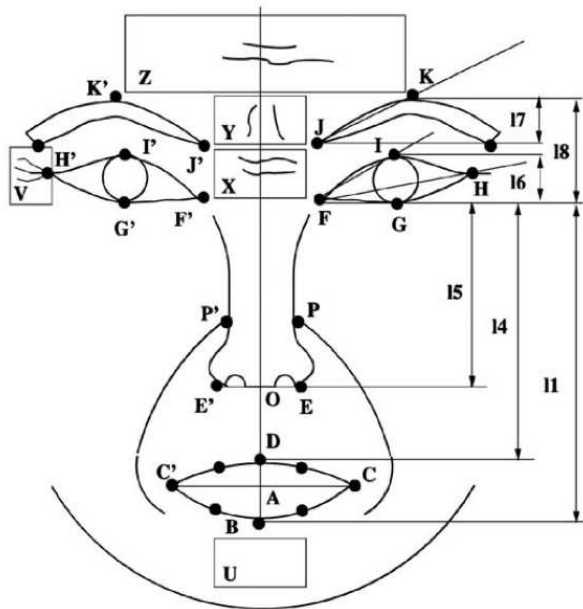
FACS



















- In 1978 Ekman and Friesen developed the Facial Action Coding System
- Measurement units: “Action Units” (AUs)
 - AUs are contractions or relaxations of one or more muscles
 - 46 AUs account for changes in facial expression
 - 12 AUs describe changes in gaze direction and head orientation



The FACS has become the leading standard for measuring facial expressions

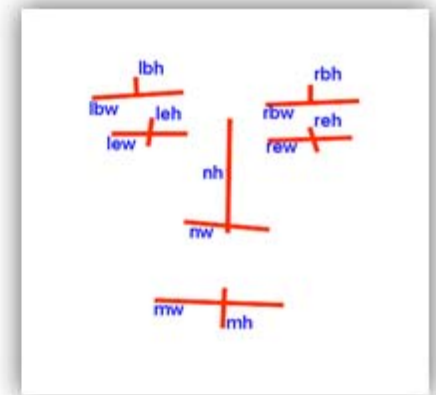
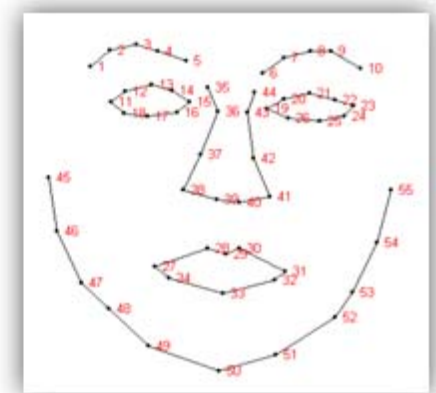
Features extraction: Active Appearance Model FACS



AU1  Inner Brow Raiser	AU2  Outer Brow Raiser	AU4  Brow Lowerer	AU5  Upper Lid Raiser	AU6  Cheek Raiser	AU7  Lid Tightener
AU9  Nose Wrinkler	AU10  Upper Lip Raiser	AU12  Lip Corner Puller	AU15  Lip Corner Depressor	AU16  Lower Lip Depressor	AU17  Chin Raiser
AU20  Lip Stretcher	AU23  Lip Tightener	AU24  Lip Pressor	AU25  Lips part	AU26  Jaw Drop	AU27  Mouth Stretch

Features extraction: Active Appearance Model EDU

- Expression Descriptive Units by Antonini, Sorci, Bierlaire and Thiran in « Discrete Choice Models for Static Facial Expression Recognition »



EDU1	$\frac{lew+rew}{leh+reh}$	EDU8	$\frac{leh+reh}{lbh+rbh}$
EDU2	$\frac{lbw}{lbh}$	EDU9	$\frac{lew}{nw}$
EDU3	$\frac{rbw}{rbh}$	EDU10	$\frac{nw}{mw}$
EDU4	$\frac{mw}{mh}$	EDU11	EDU2 / EDU4
EDU5	$\frac{nh}{nw}$	EDU12	EDU3 / EDU4
EDU6	$\frac{lew}{mw}$	EDU13	EDU2 / EDU10
EDU7	$\frac{leh}{mh}$	EDU14	EDU3 / EDU10

