IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 1, Issue 2, April-May, 2013
ISSN: 2320 - 8791
www.ijreat.org

VIDEO SURVEILLANCE FOR LIVE ACCESS USING ANDROID

D.KRIPA[1], N.BAVITHRA[2], B.SOUMIYA[3] and B.LALITHADEVI[4]

**Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College,
Avadi, Chennai – 62.**

ABSTRACT

This paper presents an enhancement in video surveillance technique. This provides high security. The video can be provided with high resolution by using Android smart phones. Here live videos are received in mobiles by the indication of sms.These live videos are accessed from any network.The required video is stored in e-mail for user's comfortability.

I. INTRODUCTION

Nowadays, visual surveillance system is becoming an essential part of our daily life. Since it is quite necessary to utilize such systems to ensure public security,almost every public facility has its own surveillance system. With the development of low-cost surveillance video cameras and high-speed computer network, it is technological feasible and economically affordable to provide such system for crime prevention and crime scene investigation. Visual surveillance is a long-standing field of computer vision, many efforts and tremendous progress has been made in this area. The CMU VSAM[1] system receives video inputs and collects and disseminate real-time information about the scene ,real-time visual surveillance system and the ADVISOR[3] system detect and track multiple people and monitoring their activities in outdoor scenes such as metro stations; Pfinder[4] also provides real-time performance in people tracking and behavior interpretation. Although there are several existing video based surveillance systems, all of them do not have the capability of acquiring clear human face images or sequences. Clear human faces can provide police more information about the criminals and help to track down them as soon as possible.Many face or body detection based method cannot achieve this because human faces will emerge multiple views and postures in different surveillance scenes. Moreover, environment illumination variation is another main factor which results these methods to fail. To solve this problem we propose a novel and robust visual surveillance system framework that exploits two main components: human head detection and object tracking. Human detector based on histogram of gradients (HoG) feature[5]. Simultaneously, motion and appearance information are extracted from the video sequence. Based Bayesian theory, we use two likelihood to evaluate the probability of a detected region represents an actual human head. The false positives are eliminated and the true positives are tracked by the EMD tracking algorithm with SURF points. The head detection and tracking results can be further utilized to control the camera PTZ to capture proper and high-resolution snapshots. The rest of the paper is structured as follows. After summarizing related work in Section 2, we demonstrate our system architecture in Section3.System components and implementation details are described in Section 4. Furthermore, experimental results and discussions are given in Section 5. Finally, we conclude the paper.

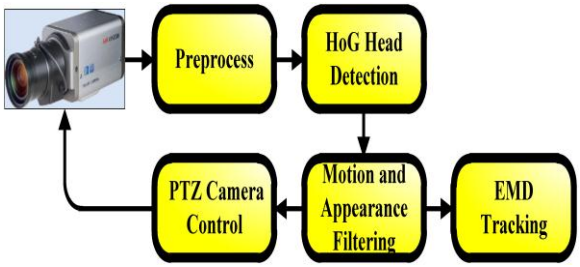
II. RELATED WORK

A.Humanface and body DetectionPedestrian detection has long been a hot research area in computer vision and has reached an impressive level[5], [6]. Many available approaches have already shown well performance on individual images. These existing methods generally consist of two steps: feature extraction and training. Viola et al[6] proposed a pedestrian detection framework that uses haar-like features inherited from[7] and then applies AdaBoost learning algorithms to get satisfactory classification results. These methods can only be applied in detecting human face which limits their popularity in actual surveillance systems. Another well-known pedestrian detector developed by Dalal et al[5] calculates a set of overlapping histogram of

gradients (HoG) descriptors fed into a support vector machine to achieve
good detection results. In this paper, we will adopt the latter approach for it is easy to be extended to head detection and robust to the various gestures of human heads. Computing HoG feature descriptor is time-consuming and also limits its application in real-time systems.
IEEE TRANSACTIONS ON IMAGE PROCESSING VOL. 34, NO. 10, JUNE 2012
2833 B. Object TrackingIn video tracking, mean shift[8], [9] and particle filter[10], [11], [12] are two main tracking methods categories. These methods utilize objects' appearance features such as color, gradient, optical flow and texture, but are not robust because they only consider the bin-to-bin comparison. Instead of using illumination-variant features, or applying transforms to makefeaturesilluminationinvariant,Zhaoetal[13]approaches illumination changes using the EMD[14] as a similarity measure to match color distributions. Their method uses color signature as the matching feature. But since it needs clustering algorithm to generate the feature in each iteration step repeatedly, it wastes the computation resource. Moreover, the color signature feature itself is also not robust to the applications in real surveillance scenes because this feature lacks position and orientation information.

