Design Concept of a Human-like Robot Head

Karsten Berns
Robotics Research Lab
Department of Computer Science
University of Kaiserslautern Germany
Email: berns@informatik.uni-kl.de

Tim Braun
Robotics Research Lab
Department of Computer Science
University of Kaiserslautern Germany
Email: braun@informatik.uni-kl.de

***Abstract*— For humanoid robots able to assist humans in their daily life, the capability for adequate interaction with human operators is a key feature. If one considers that more than 60% of human communication is conducted non-verbally (by using facial expressions and gestures), an important research topic is how interfaces for this non-verbal communication can be developed. To acieve this goal, several *robotic heads* have been designed. However, it remains unclear how exactly such a head should look like and what skills it should have to be able to interact properly with humans. This paper describes an approach that aims at answering some of these design choices. Based on parameters obtained from a simulation system that was used to test facial expressions, the design of a human-like head developed at the University of Kaiserslautern is described. Additionally, the mechatronical design of the head and the accompanying neck joint are given. Finally, a real-time capable method for image based face detection is presented, which is a basic ability needed for interaction with humans.**

I. Introduction

Worldwide, several research projects focus on the development of humanoid robots. Especially for the head design there is an ongoing discussion if it should look like a human head or if a more technical optimized head construction [1], [3] should be developed. The advantage of a technical head is, that there is no restriction according to the design parameters like head size or shape. This fact reduces the effort for mechanical construction. On the other hand, if realistic facial expressions should be used to support communication between a robot and a person, human likeness *could* increase the performance of the system. In fig. 1, humanoid robot heads are classified according to their human likeness and technical complexity[1]. The aim of our project is to develop both a very complex robot head able to simulate the facial expressions of humans while perceiving its environment by a sensor system (stereo-camera system, artificial nose, several microphones, ..) similar to the senses of a human. On the other hand, the robot head should look like a human head to examine, if its performance due to non-verbally communication will be higher compared to technical heads.

In the following, the design criteria of our head are introduced. Starting from Ekmans *Facial Action Coding System* [4], adequate action units are selected for the emotions that shall be expressible by the head. Then, our simulation system is described which was implemented to test the behavior of the skin when activating the action units. The results obtained from this are used to guide the construction of the robot head. In parallel to these experiments, software was developed to detect the head of a human being based on colour information and to extract features. This work, which is a pre-condition for the interaction of the robot head with a human, is described at the end of the paper.

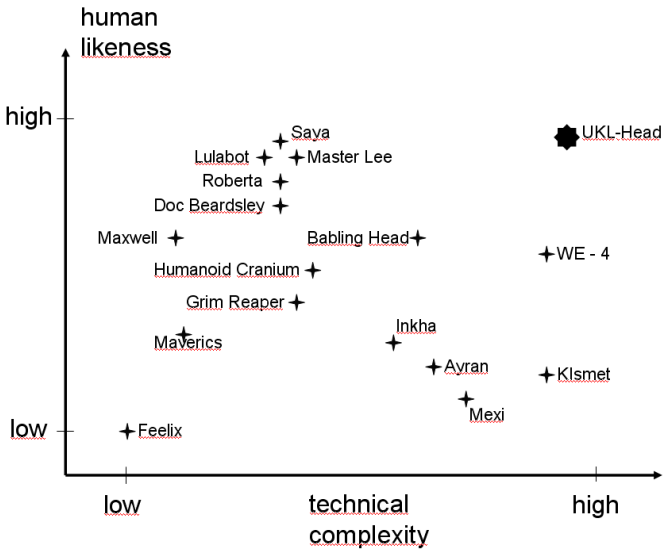

Fig. 1. Classification of humanoid robot head according to human likeness and technical complexity. The technical complexity is meassured by the number of sensors/actuators and the interaction skills with humans.

