

Human Pose Estimation in Images and Videos

P. Reddy Gurunatha Swamy^{1*}, B. Ananth Reddy²

¹Department of ECE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India.

²Department of ECE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India.

*Corresponding author E-mail: bananthreddy30@gmail.com

Abstract

Estimation of human poses is an interesting and challenging topic in the field of Computer vision. It includes some un-noticed challenges like background effect, the color of the dress, skin tones and many other unpredictable challenges. This is a workable concept because it can be used in sign language recognition, correlating various pose styles from different parts of the world and in medical applications. A deep structure which can represent a man's body in different models will help in improved recognition of body parts and the spatial correlation between them. For hand detection, features based on hand shape and representation of geometrical details are derived with the help of hand contour. An adaptive and unsupervised approach based on Voronoi region is primarily used for the color image segmentation problem. This process includes identification of key points of the body, which may include body joints and parts. The identification parts will be tough due to small joints and occlusions. Identification of Image features is described in this paper with the help of Box Model Based Estimation, Speed up robust features and finally with Optical flow tracking algorithm. In Optical flow tracking algorithm, we have used Horn-Schunk algorithm to determine featural changes in the images.

Key words: Voronoi segmentation, box model, optical flow tracking, SURF features.

1. Introduction

The process includes identification of key points of the body, which may include body joints and parts. The identification parts will be tough due to small joints and occlusions. Before we work on images and videos, it is imperative that we go through some of the basic and morphological operations and pre-processing techniques used to process images and videos, without which a proper study cannot be satisfactorily completed. Images obtained are not perfect as they are affected by a plethora of external and internal disturbances, thereby causing noise, blurs, etc. Especially, the binary images obtained by normal thresholding are affected by various amounts of noise and texture. Therefore, before we go to extract useful features from the image, we should do pre-processing operations on the image for better processing and results. The term 'Morphological operations' accounts for these basic operations such as filtering to remove noise and other operations based on the form and structure of the image, applied primarily to grayscale images. Morphological operations in image processing generally refer to a collection of non-linear operations related to the morphology or the structural features of an image. These operations primarily rely on the relative positioning, structure and contrast between the pixels or regions of pixels of an image rather than relying only on the magnitudes of the intensities of the pixels in a binary image. Morphological techniques essentially make use of a structural element which is nothing but can be thought of as a kernel. Performing such operation on an image results in a new image in which certain pixels are set to a value only if a certain condition is satisfied, according to the operation. The most common morphological operations used in Image processing are dilation and erosion. The process of dilation involves the addition of pixels, i.e., the boundary pixels are extended if conditions are met, whereas the process of erosion,

like soil erosion, involves the shrinking of extra boundary pixels, indirectly leading to the removal of salt and pepper noise.

2. Proposed Work

Working on Hand Estimation

The idea proposed by Shefali et al.,^[1] is to identify distinct points for uniqueness in hand recognition. It is observed that 9 parameters (5 hand peaks and 4 valley points) are to be used for uniqueness. The distance between the 9 points and the reference point is observed as the parameter. (The reference point calculations are provided in the below section). The shape and geometry of the hand images are obtained in a common methodology and used to improve the identity verification of the person by combining the shape and features of hand geometry to obtain scores for the so formed fusion.

The proposed algorithm involves acquiring images, followed by pre-processing, finding the orientation of the hand gesture and also orientation registration followed by extracting the reference point extraction, determination of the finger feature points based on the calculation of the distances between the reference points and the feature points. This is done to obtain peaks and valleys, based on the local minima of the distance from reference point to the contour boundary points on the image. Hand shape features can be used in 2 methods: the first one being the mapping of the distances and the next one being the orientation registration.

Voronoi Image Segmentation Technique

The idea is to segment the image into Voronoi regions and perform clustering analysis on each Voronoi region to determine the unique characteristics of whole image.

Initially the given image is adjusted as per the requirement (resizing, rotating and grey scaling) and made ready for further operations. To divide the image into Voronoi segments we need an image property such as corner point that yields segments. For this corner points are obtained by detecting Harris Features using Harris-Stephens algorithm. The features can be obtained only if the input image is grey scaled. Many other image features such as HOG features, SURF (Speed Up Robust Features), Shi-Thomson corner detection algorithm can be used. But Harris corners are preferred over other because of the observed advantages.

