# Object Representation Based On Gabor Wave Vector Binning : An Application to Human Head Pose Detection

M. Dahmane and J. Meunier

University of Montreal

# Introduction

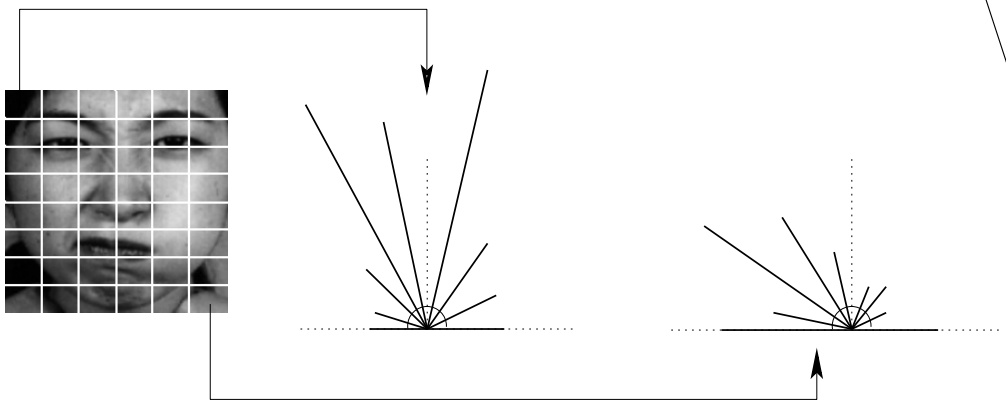
- Head pose is important:
  - Inferring important non verbal information
  - Focus of attention
  - Agreement disagreement
  - People nod, confusion etc.

