IJIIE-International Journal of Innovations & Implementations in Engineering

(ISSN 2454- 3489) 2016 July Edition Volume 1

Robust Human Tracking Using Sparse Collaborative Model in Surveillance Videos

NIJU T.S

*P G Scholar, Department of ECE,
Rajadhani Institute of Engineering and Technology, Trivandrum,
ts.niju6994@gmail.com

NANDHU.B

*Assistant Professor, Department of ECE,
Rajadhani Institute of Engineering and Technology, Trivandrum,
nandhub@rietedu.in

ABSTRACT : Robust human tracking becomes a task when the large appearance change occurs. It is due to varying illumination conditions, clustered background and camera motion. Here we use a skin detection method and a sparse collaborative model for tracking the human. Skin detection approach consists of a smooth 2-D histogram and Gaussian model for automatic skin detection in color images. The proposed method will reduce the computational costs as it requires no training. This method consists of a Sparse Collaborative Model which includes SDC (Sparse Discriminative Classifier) and SGM (Sparse Generative Model) for human tracking. In SDC, a classifier is used that separates the foreground object from the background based on the holistic templates. In SGM, an histogram based method is used. In both SDC and SGM an update schema is used in which they get updated independently.

I. INTRODUCTION

An image plays an important role in the current society. A wide range of image processing applications like gesture analysis, human–computer interaction domains was based on skin detection. For each color space components multiple ranges of threshold values are defined and the images whose pixel values that fall within these range(s) are selected as skin pixels. For detecting human skin, a training stage is required to define threshold values [1]. The first step in skin segmentation is the image pixels representation in a suitable color space. Illumination variation can degrade the skin detection systems performance. RGB, YCbCr, HSV, CIE Lab, CIE Luv, and normalized RGB color spaces are used for skin-color representation.

A generative or discriminative appearance model is needed to effectively verify state predictions. For generative methods, tracking is formulated as searching neighborhood. For discriminative approaches, tracking is posed as a binary classification problem which aims to design a classifier for distinguishing the target object from the background. Most tracking methods use rectangular image regions to represent targets, and thus background pixels are inevitably included as part of the foreground objects [10]. The holistic appearance encoded by a target template is more distinct than the local appearance of local patches. A collaborative observation model that integrates a discriminative classifier based on holistic templates and a generative model using local representations.

The process of finding skin-color pixels and regions in an image or video is Skin Detection. Skin color is an indication of the existence of humans in media. False skin detection is a problem since there are a wide variety of skin colors. In our tracking

algorithm, both the confidence value based on the holistic templates and the similarity measure based on the local patches contribute to an effective and robust probabilistic appearance model. Skin detection in images is a theme that is present in many applications [2]. This is the first step for faces recognition. Several factors need to be considered for an effective appearance model. First, an object can be represented by different features such as intensity, color, texture, super pixels, and Haar-like features. The representation schemes can be based on holistic templates or local histograms. The collaboration of the generative model and the discriminative classifier leads to a more flexible and robust likelihood function to verify the state predictions [10].

In this approach the number of discriminative features is fixed, which may not be effective for tracking in dynamic and complex scenes.

II. PROPOSED METHOD

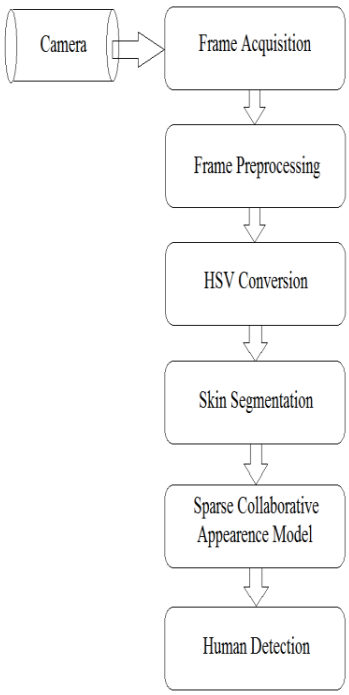


Fig 1 Block Diagram

A. FRAME ACQUISITION

To get images from any source, usually a hardware source is known as image acquisition. Using a camera, the moving images are recorded as video frames. The human detection and tracking system reads each image and process it to extract the skin features. The image can be accessed and handled as a three dimensional matrix in RGB color space. It compares each and every pixel in the subsequent frames. Images at two different instances i.e. previous image and current image are subtracted for detecting the moving object. This method continuously keep making background image using previous frames in real time. The image size depends on the resolution of the images [6]. During further processing, increasing image size will affect the entire processing of the system. So as per the requirements, the resolution of the image can be fixed.

