Facial Feature Tracking For Cursor Control

演讲人： 陈 勇

指导教师：董兰芳 老师

中国科学技术大学计算机系

2006年9月27日

论题的意义

- 动画前期预备工作
- access to the computer may be made more difficult by a person's disability
- on-screen position of a cursor may be controlled by the user changing the position of their head.
- …

内容

- detecting face objects using colour segmentation
- dominant object selection using statistical analysis
- locate nostrils' locations
- implementation specific details

Figure 1: Sample image captured using a Creative Labs Webcam Go.

Two approaches

- locate the head and then the features on it
  - generate a common head model
  - model matched, tracking can be accurate and efficient
- ignore the head and search for the features directly
  - model of the feature

  （eigentemplates have been suggested）
  - sensitive to slight variations in the image

  (i.e. variations in scale, shape and pose)

face detection methods

- depending on physiognomy of the face
- depending on the colour of face

Template matching

- define a template that resembled a facial feature and cross correlate it with a face image
- The location of the maximum response defines that feature's location
- deformable template matching method

- colour was due to two factors:
  - The amount of melanin in the skin
  - the ambient illumination (greater)
- skin colours were consistently with a fixed and quite narrow set of limiting values, after illumination were removed
- Illumination independence was achieved by deleting any one of the rgb components, the L and the Y component.

三种颜色模型

$$r = \frac{R}{R+G+B}$$
$$g = \frac{G}{R+G+B} \qquad (1)$$
$$b = \frac{B}{R+G+B}$$

$$L(x) = 105\log_{10}(x+1)$$
$$I = L(G)$$
$$R_g = L(R) - L(G) \qquad (2)$$
$$B_y = L(B) - \frac{L(G)+L(R)}{2}$$

$$Y = 0.30R + 0.59G + 0.11B$$
$$C_r = 0.50R - 0.42G - 0.08B \qquad (3)$$
$$C_b = -0.17R - 0.33G + 0.50B$$

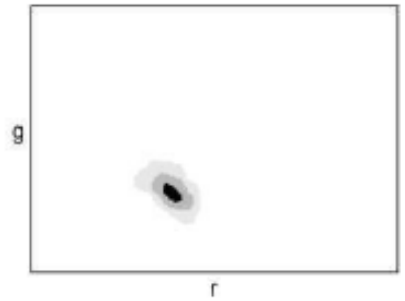


Figure 1. Distribution of skin colours of various races after intensity information is removed. The

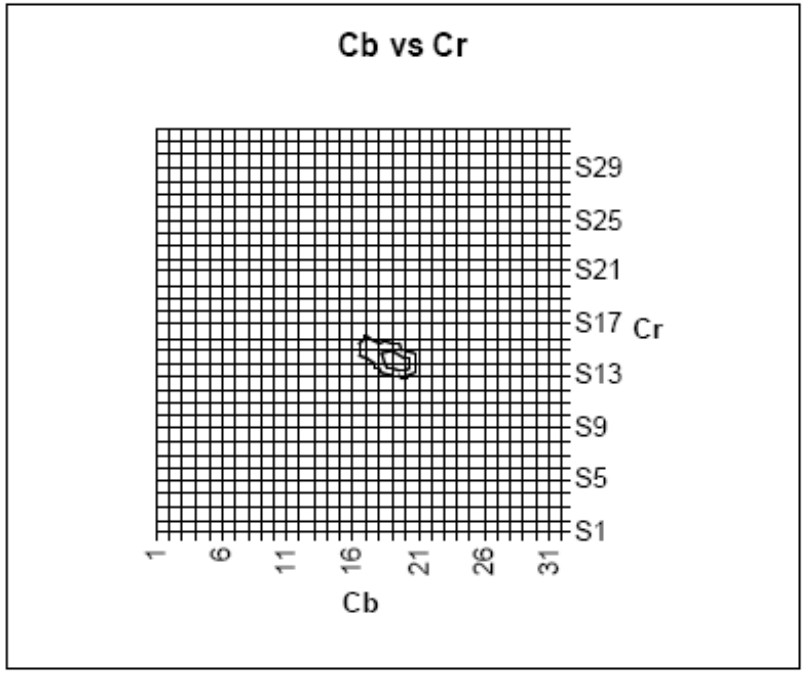


Figure 4: Clustering of skin colours in various colour spaces

Binary Image

(i) (ii)

Figure 5
a typical input
b skin map generated by colour matching.