Features extraction: Active Appearance Model Texture



Data: internet survey

- Survey conducted at the address below(English, French, Italian, Spanish):
<http://transp-or2.epfl.ch/videosurvey/>
- Respondents have to:
 - create an account
 - Socioeconomics attributes
 - label some video sequences with expressions
 - observations
- 2 databases of video are used:
 - Cohn-Kanade
 - Technical University Munich (TUM)

Data: video database

- The Cohn-Kanade database

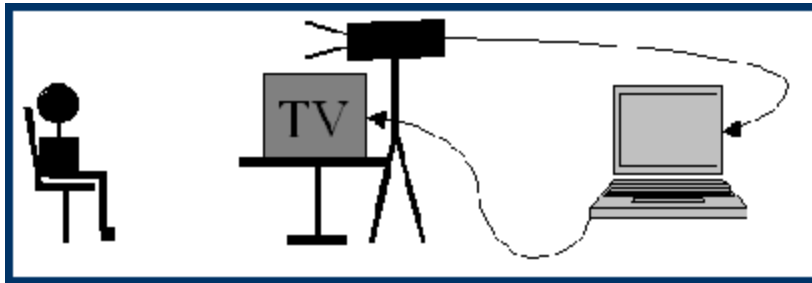
➔ Actors **playing** expressions, according to the Facial Action Coding System (**FACS**)



55 sequences, 11 subjects

Data: video database

- The Technical University Munich database (TUM)
 - ➔ Students faced to a video, natural expressions recorded



399 sequences, 18 subjects

Data: socio-economics

http://transp-or2.epfl.ch/videosurvey/index.php?include=createuser

Google t Search RS Bookmarks PageRank Check AutoLink AutoFill Send to Settings

EPFL
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

LABORATOIRE TRANSPORT ET MOBILITE
FACIAL EXPRESSIONS EVALUATION SURVEY

INTER > TRANSP-OR > Facial Expressions Evaluation Survey

Status and links

Home
You are not connected
English
Suggestions (email us)

E-mail address:
Password:
Login
Forgotten password

Why this socio-economics form?

The socio-economics fields are important in order to segment the labeler population based on those characteristics. The ethnic group is relevant to investigate the choice behavior of people when faced to videos of individuals belonging to the same or to another ethnic group.

Why a username?

An account is required so that you won't have to fill the form anytime you want to participate to the survey. If you find the survey too long you can stop whenever you want by logging off and restart from the first unlabeled image at your next login.

IMPORTANT: The e-mail address is only used to send you a new password, if you have forgotten it.

Create a new user

Birth Year: 0000
Gender: Male Female
Language: English
Studies: High School
Ethnic group: None
Current location: None
Occupational category: None
E-mail address:
Password:
Password Confirmation:
OK

Data: labels

The screenshot shows a web browser window with the URL `http://transp-or2.epfl.ch/videosurvey/index.php?include=navigator`. The browser's address bar and search bar are visible. The page header includes the EPFL logo and the text "LABORATOIRE TRANSPORT ET MOBILITE" and "EVALUATION D'EXPRESSIONS FACIALES". Below the header, there is a navigation menu with "INTER > TRANSP-OR > Evaluation d'Expressions Faciales".

