Background Masking For Real-Time Video Calling

Umang Raval Kisore S

Abstract-Techniques such as background subtraction, optical flow and temporal differencing are used to detect moving objects in a video application. Background Subtraction is the most popular technique and it utilizes the mathematical model of static background and comparing it with every new frame of the video sequence. In this project, we use methods to extract objects from the video and put it in a desirable background in real time using Masking. We also would find out the most optimal method for deploying this technique to video calling.

Keywords - Background Subtraction, Masking

I. Introduction

THERE many situations in daily life that people had to move to a particular place to do a video call when it's really important to do one. But not always we find a place when it's an emergency. We use a video capturing device (camera/ webcam in this case) to acquire images for real-time analysis and processing. Histogram Equalization is used to enhance the video quality as the frames might have low contrast and resolution. Segmentation is done to identify the foreground and masking is done to replace the background with another background of our choice.

II. Background Subtraction

Background subtraction is the method used in computational to separate foreground objects from the background in the sequence of video frames. Non-recursive techniques such as Frame differencing, Median Filter, Linear predictive filter and Non-parametric model proposed by researchers as general is a technique that uses a sliding window approach for background estimation . This technique maintain a buffer of previous video frames and estimate a background model based on the statistical properties of these frames that causes this technique consume high memory.

[1]

Most literatures on background subtraction algorithm consider four major steps namely pre-processing, background modelling, foreground detection and finally is data validation. Firstly, pre-processing is the process of changing the input video sequences into a format that can be read for the next phase. In background modelling, the new video frames are updated and background model is calculated where it provides statistical description. Unidentified pixel in the video frame in background model will be output as binary candidate foreground mask at foreground detection step. Data validation stages function as examiner and eliminator where it examines candidate mask and eliminates pixels that are not related with target moving objects and only provide the foreground masks output. [3]

Foreground segmentation, is a basic problem. In formula, foreground segmentation takes as input an image I, which is assumed to be a composite of a foreground image F and a background image B. The color of the i'th pixel can be represented as a linear combination of the foreground and background colors, where α represents the opacity value:

$$I_i = \alpha_i F_i + (1 - \alpha_i)$$

But detecting motion through background subtraction is not always as easy as it may first appear. Indeed, some videos with poor signal-to-noise ratio caused by a low quality camera, compression artifacts or a noisy environment, are likely to generate numerous false positives. False positives can also be induced by illumination changes (gradual or sudden), an animated background (waves on the water, trees shaken by the wind), or camera jitter to name a few. On the other hand, false negatives can also occur when a moving object is made of colors similar to the ones in the background (the so-called camouflage effect). With such scenarios, a simple interframe difference with global threshold reveals itself as a weak solution. In order to cope with those challenges, numerous background models and distance measures bound up to different optimization schemes have been proposed in the past decade. [4]

Different approach utilized in recursive techniques which is approximated median filter, Kalman filter, Mixture of Gaussian (MoG) do not maintain a buffer for background estimation but updated a single background based on input frame. Recursive can be described as a process of repeating objects in a self-similar way. In addition of that, it maintains a single background model that is updated with each new video frame that makes this technique have minimal memory requirements as compare to non-recursive and computational efficient.[3]

III. Foreground Detection

Next, foreground is the moving object that separated from the background model or scene after the step of background separation. The method will detect moving object and classify the process of pixels as foreground and background. It detects the foreground and compares the input video frame with the background model, and identifies it pixels from the input frame. However, the detection rate of the foreground depends on the method chosen. Normal approach is to verify the input pixel and compare it with corresponding background estimation. Another popular scheme to detect the foreground is to use threshold based on the normalized statistics. However, limitation of this technique for outdoor environment leads to another approach where two thresholds with hysteresis or known as spatial variability is introduced . [3]

There are several more existing method that have been combined with new propose method such as MoG method with gradient-base background subtraction and morphological filter that depends on contour of moving objects . Initial Background model, shadow hi-lite detection and background adaption, Background subtraction using Density and Kernel estimation, suppress false detection, updating background for short and long term and shadow detection,
trackability features and interest of region , frame differencing, adaptive median filtering with pre and post processing among others techniques and combination. [3]