II. Expression of emotion

In [4] a set of action units is introduced which are mainly used for the expression of emotions by means of facial expression. Each of these units consists of one or several muscles which are connected to specific parts of the skin. Ekman has shown that an activation of the muscles of a specific set of action units lead to an expression of the basic emotions like joy, fear, sadness, surprise or disgust. For example, the emotion 'disgust' is expressed if the action units 'brow lowerer', 'lip corner depressor' and 'chin raiser' are activated at the same time. If each of the action units is implemented by one motor and only the minimum number of units is selected for the expression of emotions, at least 30 motors must be installed in a robot head to control the skin in a humanlike way. One

[1]Also see http://www.androidworld.com/prod04.htm

reason for the big number of action units is that muscles are only able to contract and therefore only able to move the skin in one direction. On the other side, to move a bigger area of the skin it is necessary to fix several muscles on different contact points. To transfer the design principles of a human head to robot head able to express emotion like humans, it is necessary to answer the following questions:

- To properly simulate Ekmans action units, which areas of an artificial skin must be moved and in which direction?
- How important is the velocity when moving from one emotional state to another one?
- Which trajectories in the emotional space look natural when switching between emotion?
- Is it possible to express an emotion like fear more or less strongly through varying degrees of activation of the corresponding action units?

III. Simulation

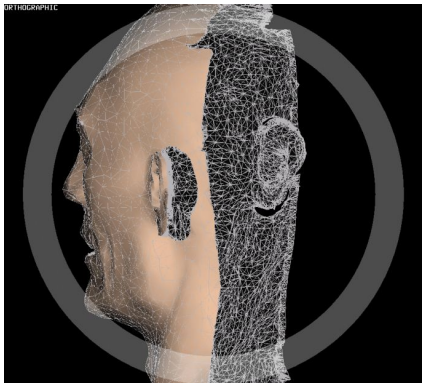

Fig. 2. Simulation of the head of the University of Kaiserslautern with the artificial skin

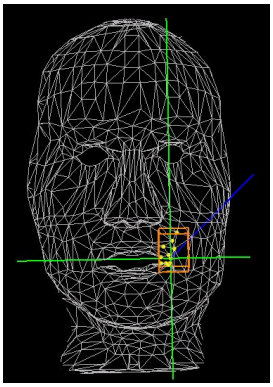

Fig. 3. Mesh points which are selected to be affected by drawn force vector. Based on the results of the simulation it was possible to select an adequate set of mesh points and the direction, in which these points should be moved when the corresponding action unit was activated.

To answer the questions raised above, a simulation of the artificial skin was implemented together with the Computer Graphics Group (Prof. Hagen) of the University of Kaiserslautern. Starting with a 3D laser scan (using the Minolta-Digitizer Vi900) of the artificial head covered with a silicon skin, a triangular mesh with about 1,6 Mio triangles was generated and reduced by standard mesh simplification algorithms to about 100000 triangles. The simplification was done in such a way that smooth areas are represented by only a few meshes while areas with a lot of curved or wrinkled skin retain a high number of triangles. Then, the resulting mesh was modelled as a spring/damper system, which describes the behavior of the silicon skin when external forces are applied. Based on this simulation of the artificial skin, several tests were performed to find out where and how the artifical skin can be moved in order to simulate an action unit. Also the influence of the magnitude of the force was evaluated. After the realisation of Ekmans action units in the simulation system, tests were performed to see if the basic emotions can be expressed adequately. For this, 60 students have been interview to classify the simulated faces according to the basic emotions. As result, about 75% of the expressed emotions are correctly classified. In fig 4 the simulated head is shown with the silicon skin attached.

IV. Mechatronics

Fig. 4. Mechanical design of the humanoid robot head with the silicon skin. The silicon mask is designed by the company Clostermann Design Ettlingen,Germany

The mechatronical system of the head is still under development. Due to simulation results suggesting that a single point of attachment would look unrealistic, the selected mesh points, which belong to an action unit, are connected with small metal plates. These plates are glued on the silicon skin and transmit the applied force over a small area, leading to a much improved 'look' of the expressed emotions. Wires are connected to these plates which allow their movements in direction of the related action unit. As actuators, 10 servos are used to pull and push the wires. Additionally, a servomotor is used to raise and lower the lower jaw. This setup was found to be adequate to express the basic emotions clearly enough for humans to understand during the simulation stage.