Blob selection

- The undesired skin coloured pixels erroneously introduced by thresholding
- application of skin filtering results in clusters of skin coloured pixels, or blobs
- One blob will correspond to the face we are seeking, others will correspond to erroneous background.

Robust statistical estimation

- find the centre point and size of the most dominant blob
- constructing two one dimensional distributions by summing the pixels in the rows and columns of the binary image

- The means and standard deviations are then calculated for each histogram

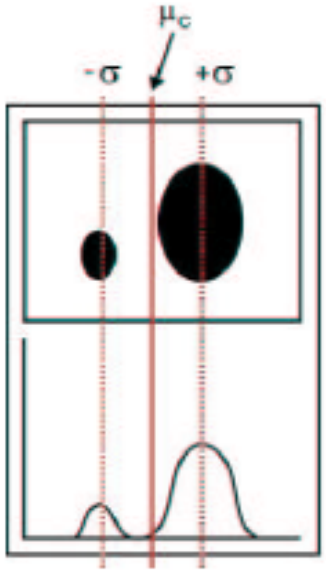
$$\mu_r = \frac{\sum_i i \cdot h_r(i)}{\sum_i h_r(i)}, \text{and } \mu_c = \frac{\sum_i i \cdot h_c(i)}{\sum_i h_c(i)}$$

$$\sigma_c = \sqrt{\frac{\sum_i (x_c - \mu_c)^2 h_c(i)}{\sum_i h_c(i)}}, \text{and}$$

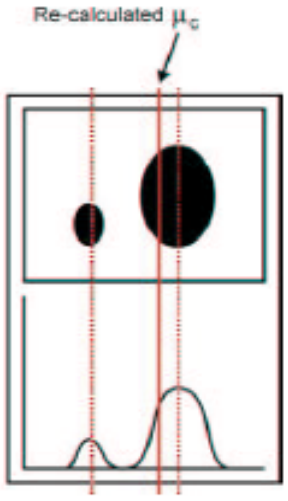
$$\sigma_r = \sqrt{\frac{\sum_i (x_r - \mu_r)^2 h_r(i)}{\sum_i h_r(i)}}$$

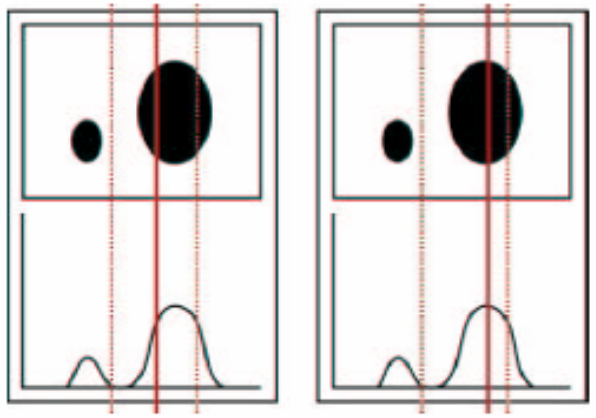
- means and standard deviations of the distributions would represent the blob's centroid co-ordinates and values related to its dimensions
- multiple blobs to be present:
- identifies that centre of the largest blob in the image.

- the mean and standard deviation of the distribution are calculated using all samples
- the mean is recalculated using samples within one standard deviation of the original mean.
- repeated until the change in the values of the two means was a negligible
- the mean will lie within the most dominant (largest) blob



the mean and standard deviation is computed.