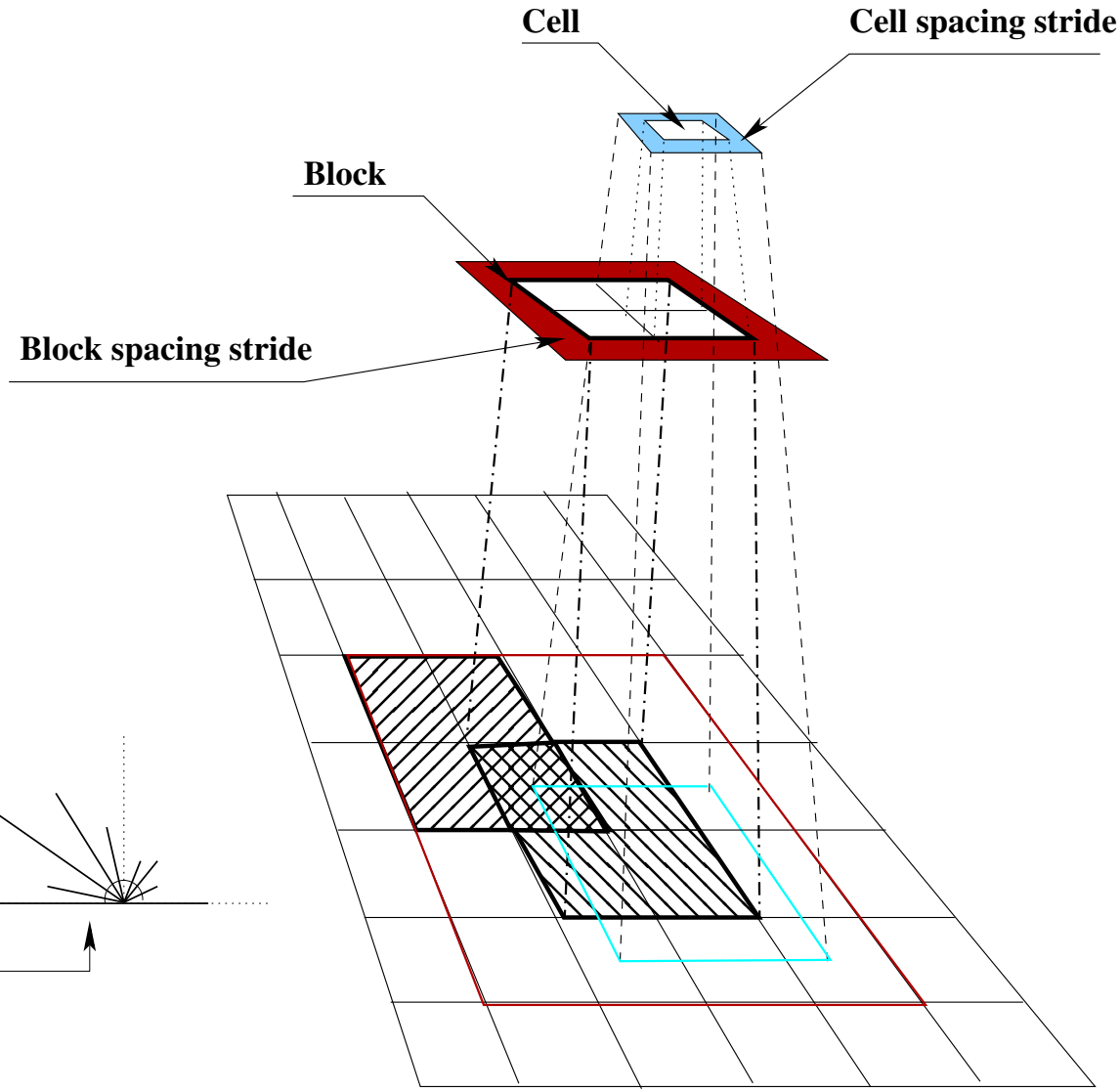
# Computerized head pose extraction

- Difficulties
  - Identity and Facial dynamics
- Approaches
  - **Geometric** : exploit properties that influence the human pose.
    - Sensitive to the location of facial points
  - **Appearance-based :**
    - Naturally avoid the problem of precise and stable localisation
    - Need a suitable descriptors

# A. Histogram of oriented gradient

- Divide the detection window into small cells
- Integrate over each cell, the magnitude of the edge gradient for each orientation bin
- Normalize the local histogram over the four-cells block





# B. The Gabor wave vector binning

- Gabor wave vector binning based descriptors consist on features generated by :
  - wavelet transform corresponding to a set of selected wave vectors (ie. orientations and scales)

# Gabor wave vectors

- The Gabor wavelet transform family is defined as:

$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} e^{-\frac{\|k_{\mu,\nu}\|^2 \|z\|^2}{2\sigma^2}} \left[ e^{\imath\, k_{\mu,\nu}\, z} - e^{-\frac{\sigma^2}{2}} \right]$$

where

$$k_{\nu,\mu} = k_\nu\, e^{\imath \phi_\mu}$$

Commonly, we have μ= {0..7} and ν {0..4} defining 40 wave vectors.

# Gabor wave vector binning

- The Basic key idea :
  - shapes
    - can be learned from local window
    - using the spatial distribution of magnitude over different frequencies and orientations.

1. A first–order image gradients is used as salient image locations
2. GWT is processed  on salient pixels
3. An image window is used to evaluate local histograms of GWT magnitude responses

# The underlying motivation for using Gabor-based descriptors

- Consistency with intrinsic characteristics of face images.
- Gabor pyramid filtering maintains:
  - continuity in the spatial frequency of the Gabor feature
  - detection ability.

# POINTING’04 dataset.

The head pose database consists of 93 images of 15 persons.

# Technical implementation

- We used a **detection window** with 100×40 pixels size.
- We have to deal with the **alignment problem** by searching for the eyes region over the entire image.
- The detection window is **partitioned** into 8 by 4 **cells** of 12 × 10 pixels.

# Technical implementation (2)

- The **voting strategy** is based on the Gabor magnitudes
- Magnitudes are collected into **40 histogram bins** (wave vector → bin).
- Histograms are then **integrated** over the cell.