III. SYSTEM ARCHITECTURE

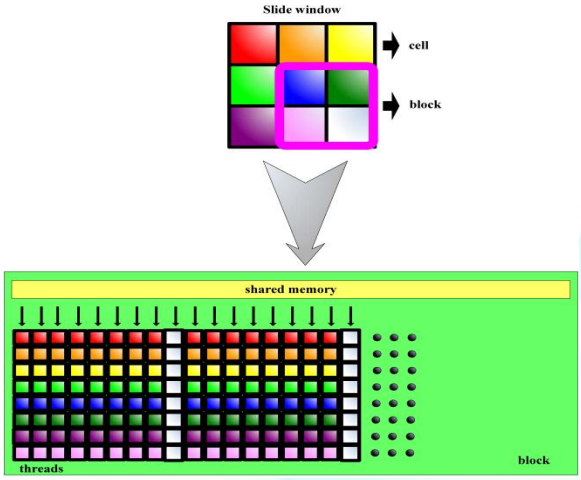
Fig. 1 depicts the system architecture. First, video signals are captured by the camera and provided to the preprocessing module(e.g. gamma correction, color balancing). Then the HoG algorithm is applied in the HoG Head Detection module to get initial head detections. In order to lower the false detection rate, we adopt motion and appearance filtering to eliminate static false positives. And then the improved head detection results is fed into the EMD Tracking module to get the actual physical position of the detected human heads.Simultaneously, based on the detection results, the system proactively control the PTZ Camera to get high-resolution human head snapshots Fig. 1. System architecture diagram.



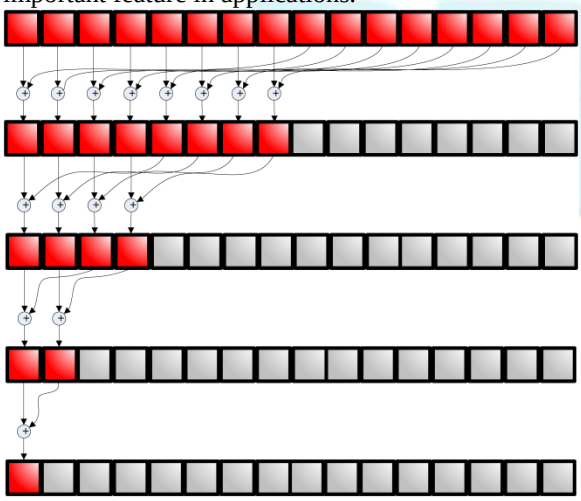
IV. ALGORITHM SPECIFICATION

A. Human Head DetectionSince the methods of human face detection cannot adapt multiple views and postures as well as illumination variation, in this paper we apply the HoG feature descriptor proposed by Dalal et al[5]. This feature describe local region's edge or gradient information in images so it is insensitive to local geometry and illumination variation.In implementation of HoG, we take the configuration as follows: input image in RGB color space (without any gamma correction);image gradient computed by applying [-1,0,1] filter along x- and y- axis without smoothing; linear gradient voting into 9 orientation bins in 0_-180_; 16_16 pixel blocks containing 2 _ 2 cells of 8 _ 8 pixel; blocks spaced with a
stride of 8 pixels; 24 _ 24 detection window and linear SVM classifier. After voting, the histogram is normalized and finally form a high dimension vector in feature space. It will take much time and computation resource to extract the HoG descriptor, so we propose a GPU-based algorithm for the purpose of real-time processing. Much time is consumed in gradient computation. When extracting HoG features, a sliding window is applied and traverse on the frames. For each
position and scale, the extracting operation is independent, which has excellent parallel property. We use CUDA for GPUbased algorithm. Since one block contains 4 cells, and the size of one cell is 8_8, then one histogram is corresponding to gradients of 256 pixels. 4 blocks will be processed in a sliding window, but the 4 blocks are overlapped with each other, so totally 9 blocks, 576 pixels need to compute gradient.
A CUDA block has 512 threads at most, so we let each CUDA block process one detection window. Especially, in each block we use 8 threads to process one extra pixel which solve the less-threads-more-pixels problem. The acceleration strategy is shown in Fig.2. When all threads finish computing gradient,

Fig. 2. The diagram of GPU based HoG feature extracting algorithm.



Different color represents different cells in the sliding window. In the CUDA blocks, different color demonstrates the threads used for processing corresponding pixels in the cells and the white boxes represent the extra pixel. we save the results in the shared memory of device and vote for generating the histogram with parallel reduction technique and normalize them, which is shown in Fig.3. In real applications, it will result to high false positive rate if we just use HoG feature. The reason is that the shapes of a certain categories of objects are similar to human heads'. To improve the detection result, we propose to include two extra information to filter the false positives. Motion and appearance are two important feature in applications.