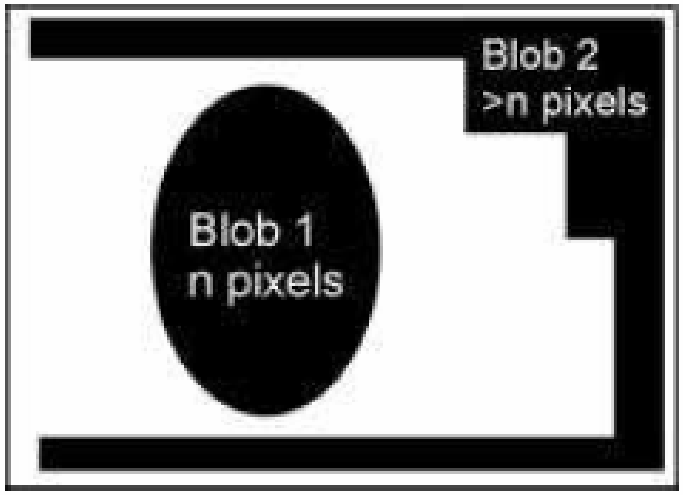




Face regions detected

values of the standard deviations can be used to compute an approximate bounding box for the face region.

Problem left



An example of connected component analysis giving incorrect results.

nostrils tracking

- lie inside the face blob.
- reduce the search space to the centre portion of this blob

  (a region whose linear dimensions are one third of the face's bounding box)
- The raw image data in the search region is thresholded with a gradually reducing threshold until two regions that match our nostril heuristics are found.

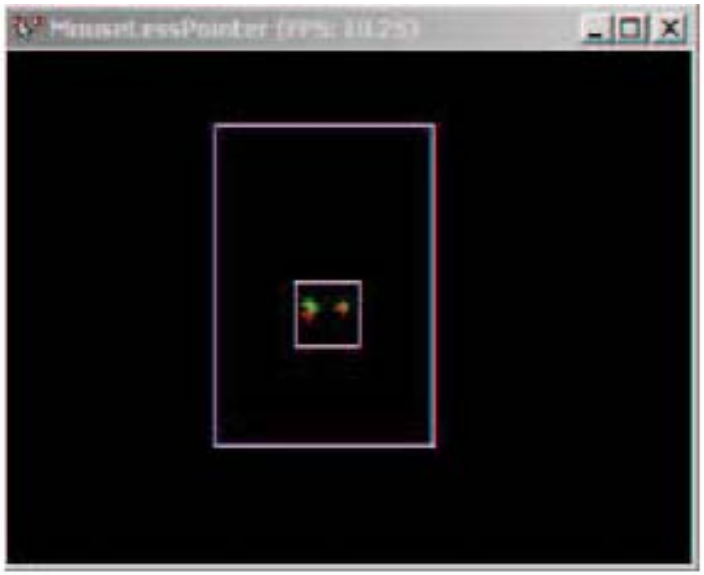


Figure 6. Tracking the nostrils in real-time. The larger square is the minimum bounding rectangle