# Technical implementation (3)

- For each block of 2×2 neighboring cells the histograms are concatenated into a **block-histogram**.
- The resulted block-histograms were concatenated into **a single** (1280 dim) **feature vector**

# SVM as base learners of poses.

- For the multiclass SVM, we used RBF-kernel :

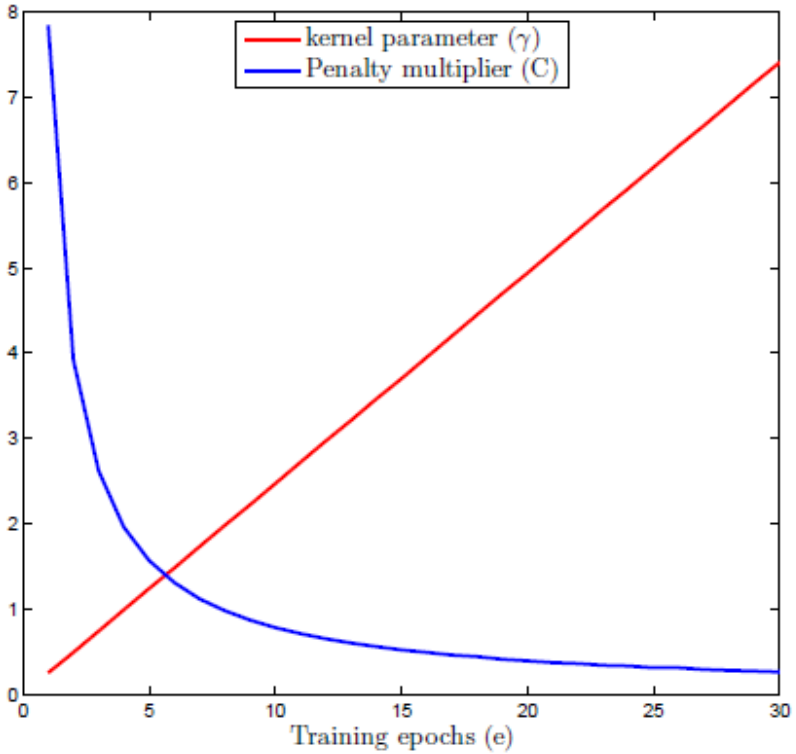
$$K(\mathrm{x}, \mathrm{x}_i^*) = \exp\left(-\gamma \|\mathrm{x} - \mathrm{x}_i^*\|^2\right)$$

- SVM parameters selection ?
  - We used a empirical **epoch-based strategy** to determine :
    - Parameters $\gamma$ and C.