Dilation and erosion processes are used complementary to each other as a sequence to remove noise in a image and the state of any given pixel in resultant image is obtained by a rule based approach to be applied on a pixel and its neighboring ones, which depends on the shape of image as well as the size of the structuring element kernel.

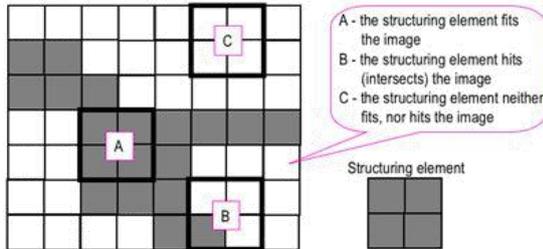


Fig. 2.1: Morphological operations using a structural element

Image Featuring

For getting a feature for an image we need to have corner points/seed points to give features. In our work, Harris Corner features are used, because the Harris corners resides its seed points majorly in the foreground areas. It failed in some case of complex imaging structures. Many features such as HOG (Histogram Oriented Gradient Features), SURF (Speed Up Robust Features), SIFT (Scale Invariant Feature Transform) have failed to work on complex background and foreground images.

Image Segmentation

The obtained corner points are used to draw polygonal segments. There are two orthodox methods available. Dirichletian Tessellation involves segmenting the image plane into a set of specified number of convex polygons subject to the condition that each polygon has exactly one corner point and that each one of the points in each polygon is nearer to its generating point than to any other. Delaunay Triangulation involves segmenting the image plane into convex polygons such that the circumcircle of the polygon does not have any seed point. Delaunay Triangulation is preferred over other models in the work due to higher observed advantages, such as non-overlapping of the feature point of the in each of their circumcircle.

Box Model Based Pose Estimation

The idea ^[3] is to segment the human body in to feature able box (A box will be drawn for elbow, arm, head, torso etc.). The box drawn can have some part of background, but it must cover the concentrated feature in the foreground. The motion parts are predicted by calculating the correlation of all the features to be examined. Even Correlation of HOG Features the features is planned for as future work. The parts/human parts with less movement will have high correlation factor due to consistency and parts in motion will have less correlation denoting featural changes in those portions. Thus, featural changes of the image are specified and are planned to be matched. But matching of the features in motion is not possible because the boxes drawn are rigid and dimensional and angular changes is not possible.

Optical Flow Tracking

Optical flow tracking is used to pattern the motion of objects in a real scenario that is occurred between observer and scene. We used Optical flow tracking algorithm to detect the featural changes in the images. Optical flow tracking can discern possibilities even in a dynamic environment. In Optical Flow Tracking, we used Horn-Schunk method to detect the image features and are represented in the form of RGB color changes. Horn-Schunk method optimizes the image on brightness constraint, and a term is used to regularize the flow field's smoothness. The changes in image are found by differentiating the image in vertical and horizontal directions with respect to time and the differenced equation is fed to color changes in RGB scale. The intensity of the color depends on the magnitude of change in image features.

3. Analysis And Experimental Investigation

Working on Hand Base Pose Estimation

The initial work is carried on pose estimation of hands (palms) and fingers. The objective is to identify unique characteristics from a person's hand. For this the image is converted into its grey scale and displayed. A 2-d median filter is used, which is a non-linear digital filter and used in removing salt and pepper noise. Removing or reducing noise is a pre-processing step to improve the image quality and reduce imperfections. Median filter is mostly used in Digital Signal Processing techniques because it is effective at eliminating noised portions, while keeping intact the edges in the image.

The image is subjected to thresholding. The thresholding values depends on the background and foreground conditions and the lighting conditions of the image. Find the image properties such as centroid and orientation of the image and rotate the image in the desired angle to have easy working for reference point. (Reference point is chosen such that its motion will be minimum when the hand posture is changed to account for linear and rotational invariance). Reference point will be usually chosen as the center of the bottom line of the hand image because it will have less effect compared to other parts when the hand is rotated or shifted. Extract reference point with the help of below algorithm. Better draw the contours of the image to have a check that the reference point will be on the contour of the hand. If the contours are not sharp use morphological transformation and obtain sharp contours and identify the size of image in-terms of pixel for reference point calculations.