B. FRAME PREPROCESSING

It is the improvement of digital image quality, without knowledge about the source of degradation. This technique improves the interpretability or perception of information in images for human viewers. It is to improve the image quality so that the resultant image is better than the original image for a specific application. Image enhancement algorithms tend to be simple, qualitative, and ad hoc. Image is also standardized by resizing the image to appropriate requirement. There are different types of noise that corrupt the image such as additive noise, Gaussian noise, impulse noise and Poisson noise etc to remove these types of noises there are various filters are available such as Gaussian filter, Median filter, High pass filter, Low pass filter.

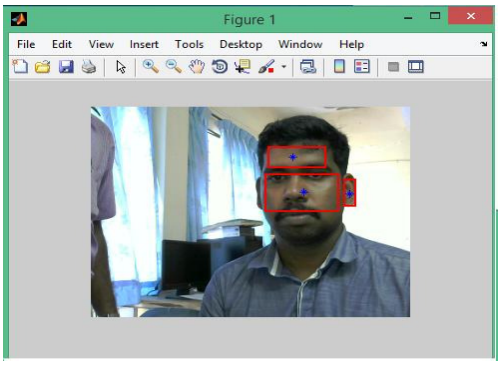
These noises can be minimized, using a 5x5 structuring element in morphological filters [6]. First used the structuring element with a dilatation filter that expands the areas in the skin regions. After that the same structuring element was used to erode the image and reduce all the imperfections that the dilatation created. These techniques were used, by approximation, to fill all the spaces that were by H channel range supposed that is skin or non-skin.

To evaluate the accuracy of this work for skin detection, scan the entire image to count

the skin and non-skin pixels location. These values are used to evaluated the false positive (FP), false negative (FN), true positive (TP), true negative (TN), the rate of success (Suc.) and errors (Err.)

C. HSV CONVERSION

In RGB model, each color appears in its primary spectral components of red, green and blue. . HSV is the representation of points in RGB model into cylindrical coordinates. Hue is an attribute with the dominant wavelength is a mixture of light waves which represented as an angle in cylindrical coordinates. Saturation refers to the relative purity or amount of white light mixed with hue [6]. Value is the intensity of pixel which ranges from 0 to 1. The radial position of the cylindrical coordinate system represents the saturation and vertical position represents the intensity value.



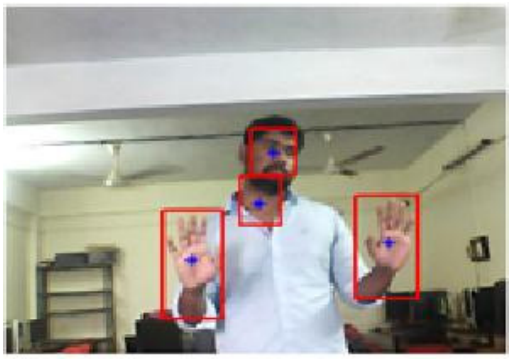


Fig 2 Original Image

The HSV conversion can be consummated by the following equations.

$I=R+G+B$

$r=R/I$

$g=G/I$

$b=B/I$

$h = \cos^{-1} \{ [R - \frac{1}{2}G - \frac{1}{2}B] / [\sqrt{(R^2 + G^2 + B^2)} - RG - RB - GB]\}$

$H= h ; B<=G$

$H= 360-h ; B>G$

Calculate the probabilities of certain colors representing skin within the rest of the given image. Using a threshold in conjunction with these probabilities will result in the successful segmentation of skin [6]. A threshold may define pixels with a skin likelihood of >0.5 as skin, and those with likelihoods of <0.5 as non-skin. Thresholds that serve more sophisticated segmentation methods can easily be applied

Fig 3 Image in HSV

Fig 4 Final Image

D. SKIN SEGMENTATION

Skin color segmentation is normally considered to be a low-level extraction. The segmentation technique, which uses all 3 color spaces was designed to boost the face detection accuracy. A significant benefit of using a Gaussian function to model a skin color distribution is that it will allow us to calculate the probability of specific colors representing skin [6]. The color distribution of skin, as it appears in an image, is the combination of its chromaticity and its illumination which varies from person to person. Applied across an entire image, we obtain a "skin likelihood image", wherein every pixel is represented by the probability of its original color being skin. We cannot achieve the segmentation of any skin without any form of thresholding. A two-dimensional Gaussian function within the normalized rg color space with our filtered pixel set, can then be used to calculate the probabilities of certain colors representing skin within the rest of the given image. Using a threshold in conjunction with these probabilities will result in the successful segmentation of skin.

