Our facial expression database: the Oulu-CASIA NIR-VIS database, consists of six expressions (surprise, happiness, sadness, anger, fear and disgust) from 80 people between 23 to 58 years old. 73.8% of the subjects are males. The whole database includes two parts, one was taken in Feb. 2008 in Oulu by the Machine Vision Group of the University of Oulu, consisting 50 subjects and most of them are Finnish people. The other was taken in April 2009 in Beijing by the National Laboratory of Pattern Recognition, Chinese Academy of Sciences, consisting of 30 subjects and all of them are Chinese people. The subjects were asked to sit on a chair in the observation room in a way that he/she is in frontal direction to the camera. Camera-face distance is about 60 cm. Subjects were asked to make a facial expression according to an expression example shown in picture sequences. USB 2.0 PC Camera (SN9C201&202) includes an NIR camera and a VIS camera which capture the same facial expression as shown in Fig. 1. The imaging hardware works at rate of 25 frames per second and image resolution is 320×240 pixels. Now the whole database is available online.

Figure 1: Capture environment.

All expressions are captured in three different illumination conditions: normal, weak and dark. Normal illumination means that good normal lighting is used. Weak illumination means that only computer display is on and subject sits on the chair in front of the computer. Dark illumination means near darkness. Facial expression database structure is shown in Fig. 2. The number of video sequences is 480 (80 subjects by six expressions) for each illumination and imaging system pair, so totally there are 2880(480×6) video sequences in the database.
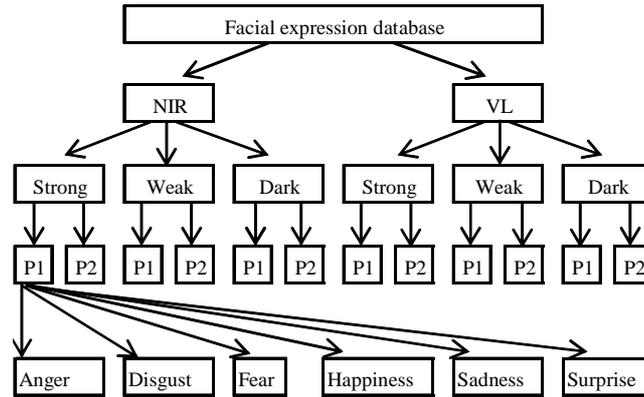
Figure 2. The structure of the facial expression database.

We have three subsets:

A: original videos, about 13GB after compression (about 28GB before compression).
B: JPEG images of original videos, about 1.2GB.
C: JPET images of cropped face images for all videos, about 150MB. The positions of the eyes in the first frame were detected automatically and these positions were used to determine the facial area for the whole sequence.