Reference Point Extraction Algorithm

1. Input: Image (I) having centroid(u), row number (x), column number (y)
2. Output: Reference point (r)
3. Initial: $j \leftarrow 0$, $tmp \leftarrow []$
4. Initialize other variables
5. begin
6. $[s1, s2] \leftarrow \text{size}[I]$
7. for $I \leftarrow 1: s1$
8. if $I(I, yu) == 1$ (i.e. if the value of I (I, yu) is 1, where I is the binary Image)
9. $j \leftarrow j+1$
10. $tmp(j) \leftarrow i$
11. end
12. $xr \leftarrow \text{maximum}(tmp)$
13. $yr \leftarrow yu$
14. end
15. return (xr, yr)
16. end

The proposed system ^[1] follows the sequence of acquiring the images, applying pre-processing operations, finding the

orientation of the hand shape, registering the orientation of it, extracting the reference point of the shape. It is then followed by determining the finger feature points. These features are then used in 2 different ways: the first one being the method for registering the orientation of shape and the second one being the map of the calculated distances, which is obtained from the local maxima and minima of the distances between the pixels on the hand contour and the reference point.

A hand orientation registration is performed here to negate the effect of rotation of hand shape images as the captured shapes are not always straight to be processed directly. Before performing hand orientation registration, we must use the pre-processing step to generate a hand mask, which may be denoted as hm. The centroid of the hand shape region is extracted, followed by the determination of the hand and major axis length from the region properties of the hm measured using region props function of the MATLAB. Centre of mass technique is used to find the coordinates (xl, yl) of the regions of the image.

Areas for Improvement in This Model

When the hand shapes are too complex the feature points cannot be extracted easily.

- When the fingers are hidden by the other portions of hand some valleys and peaks will not exist, this becomes problem to derive more features

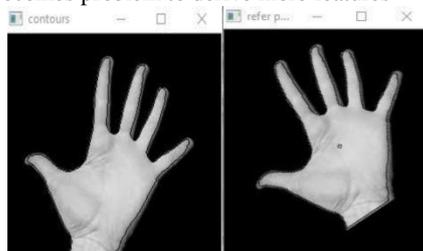


Fig. 3.1: The threshold image with contours

The rotated image of threshold image

- Valleys will be too tough to find because a normal hand will have a curvy valley where one point could not be determined as the valley point.
- The parameters such as thresholding will have the impact of light and positioning of hand in the image. When the background is complex the identification becomes feature, identification becomes very tough.

Other methods such as segmentation are preferred over feature extraction methods.

Voronoi Image Segmentation

The idea is to segment the image into Voronoi regions and perform clustering analysis on each Voronoi region to determine the unique characteristics of whole image.

Initially the given image is adjusted as per the requirement (resizing, rotating and grey scaling) and made ready for further operations. To divide the image into Voronoi segments we need an image property such as corner point that yields segments. For this corner points are obtained by detecting Harris Features using Harris-Stephens algorithm. The features can be obtained only if the input image is grey scaled. Many other image features such as HOG features, SURF (Speed Up Robust Features), Shi-Thomson corner detection algorithm can be used. But Harris corners are preferred over other because of the observed advantages.

Extract the features from the obtained corner points. The no. of corner points will be given as per the requirement. For segmenting the image based on features, segmenting methods such as Delaunay Triangulation and Dirichletian Tessellation methods can be used. We used Delaunay Triangulation where the segments are divided such that the circum circle of the Voronoi region does not

include another corner point. Dirichletian Tessellation segments the image into a set of convex polygons such that a polygon has only one feature as described in the above sections. Delaunay Triangulation is preferred over other segmentation techniques because of the transparency in choosing the segments in the image.

Each Voronoi region is pulled out individually and converted into RGB form. Clustering algorithms such as Fuzzy C Logic, K-means clustering, Hierarchical Clustering and Gaussian clustering algorithms can be used. The obtained clusters are plotted on a graph to obtain a expected features from the segmented image. Here successful clustering for obtained Voronoi regions was not as good as expected due to a multitude of reasons, explained in the following section.