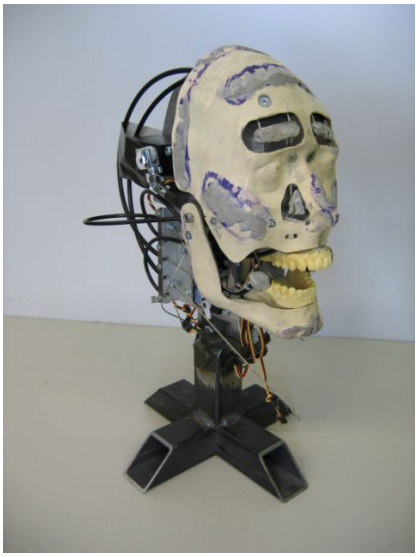

Fig. 5. Mechanical construction of the humanoid robot head

As next steps, the integration of microphones and loudspeaker in the head is in preparation. Also the mechanical eye construction with 3 DOF each and a neck construction with 3 DOF is under development (see fig. 6). The control of the servo motors as well as the determination of the pose from the inertial system is done with a DSP (Motorola 56F803) connected to a CPLD (Altera EPM 70 128). The calculation of movements of the action units according to the emotions which should be expressed is done on a Linux-PC. The DSPs are connected via CanBus to the PC. In fig. 4 the head and the carrier is shown.

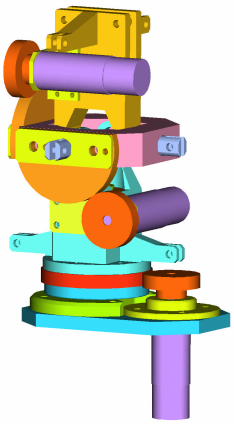

Fig. 6. CAD drawing of the neck constuction. The neck has 3 DOF; one for the rotation of the head,one for up and down, and one for moving left and right. As actuators DC motors are used. For the last 2 DOF springs are included in the construction to increase the torque in the corresponding joints.

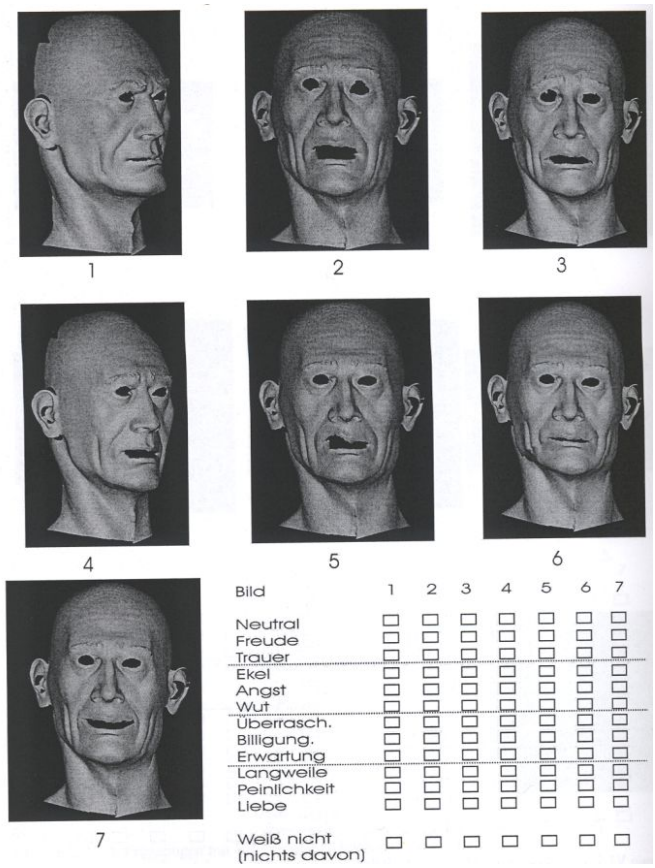

Fig. 7. This sequence of images was used as a questionre to evaluate, if the expressed emotion are recognised by the test persons.