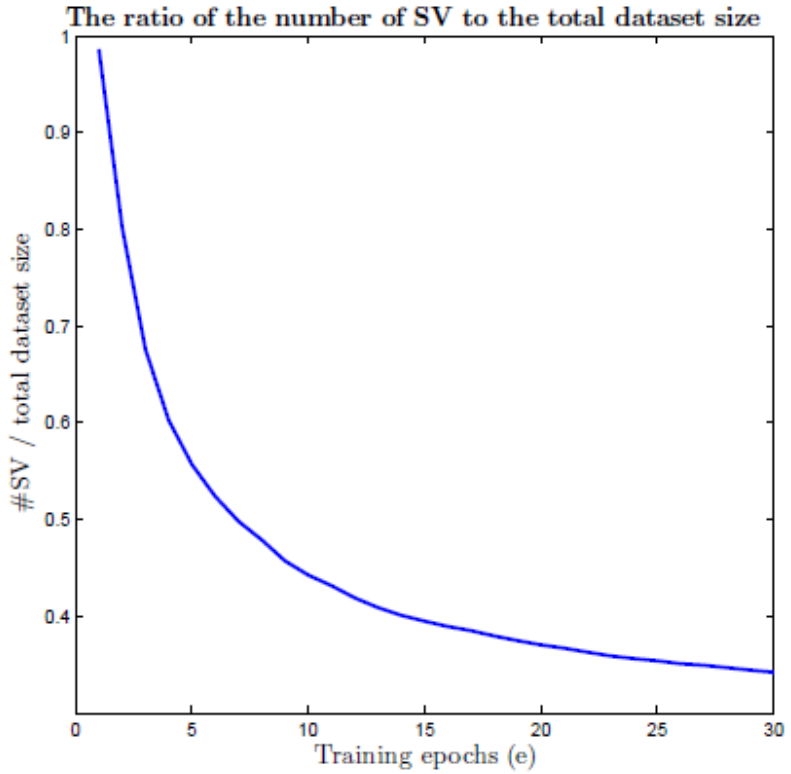
# SVM kernel parameters selection

- We select the optimal tuple $(\gamma, C)$ corresponding to the epoch with
  - highest training accuracy
  - and a reasonable number of SVs.