Fig. 3.2: Voronoi segmentation on a flower and a pulled out Voronoi region

Areas for Improvement in This Model

As the image chosen is not constrained, segmentation cannot be perfect (Even Voronoi regions cannot be expected in some cases where foreground is considerably small). Voronoi regions cannot be scaled to require sizes while working with unconstrained videos and images which include motion. The existing segmenting algorithms are not suggestable when the foreground had too many features.

Clustering of individual Voronoi regions is a complex work to be done due to the presence of one desired feature in one or more Voronoi regions. Clustering algorithms such as Fuzzy C logic and K-means analysis will fail to work due to complexity in segmentation. Assume that a feature "edge" is present in or more Voronoi region, the extraction of the featured part will be not possible. When features such as SURF are used for extracting Voronoi regions, they yield different segments (Voronoi regions) when the lighting and blur in the image changes.



Fig. 3.3: Voronoi segmentation failed as the image is too complex

Box Model Based Pose Estimation

The idea is to segment the human body in to feature able box (A box will be drawn for elbow, arm, head, torso etc.). The box drawn can have some part of background, but it must cover the concentrated feature in the foreground. The motion parts are predicted by calculating the correlation of all the features to be examined. Even Correlation of HOF Features the features is planned for as future work. The parts/human parts with less movement will have high correlation factor due to consistency and parts in motion will have less correlation denoting featural

changes in those portions. By changing the position of box by sliding it over image the correlation will be maximized (maximally fit to the feature) and settled over there.

The pair of images chosen for this process must be mostly similar and must have some change in any feature that is to be concentrated. In our work, we had chosen the images such that the left elbow will have a movement. Box model is drawn manually for one of the image and is mapped onto it and for this we can check the correlation property of the feature and make sure that the correlation is 100%. Now map the same box model for the other image and calculate the correlation percentage for every feature and arrange the correlation in an order so that we can observe the motion in the image and identify as a flaw part.

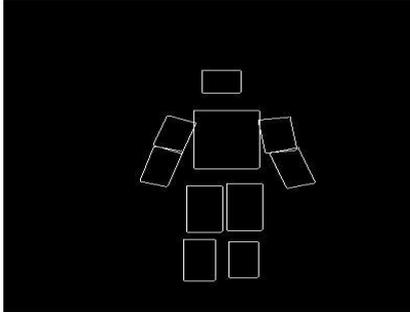


Fig. 3.4: Model based pose estimation showing various boxes for identification of different portions of body

A small difference can be observed between the two images in the left elbow. The correlation for this part is calculated as **0.9558**, whereas the correlation factor for head in the same images is recoded as **0.9983**. The Correlation of HOG Features for the same portions of the portions are recorded as [1, 0.9542; 0.9542, 1] for the left elbow and [1, 0.9886; 0.9886, 1] for the head part. Indicating that the HOG features of the two images match by 95.42% and 98.86% for left elbow and head respectively.

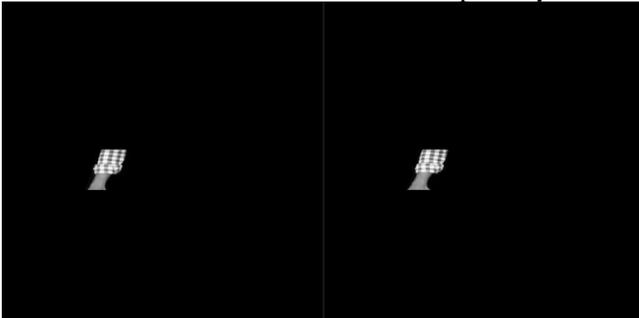


Fig. 3.5: Correlation measurement for left elbow

Speed Up Robust Features (Surf)

There are endless situations calling for image processing done by software. In general, they can be clustered into a few groups. Checking for presents object detection localization measurement as well as identification and verification. In most cases the acquired image from the cameras not directly processed within the application instead of his pre-processed to enhance the image calling to the specific talk. Examples of pre-processing on noise reduction as well as brightness and contrast enhanced.