V. Face detection for interaction with a human operator

One of the next experiments planned for the humanoid robot head is the expression of emotion based on the facial expressions and gestures of a human operator. For example, the robot should be able to 'mirror' the emotions expressed by the human. A necessary precondition for this is the detection of humans and especially their faces. For this, a detection algorithm based on the camera images that will be obtained by the artificial eyes has been implemented and will now be described.

To allow a reliable detection of faces in the expected dynamic environment, a pattern recognition method is needed that can cope well with variable lighting and differing facial poses or expressions. For this task, a pattern classifier that was previously used in the context of mobile robotics [2] has been used. This method is based on a pattern classificator called Support Vector Machine (SVM). The SVM has been shown to yield excellent face detection performance even in difficult conditions and thus seems to be a good choice in this context. However, the computational demands of a SVM-based face detection system are high compared to haar-based classifiers or neural-network approaches.

In order to reduce the computational complexity of a SVM-

based face classifier and be able to exploit its high classification ability, we have developed an approach that speeds up the face detection task using three complementing techniques:

1) A special, highly efficient *Sequential Reduced SVM (SRSVM)* is used for classification of image patches instead of a regular one.
2) Prior to the application of the SRSVM, it's search space is reduced to image parts containing *face candidates*. These candidates are determined quickly using skin-color filtering and geometrical constraints.
3) The search space is further reduced by fusing range information taken from a stereo camera system and an auxiliary infrared distance sensor (attached to the head base) with the captured image. This yields distance information for each face candidate and is used to *restrict the scale* at which to look for faces.

The **SRSVM** has been originally introduced by Romdhani and Schölkopf [6] as a way to speed up regular SVMs. Key idea behind their approach is to reduce the number of support vectors (SVs) that need to be taken into account for classification. This reduction is achieved by replacing the set of support vectors of a normal SVM with a reduced set that contains a lot less SVs, but still defines approximately the same decision surface. In a second step, the SVM evaluation scheme is modified so that the reduced set can be evaluated sequentially. Consequently, each image patch is first classified using just one SV; additional SVs are only considered if this classification does not yield a sufficiently clear result.

Fig. 8. Original image for tracking faces.

In order to confine the application of the SRSVM to image areas that are likely to contain faces, **face candidate** areas are extracted from the input image prior to classification. These candidate areas are determined by applying a skin-color filter to the color input image.

This image filter is based on results obtained by Rehg [5]. In his work, the conditional probability density function determining the likelihood that a given RGB-Color shows skin ($P(skin|RGB)$) was obtained by hand-labelling skin

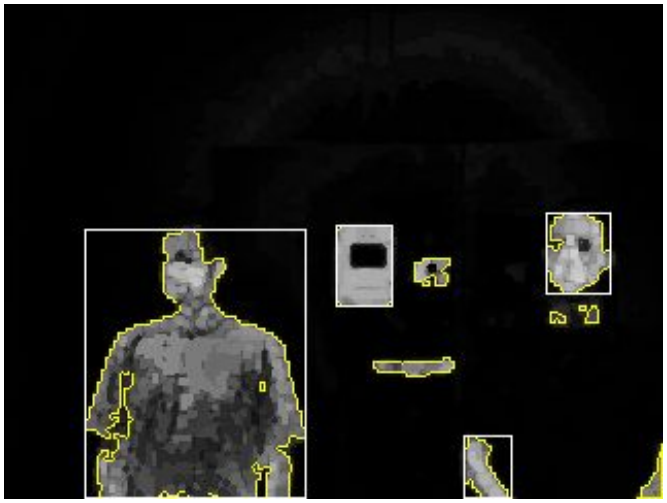

Fig. 9. Skin-filtered image. Detected outlines are marked, valid face candidates are indicated by boxes.

colored regions in images taken from the internet. This density function has been published in the form of a mixture of 16 three-dimensional gaussian functions; for the use in this realtime application, the function must be expanded into a precalculated lookup-table. Details on this process are given below.