On the left side, there is a sidebar with the following content:

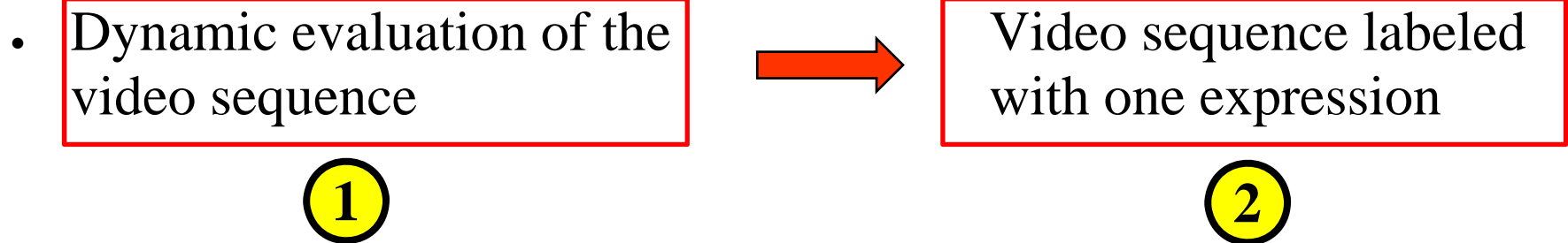
- Statut et liens**
- Accueil**
- Connecté en tant que : thomas.robin@epfl.ch
- [Se déconnecter](#)
- Suggestions (pour nous écrire)**

The main content area features a video player showing a woman's face. Below the video player, there is a questionnaire with the following options:

- Joie
- Surprise
- Peur
- Degout
- Tristesse
- Colere
- Neutre
- Autre
- Je ne sais pas

At the bottom of the questionnaire, there is a "Valider le questionnaire" button and navigation arrows.

Model: introduction



- **①** : modeling the dynamic evaluation
 → The state transition process


- **①** → **②** : Link between dynamic evaluation and label
 → The measurement equation

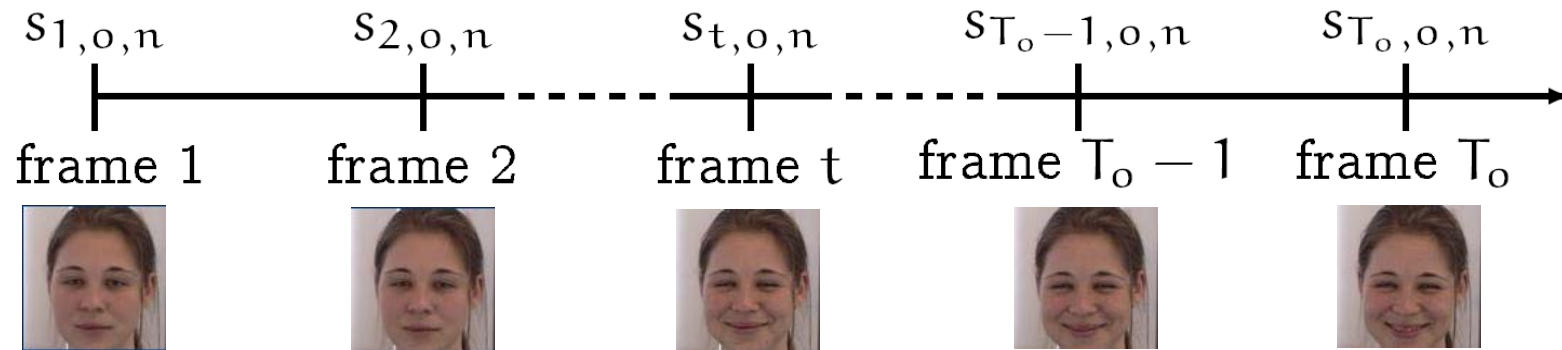
Model: state transition process

- Modeling of the dynamic evaluation of a video sequence
 - i : expression
 - n : respondent
 - N : total number of respondents
 - O_n : number of video sequences labelled by the respondent n
 - t : frame of a video sequence
 - o : video sequence
 - T_o : total number of frames in the video sequence o