Some of the steps can be done directly by the camera itself and thus save CPU load only hosts. To use the camera as a measuring device. It must be calibrated to the physical world. Camera calibration can have referred to 2 things. Geometric calibration and color calibrate. With geometric calibration, we correct the lens distortion. Furthermore, we can also determine the relationship between cameras natural units meaning pixels. And the real-world units like millimeters are inches forks. What color calibration we ensure an accurate reproduction of called us. The back to the pre-processing the battle the image quality and the results of the image processing within your application.

Now let's have a closer look at image processing. When it comes to locating pots usually matching is involved. This means looking for regions that are like all the same as a predefined template. This template can either be also an image. Or a geometric pattern which contains information regarding Acheson geometric features. These methods are colored correlation pattern matching and geometric pattern matching respect. Let's have a look at cookie inspection.

Yet template would be the image of a perfectly shaped cookie. The camera takes images of all cookies on the conveyor belt. As soon as there's a cookie that doesn't match the template. A cookie is rejected. The main use of measurement with the help of image processing is in alignment our inspection epic. Most measurement techniques rely on Texan items. And Acheson area and an image displaying a significant change in the image intensity. Or in other words high local contrast.

This means your software analyses the gray levels of the image and based on this identifies shapes measures distances and calculates the geometry. This measurement and calculation are made possible by the camera calibration. That established the relationship between pixels. And real-world units. Take the label off a bottle for example. With measurement, you can check if the label has been placed correctly. Typical applications for identifications are barcode into the matrix code reading.

Optical character recognition also called OCR. One way to manage optical character recognition is by separating the characters in the image and comparing them with a set of templates. Afterwards, the software can convert the captured data into editable and searchable data. A popular example for OCR is automatic number plate recognition also known as a and P. R. Each of these processing techniques covers a wide spectrum of machine vision applications. Combining them can give you even more possibilities.

Imagine a car is entering a parking garage. The camera takes an image of the license plate and the call. When the car is exiting the garage, the camera takes another image. Then the software compares those 2 images. The gate only opens when the license plate and the common law the same as in the first image. This was only a short overview of the information you can gain from your images by using image processing. But it offers a taste of the many different possibilities.

Optical Flow Tracking

Optical flow has its vast advantages in the field of Digital Image processing. The Optical flow theory first came into existence in 1940's by James. G. Gibson who is an American psychologist. It is a technique used to estimate the motion and motion vectors. In an Image sequence, we can also estimate the pixel motion and its dense pixel correspondence. Optical flow mainly used in the field of Object tracking. Object velocity and position of the object of next frame can also be estimated using Optical flow. It consists of two fields: Motion field and Optical flow field.

Let us consider two images of the same frame with times t and Δt at every position of the pixel to calculate and estimate the motion of object. Then we get differential methods which approximates Taylor-Series method. The brightness constancy for an image I with intensity $I(x, y, t)$ is given as:

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t)$$

where $\Delta x, \Delta y, \Delta t$ are the movements of two image frames.

Assume that the moment is very small, and applying Taylors method we get;

$$(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) + \frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t$$

The resultant equation from the above equation

$$\frac{\partial I}{\partial x} v_x + \frac{\partial I}{\partial y} v_y + \frac{\partial I}{\partial t} = 0$$

Where v_x and v_y are the velocity components of Optical flow
Optical flow mainly consists of three methods namely:

- Block matching
- Horn-Schunck algorithm
- Phase correlation method
- Lucas-Kanade algorithm

In this scenario, we used Horn-Schunck algorithm to find featural changes.

Horn-Schunck Algorithm

In this paper, we have used the most fundamental and widely used algorithm called HORN-SCHUNK ALGORITHM to estimate Optical flow in an image. It is a classical optical flow algorithm which was proposed by HORN and SCHUNK in the year 1981. It mainly solves the problems in the aperture of the image. The aperture problem is solved by using a global constraint called “smoothness vector”. It reduces the distortions in the image by using this smoothing vector.

This method provides the solution for minimization problem through newly designed framework. It is used to produce smooth vector fields by minimizing the optical flow constraint and reducing the magnitude of field variations. In this method smoothness is considered as an important condition to be satisfied. Equations involved are solved in an iterative manner. Horn Schunck method provides not only smoothness for the results but also provides global information, accurate derivatives of time and using of two frames at a time is possible. Many of the current methods in optical flow estimation encounters the problem with limited spatial resolution which can be reduced by switching to high resolution and high accurate Horn-Schunck optical flow.