A threshold may define pixels with a skin likelihood of >0.5 as skin, and those with likelihoods of <0.5 as non-skin. Thresholds that serve more sophisticated segmentation methods can easily be applied.

E. SPARSE COLLABORATIVE APPEARANCE MODEL

Most of the tracking methods uses rectangular image regions for representing targets. The classifiers based on local representations are significantly affected while considering background patches as positive ones. The holistic appearance of a target template is more distinct than the local appearance of local patches [10]. For separating foreground objects from the background the holistic templates are more effective for discriminative models. A collaborative observation model integrates a discriminative classifier based on holistic templates and a generative model using local representations.

1. Sparse Discriminative Classifier (SDC)

Motivated by the demonstrated success of sparse representation for vision tasks, it proposes a sparse discriminative classifier for object tracking. In the following, it use the vector **x** to represent intensity values of a raster scanned image

a. Templates:

Each down sampled image is stacked together to form the set of positive templates. The negative training set is composed of images away from the target location. Thus, the negative training set consists of both the background and images with parts of the target object [10]. Thus better object localization as samples containing only partial appearance of the target are treated as the negative samples and the corresponding confidence values are likely to be small

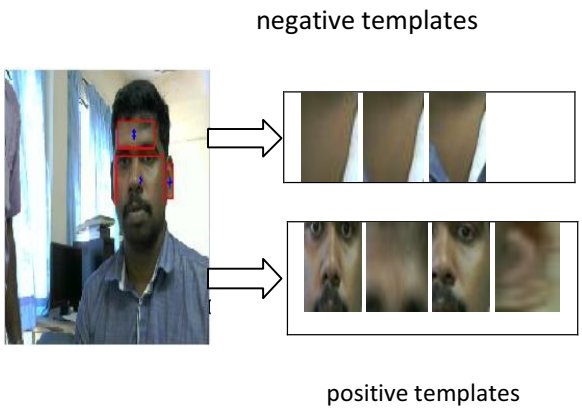


Fig 5 Positive and negative templates in the SDC Module

b. Feature Selection:

The above-mentioned gray-scale feature space is rich yet redundant, from which determinative ones can distinguish the foreground object from the background. It can be extracted by learning a classifier [10]. The feature selection scheme adaptively chooses suitable number of discriminative features in dynamic environments using the *l1* constraints. It project the features to a subspace via a projection matrix **S** which is formed by removing all-zero rows from a diagonal matrix **S** and the elements are determined by

$$S_{ii} = 0, \; s_i = 0$$

$$1, \text{ otherwise.}$$

The discriminative feature space is projected by the training template set and the candidates. The training template set and candidates in the projected space are $\mathbf{A} = \mathbf{SA}$ and $\mathbf{x} = \mathbf{Sx}$.

c. Confidence Measure:

The proposed SDC method is developed based on which a target image region can be better represented by the sparse combination of positive templates. While a background patch can be better represented by the span of negative templates. The error is computed based on the target (positive) templates, which is less effective for tracking. Both the negative and indistinguishable samples have large reconstruction errors when computed with the target (positive) template set [10]. The confidence measure exploits the distinct properties of the foreground and the background in computing the reconstruction errors to better distinguish patches from the positive and negative classes.

2. Sparse Generative Model (SGM)

The recent advances of sparse coding for image classification as well as object tracking, presents a generative model for object representation that takes local appearance information of patches and occlusions into consideration.

a. Histogram Generation:

For representing the local appearance information of a target object the grayscale features are used where each image is normalized to 32 × 32 pixels. It use overlapped sliding windows on the normalized images to obtain M patches and each patch is converted to a vector $\mathbf{y}_i \in R^{G\times 1}$, where G denotes the size of the patch

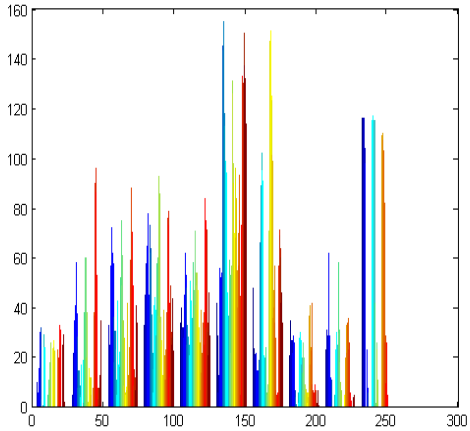


Fig 6.Histogram Generation

b. Occlusion Handling:

To deal with occlusions, the constructed histogram has to be modified for excluding the occluded patches when describing the target object [11]. A patch with large reconstruction error is regarded as occluding part and the corresponding sparse coefficient vector is set to be zero. The proposed representation scheme takes spatial information of local patches and occlusion into account, thus making it more effective and robust.

c. Similarity Function:

For computing the similarity of histograms between the candidate and the template due to its effectiveness by the vector **o** reflects the occlusion condition of the corresponding candidate by the histogram intersection function. The comparison between the candidate and the template should be carried out under the same occlusion condition, so the template and the c-th candidate share the same vector $\mathbf{o}_c$.