# SVM kernel parameters selection (2)




The SVM parameters evolution over training epochs

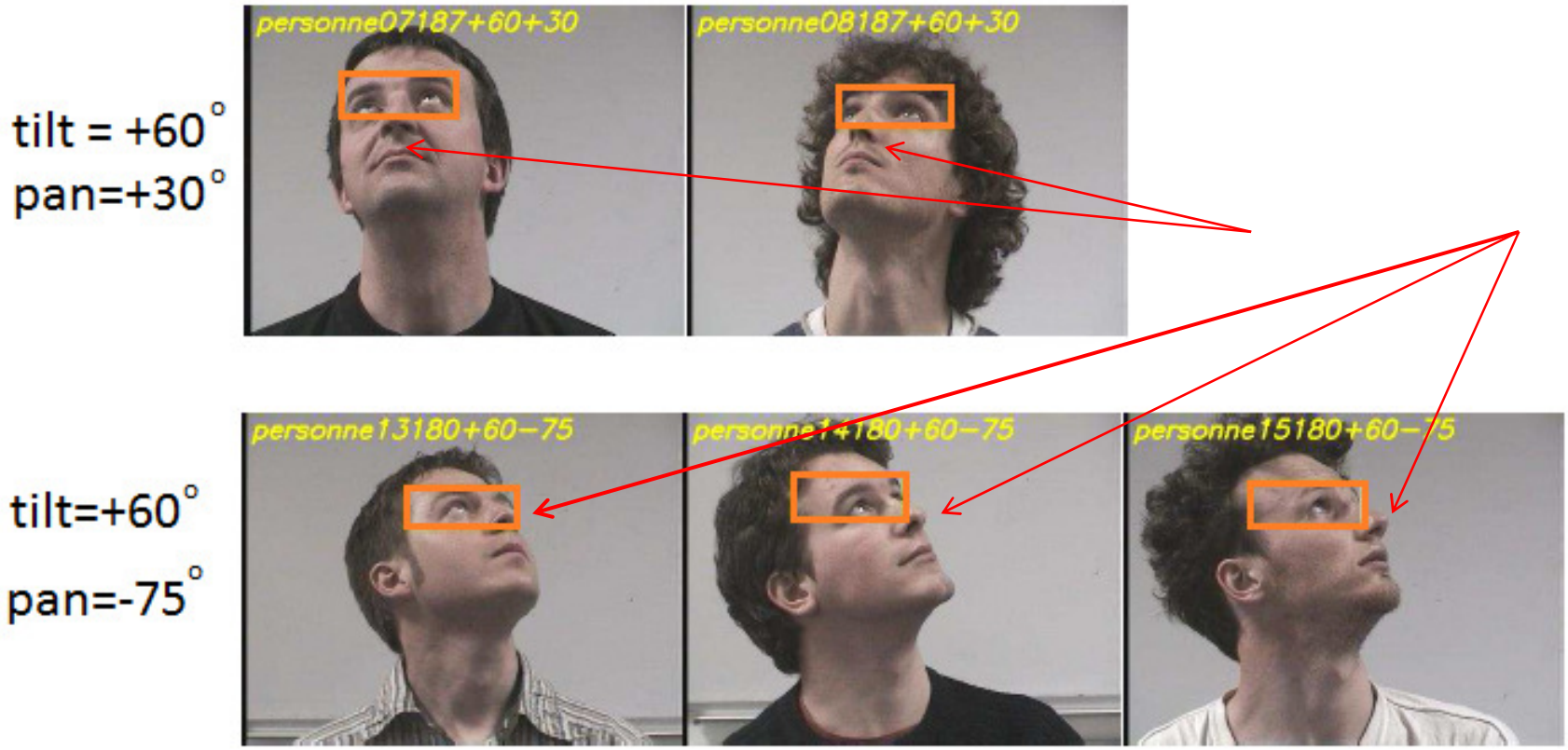



Stabilization of the number of support vectors from epoch 25

- Performances of different pose detection techniques on POINTING'04 setup.

| | Mean absolute error (°) | | Classification accuracy (%) | |
|---|---|---|---|---|
| | **Yaw** | **Pitch** | **yaw** | **pitch** |
| Human | 11.8 | 9.4 | 40.7 | 59.0 |
| Voit et al. | 12.3 | 12.7 | - | - |
| Tu et al. | 14.1 | 14.9 | 55.2 | 57.9 |
| Gourier et al. | 10.1 | 15.9 | 50.0 | 43.9 |
| **Our method** | **5.7** | **5.3** | **65.0** | **73.3** |

# Some pose ambiguity problems



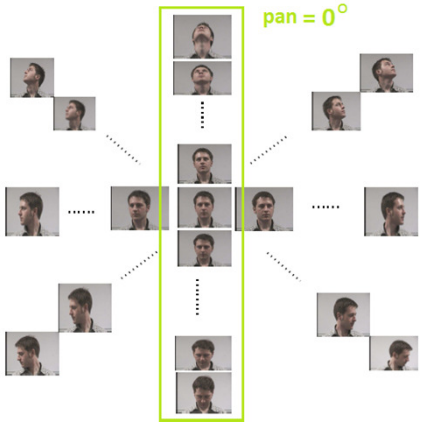
# Proposed descriptors vs. HoG performance comparison

| | | Mean absolute error (°) | | Classification accuracy (%) | |
|---|---|---|---|---|---|
| | | **Yaw** | **Pitch** | **yaw** | **pitch** |
| ±0° | HoG | 5.6 | 6.3 | 66.4 | 70.7 |
| | **Our feature set** | **5.7** | **5.3** | **65.0** | **73.3** |
| ±15° | HoG | 0.9 | 3.7 | 97.5 | 88.1 |
| | **Our feature set** | **0.9** | **2.5** | **97.5** | **91.8** |

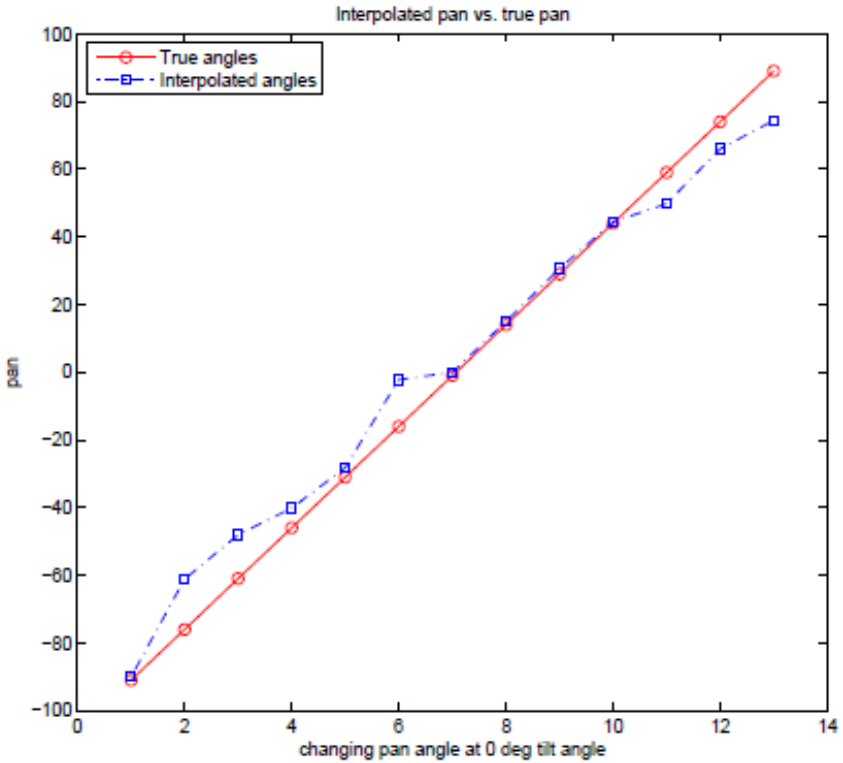
# Continuous poses inferring from POINTING'04 discrete poses