Algorithm

1. Initially we need to consider two input frames frame1, frame2 and a variable i, j, λ .
2. Now build an image pyramid and initialize the flow to zero.
3. For i =image pyramid level:1, we need to initialize the optical flow from its previous level and build gradient and Laplacian matrix.
4. Now for j =maximum wrap num: 1, we need to calculate gradient of input image i.e., I_x, I_y, I_z .
5. We need to construct a linear system and then solve the system to estimate or compute the Horn Schunck flow.
6. Finally, we need to employ a median filter to smooth the output image.

The mentioned algorithm is strong enough to handle the input images which are quantized and can also eliminate additive noise if any. Below is the clear explanation of how this method is used with the help of equations.

Consider $I(X, Y, t)$ denote the input image intensity and X, Y denote the pixel coordinates and t indicates time. $I_2(X+dx, Y+dy, t+dt)$ represent the pixel intensity after some time, dt . Here dx, dy denotes the displacement of pixel coordinates in time dt . From the initial conditions of the method we need to consider the pixel intensity is not changed, so,

$$I_1(X, Y, t) = I_2(X+dx, Y+dy, t+dt)$$

Now we to apply first order Taylor expansion on I_2 equation. Here we need to consider some parameter such as velocity (U) = $dX/dt, V = dY/dt$.

Now we will be considering partial derivatives of image intensity as I_x, I_y, I_t . We will consider an intensity constancy which is a linearized version,

$$I_x + I_y + I_t = 0$$

The above equation is undetermined and is a well-known as aperture problem. To solve the aperture problems optical flow method, add a global constraint which impose penalty on global derivation. Below is the energy functional,

$$E_d(U, V) = \int (\Omega \varphi (I_x U + I_y V + I_t))^2 d\Omega$$

Here for the second time, we impose penalty on the total flow field variations. This can be expressed in the energy functional,

$$E_s(U, V) = \int \Omega (\varphi (U_x^2) + \varphi (U_y^2) + \varphi (V_x^2) + \varphi (V_y^2)) d\Omega$$

E_d, E_s denotes data term and smoothness term respectively. As the neighbouring pixels velocity varies smoothly all over the image except at boundaries denotes image spatial domain and φ denotes the penalty function. Total energy equation can be denoted by E ,

$$E(U, V) = E_d + \beta E_s$$

In the above energy equation, β represents weighted factor. By applying the minimization method on the energy equation $E(U, V)$ optical flow can be regained.

Previously this minimization of total energy $E(U, V)$ is carried out by wrapping based method. Here the linearization of numerical scheme is postponed. As performed by conventional Horn Schunck method, modified model also get finalized spatial derivatives by obtaining the result of spatial derivatives of two sequential images by their temporal average, but this modified method varies by considering the second image was to be on the wrapped image. This wrapped image is obtained from coarse to fine wrapping process.

This modified HS method involves the pyramidal image calculation of optical flow and a high performance cubic interpolation wrapping method, considered for large displacement. Here pyramidal images used are obtained by continues calculation of gaussian pyramidal reduction of an image. It is also a multi scale image.

Horn Schunck method is stable to noise and for this optical flow field of the input image must be smooth and continuous also velocity vector of adjacent pixels must be similar to considered pixel. For estimation of 3D HS algorithm, we regularize equation. To improve the results obtained spatial pre-smoothing can be suggested because horn schunck algorithm is based on sub pixel accuracy. This method is also applicable for synthetic images and is highly validated when compared with lucas-kanade and other optical flow methods. It performs better for motion boundary images. Here the maximum allowable displacement is six pixels i.e., twice the allowable displacement of lucas-kanade method. At near zero deformation regions in an input image, HS method exhibit high sensitivity to noise. This method has highest spatial resolution when compared with LS method and correlation method in 3D deformation.