MoG method is chosen due to its low rate of complexity, memory consumption and suitability for outdoor environment along with its robustness and also it can handle multi-modal distributions. In MoG, the background is known as parametric frame of values where each pixel location is represented with number of Gaussian functions as probability distribution function . [3]

$$F\,(i_t = \mu) = \sum_{i=-1}^{k} \omega_{i,t} \,.\; \eta\,(\mu, \sigma)$$

where,
η = the i-th Gaussian component
μ = intensity mean
σ = standard deviation
$\omega_{i,t}$ = portion of the data accounted by i-th component

In addition, for MoG approach, only pixel that is within a scaling factor of background standard deviation is considered as part of background. This can be determined by comparing the pixel value with Gaussian component tracking. MoG approach is more suitable for outdoor environment due to its capability to separate and suppress non-stationary object that classify as noise such as moving leaves, sky and etc. However, illumination will be the drawbacks. Therefore, five main parameters will be evaluated and that will give significant impact.[3]

The five parameters are:
i. Ts = Background component weight threshold
ii. D = Standard deviation scaling factor
iii. ρ = Learning rate
iv. K = Total number of Gaussian components
v. M = Maximum number of components M in the background model

Furthermore, the advantage of MoG is that it can extend to colour video sequences that can solve the shadows effect . [3]

In contrast to the large amount of foreground segmentation publications, there are fewer studies on techniques for compositing the original foreground and a new background for background substitution. Since the light sources of the original video and the new background may be drastically different, directly copying the foreground to the new background will not achieve satisfactory results. Some seamless image composition techniques may seem relevant at first glance, but they require the original and new backgrounds to be similar. Other color correction techniques based on color constancy are more suitable in our context. Color constancy methods first estimate the light source color of the image, and then adjust pixel colors according to the specified hypothetical light source color. [8]

IV. Histogram Equalization

Histogram is defined as the statistic probability distribution of each gray level in a digital image . Histogram equalization (HE) is one of the well-known methods for enhancing the contrast of given images, making the result image have a uniform distribution of the gray levels . It flattens and stretches the dynamic range of the image’s histogram and results in overall contrast improvement. HE has been widely applied when the image needs enhancement, such as medical images enhancement. However, in consumer electronics such as TV, HE is rarely employed because it may significantly change the brightness of an input image and cause undesirable artifacts.[9]

In theory, the mean brightness of its output image is always the middle gray level regardless of the input mean, because the “desired” histogram is flat. This is not a desirable property in some applications where brightness preservation is necessary. Brightness preserving Bi-Histogram Equalization (BBHE) has been proposed to overcome that problem . BBHE first separates the input image’s histogram into two by its mean, and thus two non-overlapped ranges of the histogram are obtained.[9] Next, it equalizes the two sub-histograms independently. It has been analyzed that BBHE can preserve the original brightness to a certain extent when the input histogram has a quasi-symmetrical distribution around its mean. [9]

Later, equal area Dualistic Sub-Image Histogram Equalization (DSIHE) has been proposed, it claims that if the separating level of histogram is the median of the input image's brightness, it will yield the maximum entropy after two independent sub-equalizations . DSIHE will change the brightness to the middle level between the median level and the middle one of the input image. Nevertheless, neither BBHE nor DSIHE could preserve the mean brightness. Then Minimum Mean Brightness Error BiHistogram Equalization (MMBEBHE) is proposed to preserve the brightness "optimally" . MMBEBHE is to perform the separation based on the threshold level, which would yield minimum difference between input and output mean. This threshold level is essentially chosen by enumeration. Another scheme, named Recursive Mean-Separate Histogram Equalization (RMSHE), has been proposed to preserve the brightness . RMSHE uses the BBHE iteratively.

First RMSHE separates the input histogram into two pieces, by the mean. Then, to each piece, it uses this operation many times to generate 2n -pieces histograms. Finally, it equalizes each histogram piece independently. It is claimed theoretically that when the iteration level n grows larger, the output mean converges to the input mean, and thus yields good brightness preservation. Actually, when n grows to infinite, the output histogram is exactly the input histogram, and thus the input image will be output without any enhancement at all. In the consumer electronics such as TV, the preservation of brightness is highly demanded. The aforementioned algorithms (HE, BBHE, DSIHE, MMBEBHE and RMSHE) preserve the brightness to some extent, however, they do not meet that desirable property quite well. [9]

We focus our attention on the following transformation
$z = T(r), 0<=r<= 1$

that maps to a level z for every pixel value r in the original image. T(r) satisfies the following fundamental conditions:

T(r) is single-valued and monotonically increasing in the interval $0 <=r<= 1$;

$0 <=T(r) <=1$ for $0<=r<= 1$.