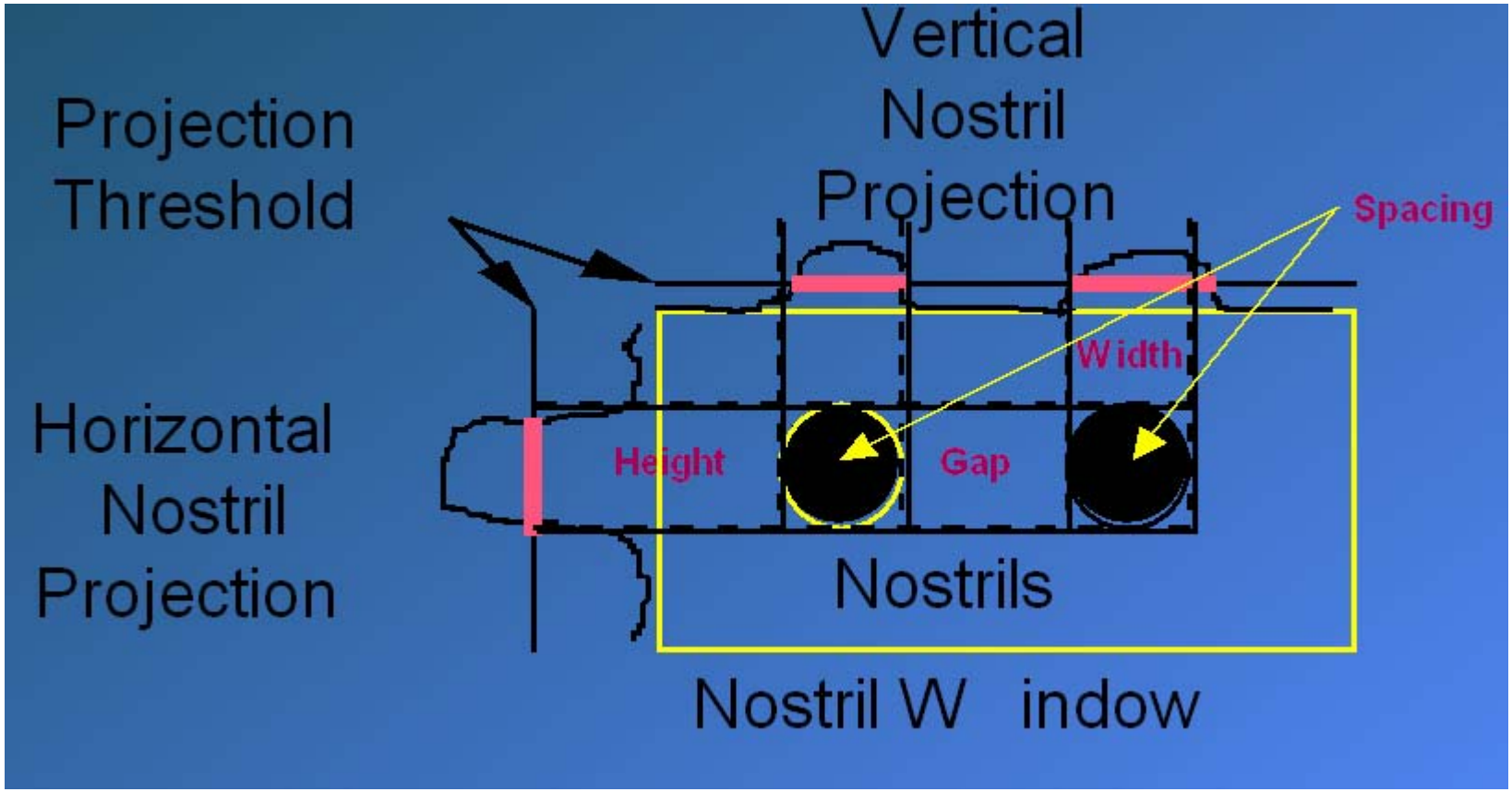
- the location of nostrils is used to update the centre of the search region for the following frame.
- If the nostril search fails, the search for the nostrils is reinitiated.

(i) (ii) (iii)

Figure 9: Sample results of the nostril tracking stage.

Another scheme:

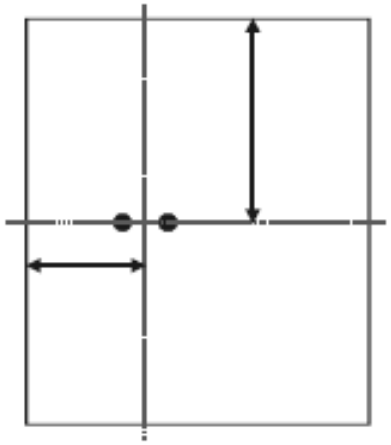
- Nostril are detected only if:
  - At least 75% of nostril window area is skin color
  - After RGB thresholding nostril window, at least 15% of area is subthreshold(nostril)
  - Min/Max constraint are met for nostrilwidth,height,gap,center spacing,orientation in thresholded projection domain


