Purpose
We have constructed a compact artificial visual system based conceptually on the fundamental design properties of human visual system. In the prototype, the foveal subsystem is an active narrow field of view camera mounted on a pan/tilt unit. The peripheral subsystem is a static wide field of view camera that provides nearly a 180 degree field of view. Events of interest detected in the panoramic sensor are selected for foveation by a high-resolution pan/tilt sensor.
Possible applications include surveillance, telepresence and remote education.

Introduction
We want to find and adapt methods that are specifically good for real-time detection of faces in panoramic images which are generally of low resolution, distorted and with great variability in scale and lighting conditions. The size of faces in the images captured by the sensor typically do not result in facial features that lead to good performance. In such conditions color may be the strongest static cue available.

Outline of the Approach
The first step in face detection is abstracted by scanning window technique that tries to match each window to template, which is a colour average face in our case.

Background Subtraction
Background subtraction is one of the most popular techniques for finding the objects which do not constantly belong to the foreground (background) or not. Most of the activity is associated with people, thus finding foreground regions will give us a very strong cue where a person is.

Background subtraction is one of the most popular techniques for finding the objects which do not constantly belong to the foreground or non-foreground. Most of the objects which do not constantly belong in the background such as change of lighting or position of objects, exponential forgetting with learning rate $\alpha_k$ was introduced. Up until now the value for $\alpha_k = 0.004$ has been chosen based on training data. Models with adaptive learning rate are at the exploration stage.

On-line learning
Approximation of EM algorithm, namely on-line k-means clustering with $k = 0$, is used.

Definition of foreground
The scene is represented by the array of pixels which are assumed to be independent from each other.
Thus, each pixel is classified as a background or foreground separately.

Examination of experimental image sequences suggests that intensity distributions of the majority of the pixels can be explained by one wide and one narrow Gaussian, which could correspond to foreground and background processes.

Usage of priors
The scale of the face ultimately depends on how far from the camera the person is. For most environments there would be a tendency for far objects to be smaller and appear higher in the images than close ones (e.g. a classroom with desks). Thus, the probability of each scale being at a certain height in the image can be calculated for the specific classroom.

Classification
For each window in the image the posterior probability of it containing a face is determined using the Bayesian inference rule and maximum a posterior hypothesis:

$$p(\text{face}|\text{data}) = \frac{p(\text{data}|\text{face}) p(\text{face})}{p(\text{data})} = \frac{p(\text{data}|\text{face}) p(\text{face})}{p(\text{data}|\text{face}) p(\text{face}) + p(\text{data}|\text{non-face}) p(\text{non-face})}$$

where $\eta(\text{data}; \mu, \Sigma)$ stands for normal distribution.

Conclusions
1. A face detection framework for panoramic images with 85% detection rate and huge potential to be developed further.
2. Few thresholds and control parameters that result in straight-forward testing and tuning of the system for desired detection and false positive rates.
3. An adaptive background subtraction module which significantly improves person detection and can be integrated in tracking procedure.
4. The future use of the foveal camera for confirmation of correct face candidates, to refine the face model and explore further processing possibilities such as face recognition.