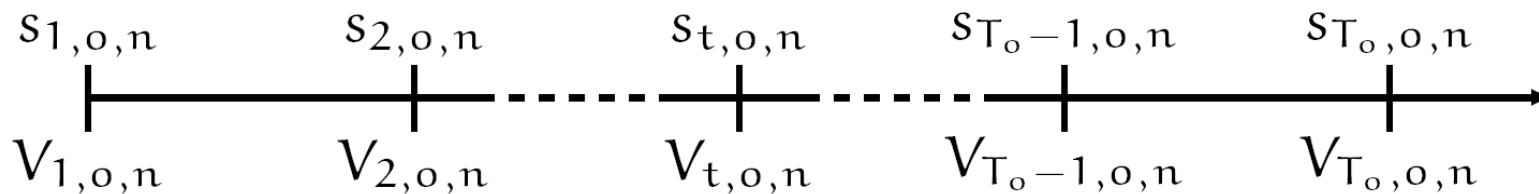
Model: state transition process

- The video sequence \mathbf{o} watched by the respondent \mathbf{n} :

 a state $S_{t,o,n}$ associated to each frame t



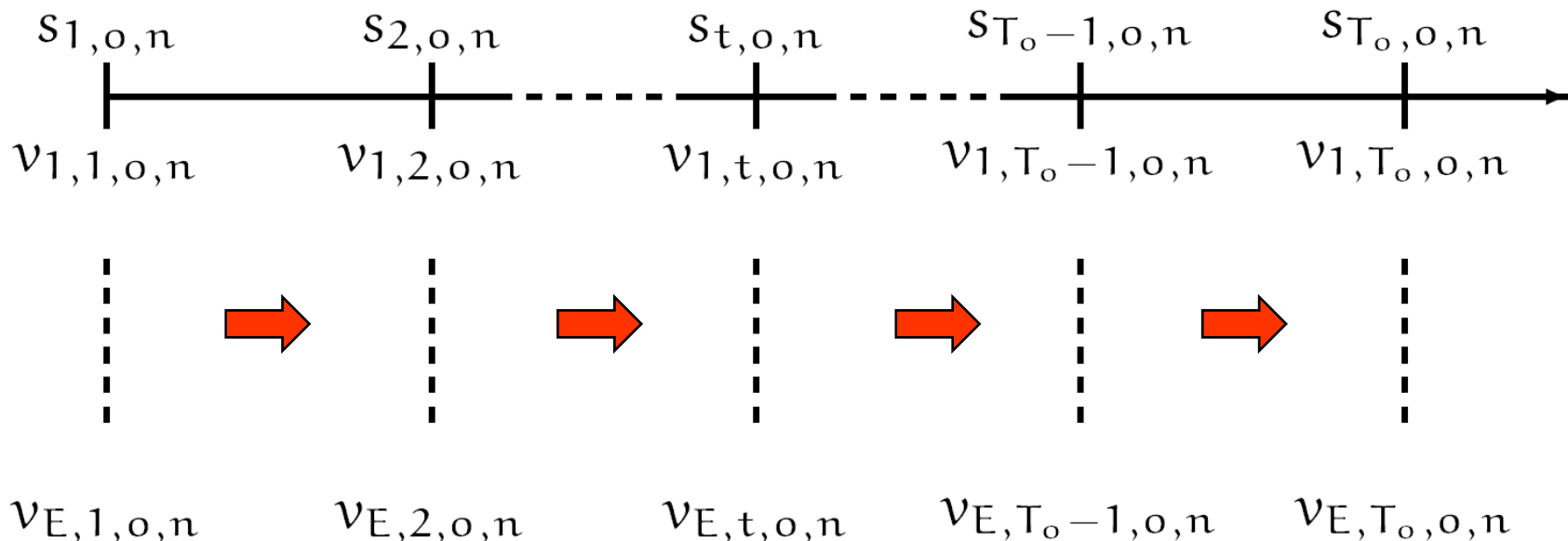
- A vector of utility functions $V_{t,o,n}$ associated to the state $S_{t,o,n}$