Vertical
Nostril
Projection
Projection
Threshold
Spacing
Width
Horizontal
Nostril
Projection
Height
Gap
Nostrils
Nostril W indow

Cursor Movement

- Given the co-ordinates of the nostrils' active point and the co-ordinates of the face search area, driving the cursor's movement.
- Jitter Problem:defined as randomised apparent movement of the nostrils, due to small amplitude, random head movements (tremor) and errors in the estimation of nostril location.
- Solution:values change by more than a predefined threshold, the new values are used to update the cursor position,

- The distances from the nostril point to the boundaries of the face region in the vertical and horizontal directions is computed
- cursor position coordinates by linear scaling



Video capture interface to PC

- the standard video camera plus video capture hardware and digital cameras that interface directly to the system
- a webcam interfaced via USB
- minimum rate of 10 frames per second must be processed

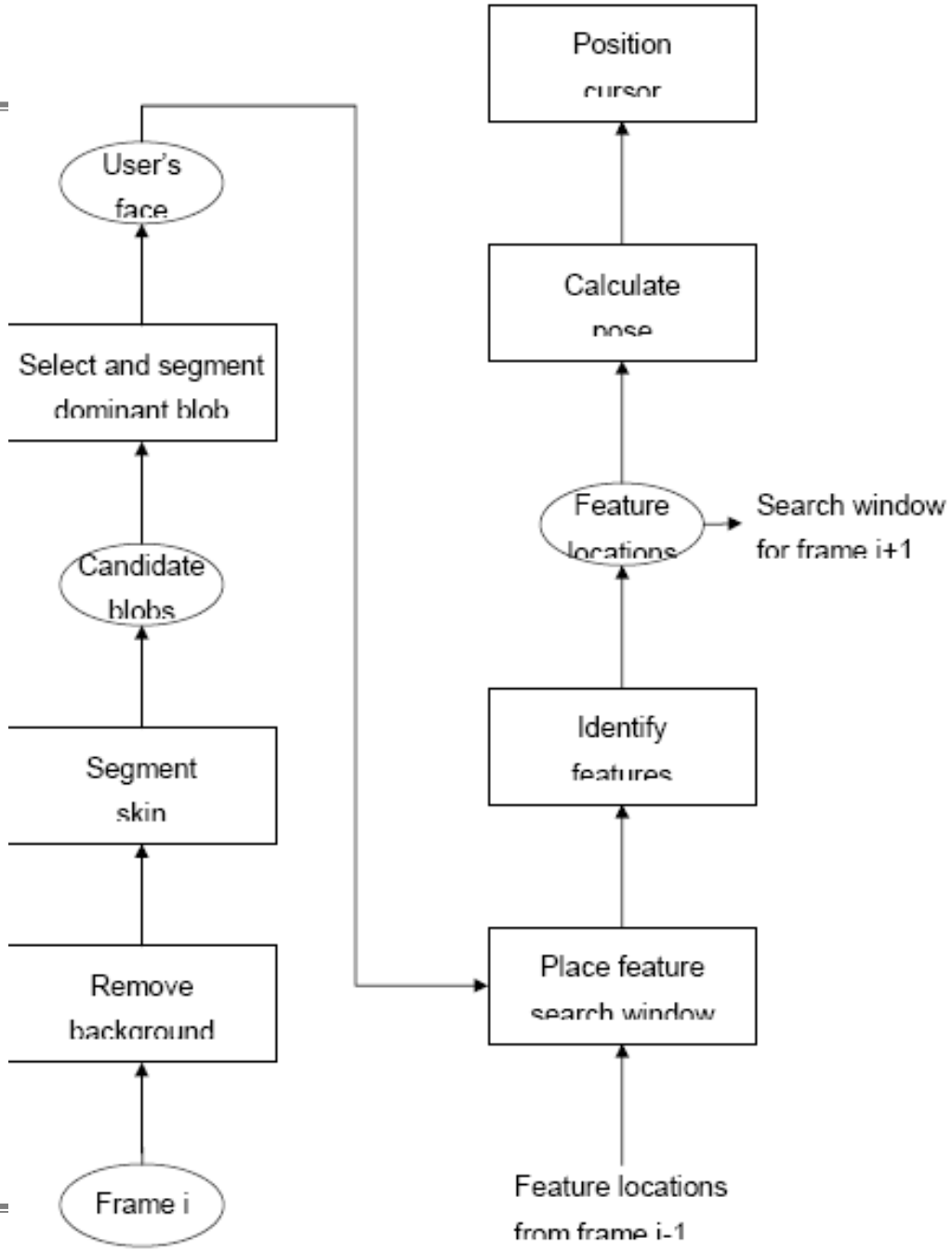