4. Results and Discussions

Hand Pose Estimation

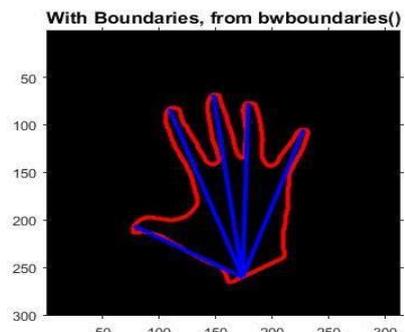


Fig. 4.1: The peaks in the hand shape

We have successfully used the hand orientation and shape-based features to extract the peaks of the hand after performing orientation registration so as to obtain shape features based on the distance between the reference point and the peaks and valleys.

Voronoi Segmentation

We used Delaunay Triangulation where the segments are divided such that the circumcircle of the Voronoi region does not include another corner point. Delaunay Triangulation is preferred over other segmentation techniques because of the transparency in choosing the segments in the image to obtain successful segmentation after which clustering was not particularly successive for complex images.

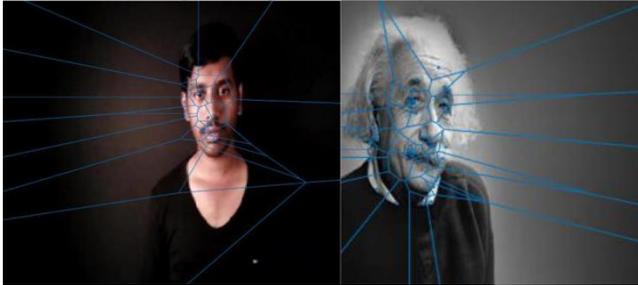


Fig. 4.2: Observe that the regions that are congested on foreground (face) denoting presence of useful information

Box Based Model Pose Estimation

The Correlation of HOG Features for the same portions of the portions are recorded as [1, 0.9886; 0.9886, 1] for the head part. Indicating that the HOG features of the two images match by 98.86% for head.

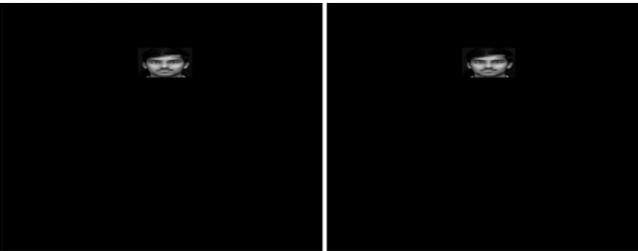
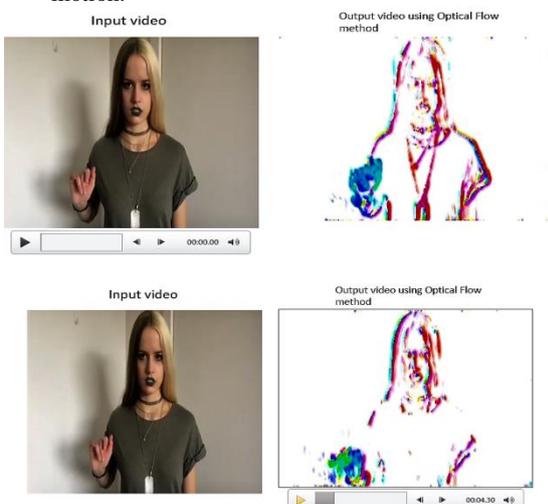


Fig. 4.3: Correlation measurement for head

Optical Flow Tracking

The main process implemented in the Horn-Schunk Algorithm are:

- 1) The first one is the estimation of partial derivatives.
- 2) The second one is a minimization of sum of the errors by an iterative process to present the final vector in motion.



5. Conclusion And Future Scope

We have explored three models for various parts of pose estimation viz., three methods i.e., hand shape and geometry based feature extraction, Voronoi region based image segmentation and model based deep structured pose estimation for primarily hand gesture recognition, finding regions of image where most of visual information cues are present, leading to increased density of Voronoi regions near the points of interest, and finally the deep-structure based model for finding the correlation between various parts of body for the motion. We have also identified areas of improvement where these methods are to be improved upon due to a multitude of reasons such as complex backgrounds, occlusion, camouflage effect, etc., and have implemented methods to minimize them.

