International Journal of Soft Computing 1(2) : 103-107, 2006
© Medwell Online, 2006

An Escalated Approach to Face Tracking and Facial Expression Recognition from Video Files

R. Baskaran and A. Kannan
Department of Computer Science and Engineering, College of Engineering,
Anna University Chennai 600025, India

Abstract: Tracking faces and identifying expression from a video files is a complex mining application due to the fact that even a delicate change in the emotion of the face could convey a great variation as an expression. This paper presents a robust novel method to track both male and female faces and recognize the facial expression in video files. Face localization in a digital image, which is being grabbed from the video source is performed using the hue-saturation and YC_bC_r based algorithms. Morphological algorithm over this region signals the eyes and mouth position in that region. Lips, eyebrows and eyelids are then identified. Distance between eyebrow and eyelid and the slope of the curvature of the lips are derived which facilitates in deciding the facial expression of the given digital image.

Key words: Facial expression recognition, face detection, face tracking

INTRODUCTION

Every emotion of an individual is being revealed out with his or her facial expression. Mild change in facial expression could convey a meaning, which is very difficult to verbally explain. This will establish a personal rapport among individuals. Automatically recognizing such expression from the information available in a digital image or video file is a highly challenging task.

Yongsheng Gatheo *et al.*[1] have represented faces as caricatures to recognize expressions. They proposed a facial expression recognition method from line-based caricatures. Edges are detected based on the algorithm of Nevatia followed by a thinning process to generate one-pixel wide edge curves. To generate the line edge map, the dynamic two-strip algorithm Dyn2S is utilized. This approach uses line edge map as expression descriptor. Similarity is obtained by computing the proposed directed line segment Hausdorff distance between the query face line edge map and the caricature models of expressions. A line-based caricature depicts a facial expression by a few lines, which are the common features of that particular expression and independent of individual subjects.

Ying-li Tian *et al.*[2] have developed an Automatic Face Analysis system to analyze facial expressions based on both permanent and transient facial features in a nearly frontal-view face image sequence. Multi state face and facial component models are proposed for tracking and modeling the various facial features, including lips, eyes, brows, cheeks and furrows. During tracking, detailed parametric descriptions of the facial features are extracted. With these parameters as the inputs, a group of action units are recognized.

In Javad Haddadnia *et al.*[3] have proposed a new face localization algorithm. This algorithm is based on the shape information considered with a new definition for distance measure threshold called Facial Candidate Threshold. In the feature extraction stage they have defined a new parameter, called the Correct Information Ratio, for eliminating the irrelevant data from arbitrary face images. Pseudo Zernike Moment Invariant is utilized to obtain the feature vector of the face under recognition. To find a face region, an ellipse model with five parameters is used: X0; Y0 are the centers of the ellipse, the orientation and lengths of the minor and the major axes of the ellipse, respectively.

Rein-Lien Hsu *et al.*[4] had two major modules in detecting faces. Initial module is face detection to identify face. Face verification module ensures that no other skin regions like hands or neck are treated as face. They first estimated and corrected the color bias based on a light compensation technique. These corrected red, green and blue color values are then subjected to nonlinear transformation to YC_bC_r color space to detect skin regions in a digital image. Once skin region is identified, eye, mouth and boundary maps are found. These maps are used for verifying the skin region to be face or not.

Donato *et al.*[5] rejected regions that do not have skin tone colors from the input image. This is based on color threshold approach and skin segmentation. Initially they rejected all the non-face regions from the image. This is achieved by making an overlapped blend of image in RGB, HSV and YC_bC_r color space. Then they used binary image processing to create clearer delineations in these regions. Template matching with both training image faces and eigen faces is followed to detect faces.

Many of the existing systems intend to recognize the facial expression from static image source. In this study we intend to recognize the facial expression from static image files and also from video files.

SYSTEM ARCHITECTURE

The overall system architecture providing interaction between various modules is depicted in Fig. 1.

Input processing: When any video file is accepted as input, its every frame is grabbed as an image. This is achieved by decoding the video file and virtually running it to get video buffer information that can be taken as an image input. Location of the face in a particular frame is also used to track the face in next frame. This is done to fasten the face detection process in case of video files.

Face localization module: Face Detection in performed using the hue-saturation and YC_bC_r values of the regions in the images[6]. This will now help to remove all unwanted background regions other than skin region.