Figure 3: Data flow diagram.

Implementation and Performance Evaluation

- Two phases:
  - initial calibration phase:
    - Aquire the background of image,if warranted
    - captures the skin colour values for this particular user [1]
  - Real-time tracking phase:

result

- at the time of development, a PIII processor witha clock speed of 500 MHz
- achieved throughput rates of 30 frames per second at a resolution of 160 by 120 pixels, and 18 frames per second at 320 by 240 pixels.
- As the illumination is reduced, darker skin tones result in tracking failure sooner than lighter tones due to the lower contrast between the skin and the nostril areas.

- The system can be forced into tracking errors by introducing any large skin-coloured object into the field of view.

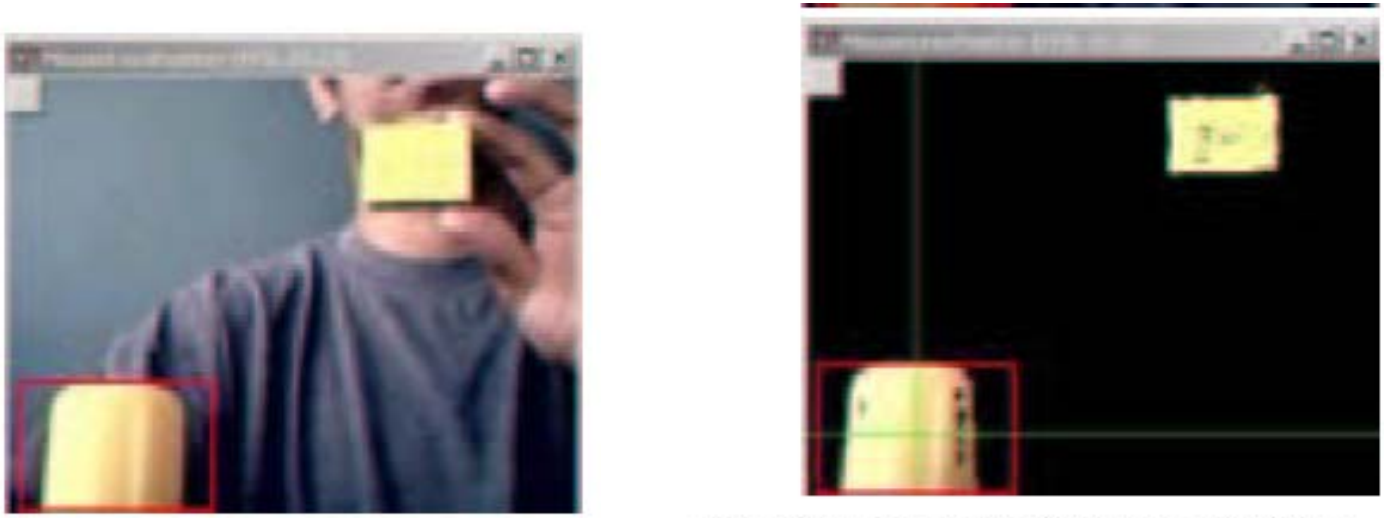


Figure 10. Tracking the dominant object.

参考文献

1.Dr. T. Morris,Facial Feature Tracking For Cursor Control, **Journal of Network and Computer Applications (2006)**

问题与回答

谢谢！