Be x an RGB colour vector and the i^{th} gaussian function $1 < i < 16$ determined by a scalar weight ω_i, a mean vector μ_i and a diagonal covariance matrix $\sum_i$. The distributions $P(RGB|skin)$ and $P(RGB|\neg skin)$ can be evaluated from the published mean and variance values using the following equations:

$$P(x) = \sum_{i=1}^{n} \omega_i \frac{1}{(2\pi)^{\frac{3}{2}} |\sum_i|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_i)^T \sum_i^{-1}(x-\mu_i)} \quad (1)$$

$$= \sum_{i=1}^{n} \omega_i \frac{1}{(2\pi)^{\frac{3}{2}} (\sigma_{R_i} \cdot \sigma_{G_i} \cdot \sigma_{B_i})^{\frac{1}{2}}} e^{t} \quad (2)$$

with

$$t = -\frac{1}{2}\Big(\frac{(x_R - \mu_{R_i})^2}{\sigma_{R_i}} + \frac{(x_G - \mu_{G_i})^2}{\sigma_{G_i}} + \frac{(x_B - \mu_{B_i})^2}{\sigma_{B_i}}\Big)$$

Given these two probability densities for an RGB colour vector, the probability $P(skin|RGB)$ needed for the image filter can be calculated using bayes' rule. The total percentage of skin pixels in the data set required for this amounts to about 10%.

$$P(skin|RGB) = \frac{p_a}{p_a + p_b} \quad (3)$$

with

$$p_a = P(RGB|skin) \cdot P(skin)$$
$$p_b = P(RGB|\neg skin) \cdot P(\neg skin) \quad (4)$$

and

$$P(skin) = 0.1, \quad P(\neg skin) = 1 - P(skin) = 0.9$$

Further processing by morphological smoothing, binarization and contour extraction yields outlines of potential faces.

These outlines are then tested using several geometrical validation rules:

1) The contour bounding box must have a minimum size of 19x19 pixels.
2) The width of the bounding box must lie between it's height and half of it's height.
3) The relative amount of filled pixels in the bounding box after binarization must be above 70 percent.

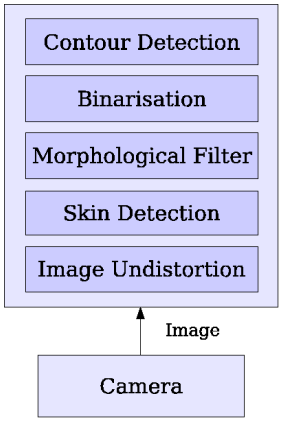

Fig. 10. Structure of Image Processing Part

Figure 8 and Figure 9 show an example for the face candidate generation before and after the application of the skin color filter. Figure 10 outlines the complete candidate generation stage.

Since the employed SVM can only classify image patches with a fixed size[2], an input image containing faces of unknown size must be examined at several scales. To avoid this time consuming multi-scale analysis, the distance between the robotic head and the depicted human will be calculated with stereo-image processing. At the moment a testbed (see fig. 11) for the stereo-camera system equipped with 2 dragonfly cameras is used to directly determine the appropriate scale for the face detection. This setup will be integrated later as part of the eyes of the humanoid head.

To actually detect faces in realtime using the collected information, the extracted face candidates are scaled according to their corresponding distance measures. Then, the area occupied by the candidate bounding box is split into 19x19 subwindows and classified by the SRSVM into faces and non-faces. The basic SVM needed for the face detection system has been trained on over 20000 face examples and 102000 non-face examples which have been collected from various image databases. The initial training using a gaussian kernel with a width of $\sigma = 0.04$ resulted in a SVM with 27267 support

[2]The image window used in our implementation is 19x19 pixels wide.