Morphological operation module: Facial part of the video frame is available as binary image with skin-tone region and non-skin regions like background, hair eye and teeth. This binary image is dilated and then eroded. This will help to create mask for the eyes and mouth.

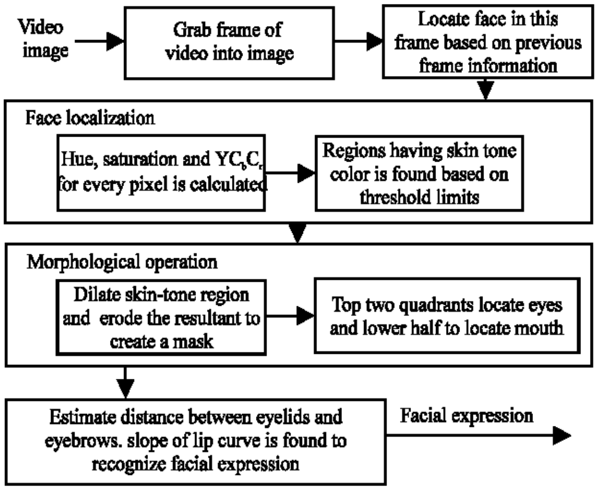
Video image
Grab frame of video into image
Locate face in this frame based on previous frame information
Face localization
Hue, saturation and YCbCr for every pixel is calculated
Regions having skin tone color is found based on threshold limits
Morphological operation
Dilate skin-tone region and erode the resultant to create a mask
Top two quadrants locate eyes and lower half to locate mouth
Estimate distance between eyelids and eyebrows. slope of lip curve is found to recognize facial expression
Facial expression

Fig 1: System architecture of facial expression recognition using quadra color representation

Recognizing facial expression: Left eye is located by applying the mask over top left quadrant of the extracted video fragment. Similarly right eye is located in the top right quadrant and mouth is located in the bottom half. We then find the distance between eye and eyebrow. Abnormal length indicates that the person is surprising. Otherwise with the lips, we try to find the group of pixels in the mouth mask region. We then, find the slope of such a curve shaped group of pixels.

FACE DETECTION

When we deal with video files, we will get image by storing each and every frame as a static image. This module can be viewed as two major sub modules namely face localization and face verification. Initially face identification is done to reject non-skin regions[7] in the image and then face verification process to ensure the skin region is face.

Face localization: In this module, non-face region rejection is the primary target. In order to achieve this Hue-saturation and YC_bC_r values[8] are found using the following expressions.

$$h = \frac{\cos^{-}\left(\frac{(R-G)+(R-B)}{2}\right)}{\sqrt{(R-G)^{2}+(R-B)(G-B)}} \qquad (1)$$

$$s = 1-\left(\frac{3}{R+G+B}\right)\times\min(R,G,B) \qquad (2)$$

$$Y = 0.257*R+0.504*G+0.098*B+16 \qquad (3)$$

$$C_b = 0.148*R+0.291*G-0.439*B+128 \qquad (4)$$

$$C_r = 0.439*R-0.368*G-0.071*B+128 \qquad (5)$$

Once the hue-saturation YC_bC_r values are obtained, pixels having values similar to skin are subjected to region growing based segmentation algorithm. Configurable parameters are used to set threshold values to drop the regions that do not have minimum number of pixels, area of the region, length and breadth of the regions. Shape descriptors are used to decide on the shape of the region having the face.

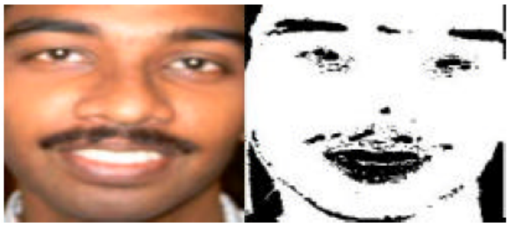

Fig. 2: Skin tone region and its binary representation

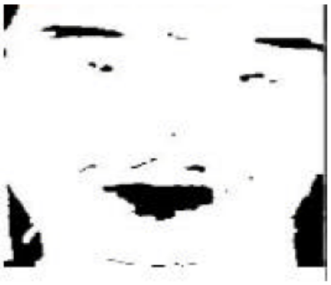

Fig. 3: Dilated representation of the Skin-tone representation

Face verification: Face verification is performed to ensure that the identified skin color region[9] is face but not hands or neck. For this we initially create a mask by creating a binary image of skin and non-skin region. This is a rectangle inside the main video frame rectangle. In this binary image, white pixels will represent skin-tone region in the video frame image. Non-skin regions, which include background, nails, hair, eyebrows and teeth, will be represented by black pixel. A sample video frame fragment and its skin-tone binary representation are given in Fig. 2.