Model: state transition process

- Model the transition between the states $\{s_{t,o,n}\}_{t \leq T_o}$

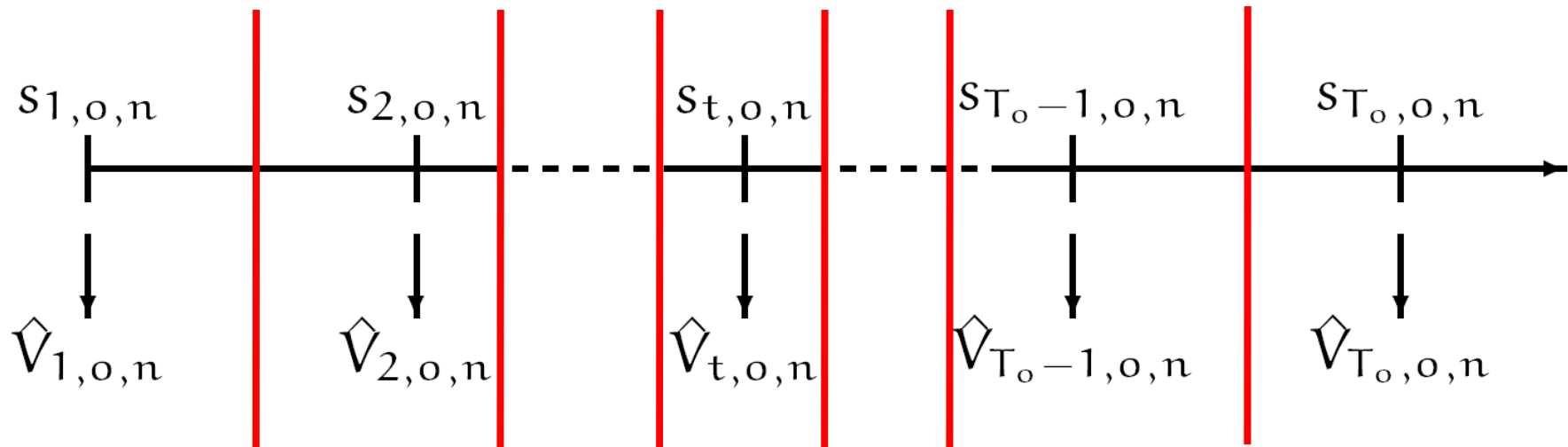
$$V_{t,o,n} = \{v_{1,t,o,n}, v_{2,t,o,n}, \dots, v_{E,t,o,n}\}$$



Model: state transition process

- $\hat{V}_{t,o,n}$: specific vector of “static” utility functions capturing the respondent perception of the frame t

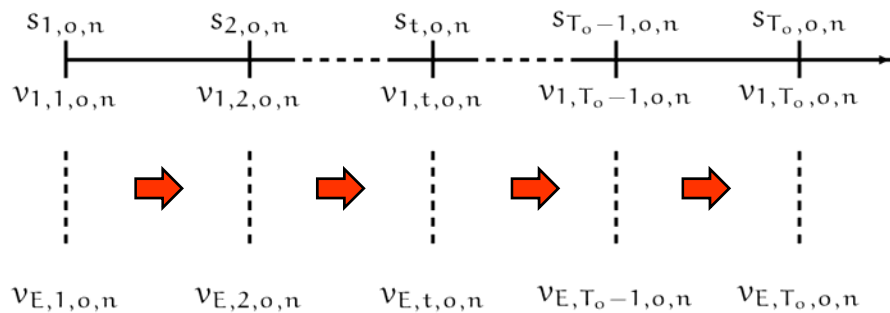
$$\hat{V}_{t,o,n} = \{\hat{v}_{1,t,o,n}, \hat{v}_{2,t,o,n}, \dots, \hat{v}_{E,t,o,n}\}$$