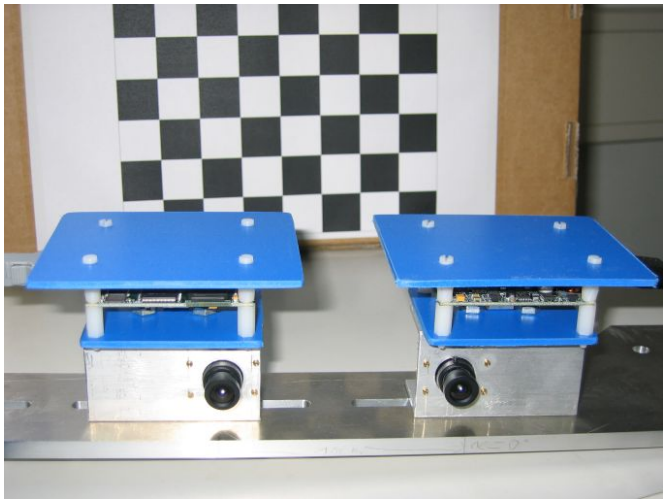

Fig. 11. A testbed for the stereo-camera system, which will be included as eyes in the humanoid robot head.

vectors. This set was subseqently reduced to 100 support vectors that formed the basis of the SRSVM.

We have found that an image patch can be classified by the SRSVM with only 4 SVs on average, taking $12\mu s$ per patch on a 1.6 GHz P4. Compared to a normal SVM, the resulting speed gain is immense. The search space reduction achieved by skin-color filtering and the incorporation of range data is also substantial. While the number of image patches that would have to be examined each frame without any reduction amounts to about 200000 (6 scales of a 320x240 pixel image), the implemented approach does only test about 700 patches per face candidate. Since in most frames, at most one or two candidates are present, this is again a big saving in computational time. Nevertheless, the detection performance is only slightly degraded.

In total, the face detection part is able to process 3 fps while leaving enough processing time for the remaining image processing, motion and safety systems.

VI. Conclusion

In this paper a humanoid head construction is introduced, which will be used to interact with humans. One focus of the present research is how the facial expressions of a human being can be transferred to a robot head. From our point of view the complexity of this problem will be reduced, if the robot head is human-like. Based on a simulation of the silicon skin of the head and the implementation of Ekmans action units facial expressions were simulated. In experiments with several students it was shown that the basic facial expressions (neutral, fear, disgust, happyness, anger, surprise, sadness) were in general classified correctly. The simulation was also used to reduced the number of actuators necessary to express emotions.

The second focus of this paper lies on a face detection approach that was implemented in order to detect possible interaction partners. Starting from images taken with a stereo

camera system, skin coloured face candiates are extracted and verified using a Support Vector Machine classifier. This SVM was modified in several aspects to allow for real-time face detection. It was shown that under illumination conditions normally found in indoor environments, human faces can be reliably detected in real-time. The next steps taken in the course of this work will include the addition of facial expression analysis to the image processing subsystem, and in parallel to this the completion of the mechatronical design as well as the implemention of a behaviour based control concept for interaction with humans.

REFERENCES

[1] Hideaki Takanobu Atsuo Takanishi, Hiroyasu Miwa. Development of human-like head robots for modeling human mind and emotional human-robot interaction. *IARP International workshop on Humanoid and human Friendly Robotics*, pages 104–109, Dec 2002.

[2] Tim Braun, Kristof Szentpetery, and Karsten Berns. Detecting and following humans with a mobile robot. In *Proceedings of the EOS Conference On Industrial Imaging and Machine Vision*, june 2005.

[3] Cynthia Breazeal. Emotion and sociable humanoid robots. *Int. J. Hum.-Comput. Stud.*, 59(1-2):119–155, 2003.

[4] P. Ekman and W. Friesen. *Facial Action Coding System*. Consulting psychologist Press, Inc, 1978.

[5] M. Jones and J. Rehg. Statistical Color Models with Application to Skin Detection. *IEEE Conference on Computer Vision and Pattern Recognition*, 1:274–280, 1999.

[6] S. Romdhani, P. Torr, B. Schölkopf, and A. Blake. Computationally Efficient Face Detection. *Proceedings of the 8th International Conference on Computer Vision*, 2:695–700, 2001.