This binary image is subjected to dilate operation, which is performed to close all small holes in the image. The image in Fig. 3 provides the dilated version of the above shown binary image shown in Fig. 2.

The algorithm used to perform the dilate operation is given below.

```
for every target pixel location (x,y)
{
   target[x][y] = 0x0;
   for (i = -pivotx; i < kernelRowCount - pivotx; i++)
   {
      for (j = -pivoty; j < KernelColumnCount - pivoty; j++)
      {
         if((x+i, y+j) are in bounds of source && (pivotx+i, pivoty+j) are in bounds of Kernel)
         {
         tmp = source[x + i][y + j]+ Kernel[pivotx + i][pivoty + j];
target[x][y] = max{tmp, target[x][y]};
         }
      }
   }
}
```

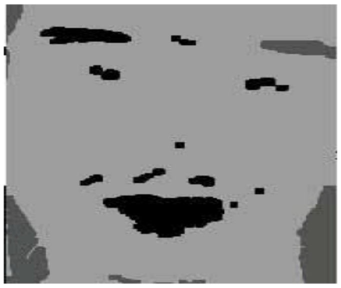

Fig. 4: Erosion applied over dilate representation of the skin-tone representation

The resultant values of the dilate operation are subjected to erode operation which is performed to magnify holes. After these operations are performed we get an image with large non-skin portions like eyes, eyebrows and mouth to be represented as holes, which will serve as mask. When there is no hole, then this signals the fact that this video frame does not hold any facial information. Figure 4 shows the erosion over dilation of the binary representation of the skin tone region.

The algorithm in erode operation is given below.

```
for every target pixel location (x,y)
{
  tmp = 0xff;
  for (i = -pivotx; i < kernelRowCount - pivotx; i++)
  {
    for (j = -pivoty; j < kernelColumnCount - pivoty; j++)
    {
      if((x+i, y+j) are in bounds of source)
      {
       tmp = min{tmp, source[x + i][y + j] - Kernel[pivotx + i][pivoty + j]};
      }
    }
  }
  target[x][y] = tmp;
}
```

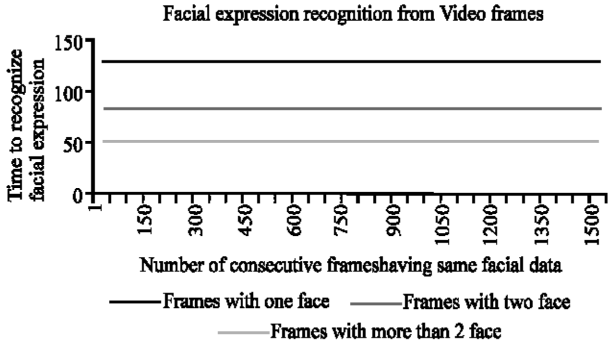

Fig. 5: Graph depicting the analysis of video files having frames with one, two or more than 2 faces

Now this morphologically operated image will serve as a mask to confirm whether the video frame imagefragment has facial data or not[10]. If the top left quadrant has a black hole and also the actual image corresponding to that hole has eyebrow or eyelid or eye's hue, saturation and YC_bC_r values, we confirm left eye is present. Similarly if the top right quadrant has a black hole and also the actual image corresponding to that hole has eyebrow or eyelid or eye's hue, saturation and YC_bC_r values, we confirm right eye is present. Also the bottom half should have a black hole and also the actual image corresponding to that hole has mouth or lip or teeth or tongue's hue, saturation and YC_bC_r values, we confirm the presence of mouth. We take the facial portion of the digital image for the processing in this module. If any of this fail to work out positively, then we decide that there exists no facial data.