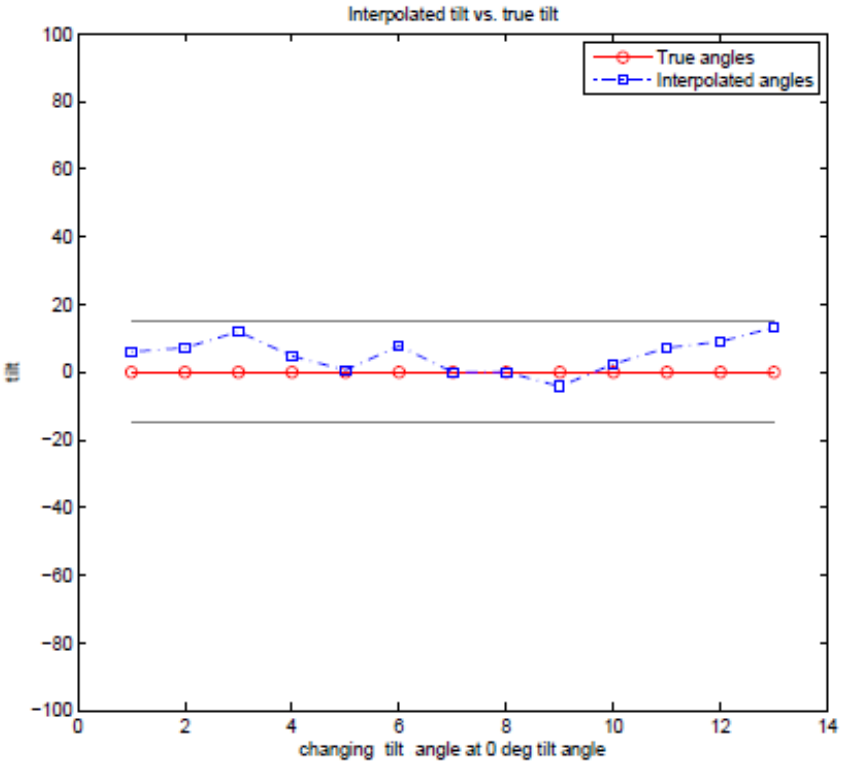
- Gabor response continuity
  - establish a **mapping** between the space of the discreet poses and the descriptors space.
- Continuous pose consists on:
  - Interpolating the 3×3 neighboring poses (poses within ±15° range) of the **winner pose** using the respective SVM-scores as weights.

# Continuous poses




- Interpolated pan and tilt at pan= 0°



# Conclusion

- We presented a **Gabor wave vector binning** based descriptors.
- We show that they
  - present for pose estimation a **suitable feature set**.
  - perform **better classification accuracy** vs. existing algorithms and even Human performance

# Conclusion (2)

- Better classification accuracy against the HoG detector is obtained
- Able to infer a smooth continuous estimate of the pan and tilt angles
- We need to optimize the processing time to generate  the 40 integral images.

Thanks