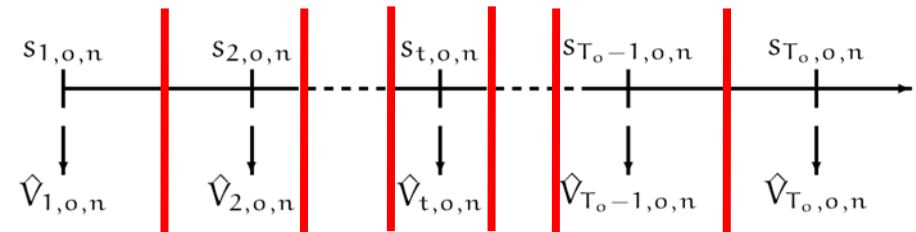
Model: state transition process

- Link between $V_{t,o,n}$ and $\hat{V}_{t,o,n}$

dynamic



static



$$\longrightarrow V_{t,o,n} = \sum_{a=1}^t A^{t-a} \hat{V}_{a,o,n} + \xi_n$$

Model: state transition process

$$V_{t,o,n} = \sum_{a=1}^t A^{t-a} \hat{V}_{a,o,n} + \xi_n$$

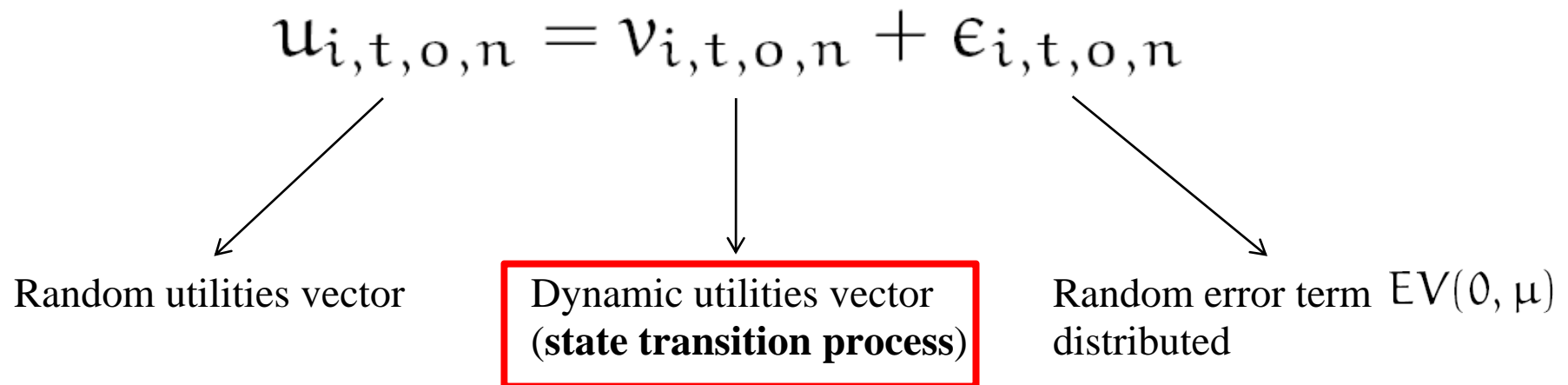
Diagram illustrating the components of the state transition process equation:

- $V_{t,o,n}$: Dynamic utilities vector
- A : Transition matrix with attenuation memory effect
- $\hat{V}_{a,o,n}$: Static utilities of the frame a
- ξ_n : Respondent n unobserved influences

- Remarks: A : in $\mathbb{R}^{E \times E}$ can be set diagonal and universal to ease the model identification
- ξ_n : depends only on the respondent, we supposed it $N(0, \sigma)$ distributed

Model: measurement equation

- Association of a random utility $u_{i,t,o,n}$ for each frame t of the video sequence o watched by the respondent n and for each expression i



 **Discrete choice models framework**

Model: measurement equation

- $P_{o,n}(i/t, \xi_n)$: probability for the respondent n of choosing the expression i in the frame t of the video sequence o , given ξ_n
- $\epsilon_{i,t,o,n} \sim EV(0, \mu)$: mixture logit for panel data

$$P_{o,n}(i/t, \xi_n) = \frac{\exp(v_{i,t,o,n}(\xi_n))}{\sum_{j=1}^E \exp(v_{j,t,o,n}(\xi_n))}$$

How link $P_{o,n}(i)$ with $P_{o,n}(i/t, \xi_n)$?