If we choose

$T(r)=\int p_r(w)dw$

it is not difficult to find that the PDF of the output gray level z follows a uniform distribution ranging from 0 to 1.

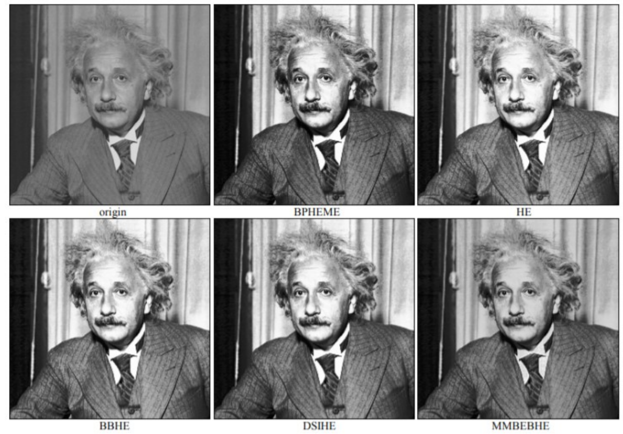


Figure 1 : Result of Different HE techniques

Or, we may say that the determination for the variable z at a given pixel will provide us the maximum information, because z equals to any gray level with equal probability, thus, it contains the most uncertainty. HE can be regarded as a special case of HS when we choose the target output histogram as a uniform one, $p_z(z)=1$, $0<=z<=1$. [9]

Figure 2: HE techniques applied on a different sample

V. Video Matting

Unlike interactive video matting methods, which need user interaction during video playback, automatic video matting is more appropriate for live video. The earliest kind of automatic video matting problem is constant color matting , which uses a constant backing color, often blue, so is usually called blue screen matting. Although excellent segmentation results can be achieved by blue screen matting, it needs extra equipment including a blue screen and careful setting of light sources. More recent video matting methods loosen the requirement for the background to have constant color, and only assume that the background can be pre-captured and remains static or only contains slight movements. They model the background using either generative methods, such as a Bayesian model , a self-organized map, a Gaussian mixture model, independent component analysis , a foreground–background mixture model, or non-parametric methods .[8]

Such models allow prediction of the probability of a pixel belonging to the background. These methods can create holes in the foreground and noise in the background if the colors of the foreground and background are similar, because they only make local decisions. Some recent techniques utilize the power of graph-cut to solve an optimization problem based on a conditional random field using color, contrast, and motion cues; they are able to create more complete foreground masks since they constrain the alpha matte to follow the original image gradient. Other work focuses on foreground segmentation for animation.[8]

In our case, in order to acquire real-time online matting for live video, it is inappropriate to include motion cues. Thus our model is only based on color and contrast. We also find that stronger shadow resistance can be achieved by employing a color line model. Another drawback of existing online methods is that they only acquire a binary foreground segmentation and then use approximate border refinement techniques such as feathering or border matting to compute fractional opacities along the boundary.[8]

Interactive video matting is another popular video matting approach. It no longer requires a known background and static camera, and takes a user drawn trimap or strokes to tell if a pixel belongs to the foreground/background/unknown region. For images, previous methods are often sampling-based , affinity-based , or a combination of both , computing alpha values for the unknown region based on the known region information. For video, use optical flow to propagate the trimap from one frame to another. VEach classifier subsequently solves a local binary segmentation problem, and classifiers of one frame are propagated to subsequent frames according to motion vectors estimated between frames. However, they need to take all frames all at once to computer reliable motion vectors, which takes a huge amount of time, so are unsuitable for online video matting.Near realtime performance is available with the help of a GPU, but they still need an input user trimap and an extra training stage, so are inconvenient for live video applications.

VI. Results Of application

The first stage of the project was to separate the foreground from the background. Multiple test values for the rectangle were used until the one with the least amount of inaccuracy was obtained.