3. Collaborative Model

It proposes a collaborative model using the SDC and the SGM modules within the particle filter framework. In this both the confidence value based on the holistic templates and the similarity measure based on the local patches contribute to an effective and robust probabilistic appearance model [11]. The confidence value *Hc* gives higher weights to the candidates considered as positive samples and penalizes the others. It can be considered as the weight of the local similarity measure *Lc*

4. Update Scheme

For the SDC module, update the negative templates every several frames (5 in our experiments) from image regions away (e.g., more than 8 pixels) the current tracking result [10]. The positive templates remains the same in the tracking process. As the SDC module aims to distinguish the foreground from the background, it is important to ensure that the positive and negative templates are all correct and distinct. In order to capture the appearance changes and recover the object from occlusions, the new template histogram ψn is computed by

$$\psi n = \mu\psi f + (1 - \mu)\psi l \text{ if } On < O0$$

F. HUMAN DETECTION

For identifying the presence of human, skin detection is the fastest method. Other parts like nose, eyes, hair, mouth etc can be detected but these are sometimes very difficult if the object is far from the c. As skin occurs more in area than others, it is better to detect skin tone for human detection. Human skin color ranges from darkest brown to pinkish-white hues. The human skin colors pigments will gather in small regions and differ more in brightness than in colors [6]. Color component values are normalized with intensity values of the image.

III. RESULTS

In this proposed system human skin detection is done. Based on hsv color conversion and tracking the image on various conditions like rotation motion blur, complex background. A sparse collaborative appearance model is used to achieve a better tracking and to reduce the drifts like rotation, motion blur, complex background

IV. CONCLUSION AND FUTURE SCOPE

A fusion framework based on smoothed 2D histogram and Gaussian model has been proposed to automatic detect human skin in images. Here an effective and robust tracking method based on the collaboration of generative and discriminative modules. In the proposed tracking algorithm, holistic templates are incorporated to construct a discriminative classifier that can effectively deal with cluttered and complex background. The proposed method outperforms state-of-the-art methods in terms of accuracy in different conditions: background model, illumination, and ethnicity.

REFERENCES

[1] M.-J. Seow, D. Valaparla, and V. K. Asari, “Neural network based skin color model for face detection,” in *Proc. Appl. Image Pattern Recognit.Workshop*, 2003, pp. 141–145.

[2] D. Brown, I. Craw, and J. Lewthwaite, “A SOM based approach to skin detection with application in real time systems,” in *Proc. Brit. Mach.Vis. Conf.*, 2001, pp. 491–500

[3] N. Sebe, I. Cohen, T. S. Huang, and T. Gevers, “Skin detection: A Bayesian network approach,” in *Proc. Int. Conf. Pattern Recognit.*, 2004, pp. 903–906.

[4] R. Khan, A. Hanbury, and J. Stoettinger, “Skin detection: A random forest approach,” in *Proc. Int. Conf. Image Process.*, Hong Kong, 2010, pp. 4613–4616.

[5] U. Yang, B. Kim, and K. Sohn, “Illumination invariant skin color segmentation,” in *Proc. 4th IEEE Int. Conf. Ind. Electron. Appl.*, May 2009, pp. 636–641.

[6] Aneesh R P, Arun p- Implementation of an Effective and intelligent system for human detection and targeting, IRACST international Journal of advanced Computing, Engineering and

Application(IJACEA), ISSN, vol. 3,No.4, August 2014

[7] Jianpeng Zhou, Real Time Robust Human Detectionand Tracking System, Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference

[8] Swapnil V. Tathe and Sandipan P. Narote,Real-Time Human Detection and Tracking, 2013 Annual IEEE India Conference (INDICON)

[9] S. Gao, I. W.-H. Tsang,L.-T. Chia, and P. Zhao. Local features are not lonely - laplacian sparse coding

[10] B. Liu, J. Huang, L. Yang, and C. Kulikowsk. Robust tracking using local sparse appearance model and k-selection. In CVPR, 2011.