Model: measurement equation

- $P_{o,n}(i)$: probability for the respondent n of choosing the expression i to label the video sequence o
- $P_{o,n}(t)$: probability for the respondent n of making his final expression choice for the video sequence o , when watching at the frame t
- $f(\xi_n)$: multivariate density function of ξ_n

$$P_{o,n}(i) = \int \sum_{t=1}^{T_o} P_{o,n}(i/t, \xi_n) P_{o,n}(t) f(\xi_n) d\xi_n$$

Model: measurement equation

$$P_{o,n}(i) = \int \sum_{t=1}^{T_o} \underbrace{P_{o,n}(i/t, \xi_n)}_{\substack{\text{Probability of choosing} \\ \text{the expression } i, \text{ for the} \\ \text{individual } n, \text{ watching the} \\ \text{video sequence } o, \text{ in the} \\ \text{frame } t}} \underbrace{P_{o,n}(t)}_{\substack{\text{Probability for the} \\ \text{individual } n, \text{ when} \\ \text{watching at the video } o \text{ to} \\ \text{make his choice when} \\ \text{faced to the frame } t}} f(\xi_n) d\xi_n$$

multivariate density
function of ξ_n

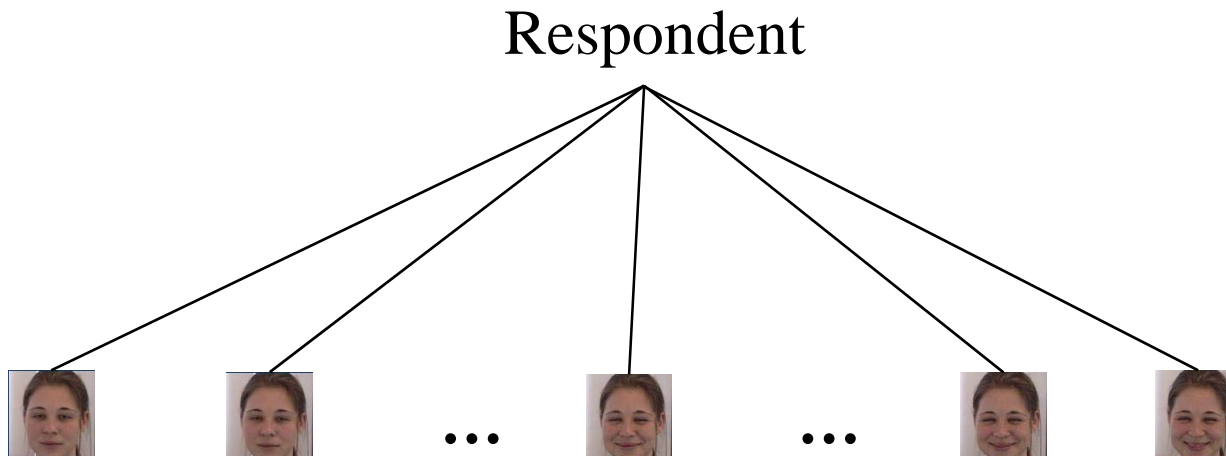
Probability of choosing the expression i , for the individual n , watching the video sequence o

Probability for the individual n , when watching at the video o to make his choice when faced to the frame t

Model: measurement equation

- $P_{o,n}(t)$: probability for the respondent n of making his final expression choice for the video sequence o , when watching at the frame t

 **Discrete choice model**



Model: measurement equation

- $\bar{v}_{t,o,n}$: utility measuring the dynamic of the frame t of the video sequence o , watched by the respondent n

$$P_{o,n}(t) = \frac{\exp(\bar{v}_{t,o,n})}{\sum_{a=1}^{T_o} \exp(\bar{v}_{a,o,n})}$$