Figure 3: Initial test results of foreground and background separation

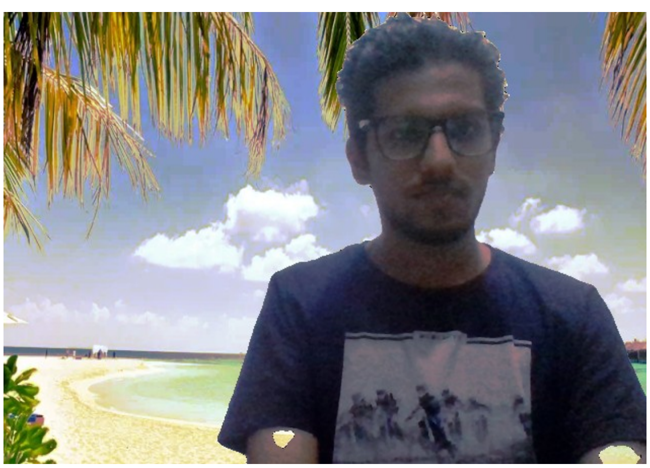
It was observed that the inconsistency in lighting in normal environments would be leading to inaccuracy while recognizing the foreground and separating the background. After that histogram equalization and video matting were applied to get the final output. It was also observed that when the user wears a dress similar to the background, the algorithm assumes most part of it background and we get an output only including the neck and face of the user. So, it is advised to make sure there is a minimum amount of contrast between the background and object colors during the process.

Backgroung Mask

VII. Conclusion

Foreground Mask

Final Combined Output



After Histogram Equaliization

In this work, we used processes such as Background Subtraction, Foreground Detection, Histogram Equalization and Video Matting to separate an object from the original background and place it in a desirable background during a real time video call. We tested multiple algorithms and came up with the least inaccurate method to do this process. Apart from hardware limitations, having a static background would be the main limtiation for this to be done. 80% accuracy was obtained as a result of application of above mentioned methods.

Further studies would include increasing the accuracy of the result and meanwhile keeping the process less dependant on powerful hardware.

References

[1] Sen-Ching S. Cheung and Chandrika Kamath, “Robust Techniques for Background Subtraction in Urban Traffic Video,” EURASIP Journal on Applied Signal Processing, vol. 2005, pp. 2330-2340, Jan. 2005.

[2] Piccardi, M.. (2004). Background Subtraction Techniques: A Review. IEEE Trans Syst Man Cybern. 4. 3099 - 3104 vol.4. 10.1109/ICSMC.2004.1400815.

[3] Shaik Mohamed, Shahrizat & Tahir, Noorita & Adnan, Rahmi. (2010). Background modelling and background subtraction performance for object detection. 1 - 6. 10.1109/CSPA.2010.5545291.

[4] Benezeth, Yannick & Jodoin, Pierre-Marc & Emile, Bruno & Laurent, Hélène & Rosenberger, Christophe. (2010). Comparative study of background subtraction algorithms. Journal of Electronic Imaging. 19. 033003-033003. 10.1117/1.3456695.

[5] Mohapatra, Badri & Panda, Prangya. (2017). Histogram equalization and noise removal process for enhancement of image. ACCENTS Transactions on Image Processing and Computer Vision. 3. 22-25. 10.19101/TIPCV.2017.39015.

[6] Tarika, Bindia & Kaur, Manmeet & Singh, Inderjeet. (2014). Review of Histogram Equalization Techniques. International Journal Of Engineering And Computer Science. 3. 9205-9210.

[7] Kaur, Ramandeep. (2015). Histogram Equalization Tool: Brightness Preservation and Contrast Enhancement using Segmentation with Opening-by-Reconstruction. International Journal of Computer Applications. 111. 11-23. 10.5120/19509-1126.

[8] Practical automatic background substitution for live video ,Haozhi Huang , Xiaonan Fang , Yufei Ye , Songhai Zhang and Paul L. Rosin

[9] Chao Wang and Zhongfu Ye, "Brightness preserving histogram equalization with maximum entropy: a variational perspective," in *IEEE Transactions on Consumer Electronics*, vol. 51, no. 4, pp. 1326-1334, Nov. 2005. doi: 10.1109/TCE.2005.1561863