Future scope of this work includes the extension of the deep-structure based models to the boxes fused with the motion of various parts of the body. This may be achieved using correlation-based measurements for the body parts in motion. Another interesting task that is to be considered is the application of the above model for the extracted HOG features. This method of feature extraction can be carried out by using a neural network-based approach to improve accuracy and reach leading results in the field. Especially since our applications focus on the use of pose estimations, it will be good to use a Convolutional Neural Network (CNN) for classifying the poses from videos, for which a large dataset of image data is to be collected.

References

- [1] Sharma S, Dubey SR, Singh SK, Saxena R & Singh RK, "Identity verification using shape and geometry of human hands", *Expert Systems with Applications*, Vol.42, No.2,(2015), pp.821-832.
- [2] Hettiarachchi R & Peters JF, "Voronoi region-based adaptive unsupervised color image segmentation", *Pattern Recognition*, Vol.65, (2017), pp.119-135.
- [3] Zhao L, Gao X, Tao D & Li X, "A deep structure for human pose estimation", *Signal Processing*, Vol.108, (2015), pp.36-45.
- [4] Chen Z, Qi Z, Meng F, Cui L & Shi Y, "Image segmentation via improving clustering algorithms with density and distance", *Procedia Computer Science*, Vol.55, (2015), pp.1015-1022.
- [5] Sapp B & Tasker B, "Multimodal Decomposable Methods for Human Pose Estimation", *IEEE Computer Vision and Pattern Recognition*, (2013).
- [6] Marin-Jimenez MJ, Munoz-Salinas R & Medina-Camicer R, "Mixing body parts model for 2D Human Pose Recognition", *IET Computer Vision*, (2017).
- [7] Eichner M, Marin-Jimenez M, Zisserman A & Ferrari V, "2D Articulated Human Pose Estimation and Retrieval in Unconstrained Still Images", *Computer Vision*, (2012).
- [8] Shi Q, Di H, Lu Y, Lv F & Tian X, "Video pose estimation with global motion cues", *Neurocomputing*, Vol.219, (2017), pp.269-279.
- [9] Kishore PVV, Kumar EK, Manjula B & Kumar PR, "Sign Video Segmentation using region, Boundary based active contours with shape priors", *Computer Science and Information Technology*, (2012).
- [10] Sedai S, Bennamoun M & Huynh DQ, "Discriminative fusion of shape and appearance features for human pose estimation", *Pattern Recognition, Elsevier*, (2013).
- [11] Toshev A & Szegegy C, "Human Pose Estimation via Deep Neural Networks", *IEEE Conference Computer Vision and Pattern Recognition*, (2014).
- [12] Pishchulin L, Andriluka M, Gehler P & Schiele B, "Strong appearance and expressive spatial models for human pose estimation", *Proceedings of the IEEE international conference on Computer Vision*, (2013), pp.3487-3494.
- [13] Kishore PVV, Kumar PR, Kumar EK & Kishore SRC, "Video audio interface for recognizing gestures of indian sign", *International Journal of Image Processing*, Vol.5, No.4,(2011).
- [14] Kishore PVV, Prasad MVD, Kumar DA & Sastry ASCS, "Optical flow hand tracking and active contour hand shape features for continuous sign language recognition with artificial neural

- networks”, *Proceedings of the 6th International Advanced Computing Conference, IACC*, (2016), pp.346-351.
- [15] G, Abikhanova, A Ahmetbekova, E Bayat, A Donbaeva, G Burkitbay (2018). International motifs and plots in the Kazakh epics in China (on the materials of the Kazakh epics in China), *Opción*, Año 33, No. 85. 20-43.
- [16] Lei Y, Jiang X, Shi Z, Chen D & Li Q, “Face recognition method based on SURF feature”, *International Symposium on Computer Network and Multimedia Technology*, (2009), pp.1-4.
- [17] Lucena MJ, Fuertes JM, Gomez JI, de la Blanca NP & Garrido A, “Tracking from optical flow”, *Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis*, (2003), pp.651-655).
- [18] Illeperuma GD & Sonnadara UJ, “An autonomous robot navigation system based on optical flow”, *6th Int. Conf. Ind. Inf. Syst*, (2011), pp.489-492.