FACIAL FEATURE EXTRACTIONS

Once it is sure that the skin-tone region under consideration is surely a facial region, we move for recognizing the facial expression. The algorithm given below provides a step-by-step procedure to recognize the facial expression from this centre point.

Step 1: From the centre point, we move down the pixels to find a point a where a 0×00000000 value is available. This signals the location of lips. We record this position.

Step 2: Similarly moving in the north east direction, 45 degrees top right to the nose top find a pixel with 0×00000000 value which signals the location of the eyes.

Step 3: This point is taken as pivot point to perform region growing to find the pixels contributing to the eyes and eyelids.

Step 4: With this top centre point of the eyelid is found.

Step 5: Keep moving up to find a pixel with 0×00000000 value, which indicates the presence of eyebrow.

Step 6: The distance between eyelid and eyebrow is calculated.

Step 7: If this is abnormal it denotes surprising expression.

Step 8: If it is normal, proceed with lips. In this case, traverse from the centre pixel of the bottom row of the image up and up to location of pixel having 0×000000 value. This is the pivot point for region growing of lips portion.

Step 9: Find the pixels contributing to the lips part of the face.

Step 10: Now the extreme points in this face are identified as x_{min} y_{min} x_{max} and y_{max}.

Step 11: Now the slope is found using the relation.

$$1 = \frac{y_{max} - y_{min}}{x_{max} - x_{min}} \qquad (13)$$

This slope is used to decide on the facial expressions as neutral, smiling or screaming.

PERFORMANCE EVALUATION

This algorithm is tested with facial expression of 30 different individuals using a 2.4GHz system having 512MB RAM. The video collection of images includes both male and female faces. There are faces, which had orientation in the face. Male face images with and without the presence of mustache and beard are tested.

Video files are tested to explore the rate of recognizing the facial expressions. Frames that contain more than one face are also tested. Tracking of the facial part is performed. When the number of subsequent frames containing facial region increases, rate of expression recognition also increases. Figure 5 presents a graph that provides the testing results with video files.

CONCLUSION

This approach provides promising results with various types of video files that include both male and female faces. This also stands strong with orientation of the face. This is stable with video files having faces with mild beard and mustache also. We aim at indexing video files based on the expression, which could help in faster retrieval of image and/or video files. Also indexing the streamed video data indexing is also targeted.

REFERENCES

1. Yongsheng Gao, K. H. Maylor Leung, Siu Cheung Hui and M.W. Tananda, 2003. Facial Expression Recognition From Line-Based Caricatures. IEEE Transactions On Systems, Man and Cybernetics, 33: 407-412.
2. Ying-li Tian, K. Takeo and F.C. Jeffrey, 2001. Recognizing action units for facial expression analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23: 97-115.
3. Haddadnia, J., F. Karim and A. Majid, 2003. An efficient human face recognition system using pseudo zernike moment invariant and radial basis function neural network. Intl. J. Pattern Recognition and Artificial Intelligence, 17: 41-62.
4. Hsu, R.L., M. Abdel-Mottaleb and A.K. Jain, 2002. Face detection in color images. IEEE Transactions on Pattern Analysis and Machine Intelligenc.
5. Donato, M.S., J.C. Bartlett, P. Hager, Ekman and T.J. Sejnowski, 1999. Classifying facial actions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21: 974-989.
6. Yang, M.H., D.J. Kriegman and N. Ahuja, Detecting faces in images: A Survey IEEE Transactions on Pattern Analysis and Machine Intelligence, 24:
7. Teodorescu, T.D., V.A. Maiorescu, J.-L Nagel and M. Ansorge, 2003. Two color based face detection algorithms: A Comparative Approach. International Symposium on Signals, Circuits and Systems, 1pp: 10-11.
8. Lee, C.H., J. S. Kim and K. H. Park, 1996. Automatic human face location in a complex background using motion and color information. IEEE Transactions on Pattern Recognition, 24: 1877-1889.
9. Sanjay, Kr., 1 Singh, D.S. Chauhan, V. Mayank and S. Richa, 2003. A robust skin color based face detection algorithm. Tamkang J. Sci. Eng., 6: 227-234.
10. Yacoob, Y. and L. Davis, 1996. Recognizing human facial expression from long image sequences using optical flow. IEEE Transactions on Pattern Analysis and Machine Intelligence, 16: 636-642.