 Derivatives of features in $\bar{v}_{t,o,n}$

Model: likelihood function

- Estimation made by likelihood maximization
- $C_{i,o,n}$: indicator of choice equals to one if respondent n chose to label the video sequence o with the expression i

$$l = \prod_{n=1}^N \prod_{o=1}^{O_n} P_{o,n}(i)$$

$$l = \prod_{n=1}^N \prod_{o=1}^{O_n} \left(\prod_{i=1}^E \int \sum_{t=1}^{T_o} P_{o,n}(i/t, \xi_n) P_{o,n}(t) f(\xi_n) d\xi_n \right)^{C_{i,o,n}}$$

Model: likelihood function

$$l = \prod_{n=1}^N \prod_{o=1}^{O_n} P_{o,n}(i)$$

$$l = \prod_{n=1}^N \prod_{o=1}^{O_n} \left(\prod_{i=1}^E \int \sum_{t=1}^{T_o} P_{o,n}(i/t, \xi_n) P_{o,n}(t) f(\xi_n) d\xi_n \right)^{c_{i,o,n}}$$

$$l = \prod_{n=1}^N \prod_{o=1}^{O_n} \left(\prod_{i=1}^E \int \sum_{t=1}^{T_o} \frac{\exp(v_{i,t,o,n}(\xi_n))}{\sum_{j=1}^E \exp(v_{j,t,o,n}(\xi_n))} \frac{\exp(\bar{v}_{t,o,n})}{\sum_{a=1}^{T_o} \exp(\bar{v}_{a,o,n})} f(\xi_n) d\xi_n \right)^{c_{i,o,n}}$$

Model: specifications

- Discrete Choice Model framework
- Attributes

- $\hat{V}_{t,o,n}$: **FACS, EDU, Texture, Socio-economics**

→ M. Sorci et al, “Static facial expression recognition”

- $\bar{v}_{t,o,n}$: **Derivatives** of features

→ measure the frame dynamic

Conclusions and Perspectives

- Conclusion:

- database of face video annotations
- new model framework
- estimation by likelihood maximization

- Perspectives:

- implementation of the likelihood maximization
- model estimation: find a satisfactory specification
- model validation: measure the prediction power

Conclusions and Perspectives

- Conclusion:

- database of face video annotations
- new model framework
- estimation by likelihood maximization

- Perspectives:


- implementation of the likelihood maximization
- model estimation: find a satisfactory specification
- model validation: measure the prediction power

Thank you for your attention

Data: data file

- Face video annotations data base  Data file for model estimation

Observation		Expression	Individual				Video			
obs.	label	id	gender	educ	...	frame 1 X1 X2 ...	frame 2 X1 X2		
1	7	1	0	4	...	2.05 4.36	3.43 4.10	...		
2	3	1	0	4	...	1.20 3.52	1.15 3.12	...		



Model: state transition process

- $s_{t,o,n}$: state associated with the frame t of the video sequence o watched by the respondent n
- $\hat{V}_{t,o,n}$: vector of utilities characterizing the frame t of the video sequence o for the individual n (dimension E)
- $V_{t,o,n}$: vector of utilities associated with the state $s_{t,o,n}$ (dimension E)
- ξ_n : vector of error terms specific to the individual n , interfering in the transition process (dimension E)
- σ : vector of standard errors of ξ_n (dimension E)
- A : squared appreciation matrix of dimension $E \times E$ associated to the respondent n faced to the video sequence o

Model: measurement equation

- Link the observation choice $y_{o,n}$ with the states sequence $\{s_{t,o,n}\}_{t \leq T_o}$
 - $U_{t,o,n}$: vector of random utilities associated with $s_{t,o,n}$ (dimension E)
 - $\epsilon_{t,o,n}$: vector of unobserved attributes interfering in $U_{t,o,n}$ associated to $s_{t,o,n}$ (dimension E)
 - $\bar{v}_{t,o,n}$: utility associated with the frame t of the video sequence o for the individual n , measuring the frame dynamic
 - $y_{o,n}$: choice made by the respondent n when faced to the video sequence o