Most part of human head is hair or skin, which means human head have specific appearance feature. Even human body is static, their head may be motorial while most of head-like objects are static, which means motion information is another good filter. We use Bayesian posterior to represent the probability of a detected region belongs to actual human head regions: p(!j jx) / p(xj!j)p(!j) j = 1; 2 (1) in which x means the detected region and !1 and !2 mean human head class and non-human head class respectively. 2834 Fig. 3. The diagram of parallel reduction algorithm. At the first level, ith element is added with i+offset element; at the second level, ith element is added with i+offset/2lth element, in which l represents the level.After finishing each level's computation, all the related threads should be synchronized. The process is repeated until there is only one element. p(!j) is prior, and we set p(!1) = p(!2) for simplicity. p(xj!j) is the likelihood and formulated as follows: p(xj!1) = pa(xj!1)pm(xj!1) (2) in which pm(xj!1) is defined as:pm(xj!1) =P xi2M T _(xi)jj(3)in which M means the mask generated from time differential motion detection[15]. means current head candidate region, so jj is the area of current head candidate region. _(xi) is Kronecker delta function. The whole likelihood means the ratio of two areas between motion region and head candidate region.Meanwhile,pa(xj!1) can be computed as below:pa(xj!1) = _1N(xj_1;_1) + _2N(xj_2;_2) (4)in which N(_) represents a Gaussian distribution, _ as mean and _ as covariance matrix. _1 and _2 are weight coefficients and _1 + _2 = 1. All the parameters can be estimated by expectation maximization algorithm[16]. Finally, we have the probability to evaluate a candidate region:p_(xj!1) = minf 1 _p(xj!1); 1g (5) The factor _ is designed for preventing a low multiplication between two high confidence. If the final confidence is greater than 0.5, then the candidate region is considered a real human head region.B. Object Tracking After detecting human head response and control the camera PTZ to capture the object, we will track it continuously until the object leaves the screen. We propose a robust tracking method based on EMD and SURF feature. Compared with SIFT[17], SURF has almost the same matching accuracy and rotation-scale invariance. Otherwise, SURF has more advantages in real applications: 1) the speed of computing SURF feature is much faster than that of SIFT because of the integral image method; 2) it can catch the points which have high contrast in different scales. Typical

SURF point detection result is shown in Fig.4. More detailed description about SURF can be found in [18]. (a) (b) Fig. 4. SURF feature. a) Original image. b) The extracting result of SURF feature points at different location and scale. One can see all the prominent regions (stamen or pistil) of the sunflowers have been detected. We formulate the tracking problem as getting the minimum EMD given a set of costs and flows, which is equivalent to solving a linear programming problem. The goal is to find the location that corresponds to the smallest EMD:C(fij) =Xmi=1Xnj=1d(pi; qj)fij (6)in which d(pi; qj) is the Euclidean distance between ith SURFpoint in the object's template and jth SURF feature point in the candidate region. fij is the flow between the two class of feature points and subject to: fij _ 0; 1 _ i _ m; 1 _ j _ n (7) Xn j=1 fij = wip 1 _ i _ m (8) Xm i=1 fij = wjq 1 _ j _ n (9) Xm i=1 Xn j=1 fij = 1 (10) in which wip is the response weight of ith SURF feature point in the object's template and wjq is the response weight of jth SURF point in the candidate region. In implementation, these weights are generated from choosing k feature points which have highest response values, normalizing these responses and sorting them in decreasing order. Simplex method can be applied to solve this problem. Before using EMD to match patch, we first use KLT optical flow[19] to get a coarse location of the object. Instead of tracking each point according to gradient in x- and ydirection, we tracking each point by the SURF responses. Based on the optical flow result, we move the track window onto the location and compute EMD measurement. If the 2835 computed EMD is less than current EMD, then we take the location as current tracked location and iterate this process until the computed EMD is larger than current EMD. When tracking multiple objects, once occlusion is detected, we apply Bayesian framework to identify the subordinations of these extracting feature points. Our proposed method can be fast and efficient because it
only needs one process of extracting SURF points for every frame while the color signature feature will be computed at every iteration step for one frame. Moreover, SURF feature is more robust than color so

and appearance filters, which will show the importance of introducing motion and appearance likelihood filters. The experiment result can prove that this step is necessary to real surveillance applications. Comparison results are shown in Fig. 5. Finally, we show some correctly Fig. 5. The our method have better performance than the one based on color signature.

IV. EXPERIMENT RESULTS

Extensive and comparative experiments are carried out and reported in this section. We first show human head detection results and the detection performance between the methods with and without motion and appearance filter as well as CPU/GPU comparison. Then we show the EMD and SURF tracker's performance and compare it with the method proposed by Zhao et al[13]. All these experiments are performed in following hardware environment: Core2 Duo 2.83GHz CPU, 4G RAM memory and GeForce GTX 460.A. Human Head Detection Firstly, we compare the speedup between our GPU algorithm and original CPU HoG implementation. We choose 4 sequences of different resolution videos, test the average runtime.Secondly, we test our algorithm in different surveillance videos, and calculate the precision and recall, which are shown in Table. 2. TABLE II
THE PERFORMANCE OF PROPOSED ALGORITHM. IN THIS TABLE, TF MEANS THE NUMBER OF TOTAL FRAMES. TP, FP AND FN REPRESENT TRUE POSITIVE, FALSE POSITIVE AND FALSE NEGATIVE, RESPECTIVELY THE PRECISION IS COMPUTED AS TP=(TP + FP) AND RECALLTP=(TP + FN).TF TP FP FN Precision(%) Recall(%)
Video1 1520 291 60 29 82.9 90.9
Video2 3936 2912 69 324 97.7 90.0
Video3 480 214 19 22 91.8 90.7
From Table. 2 we can see in most situations the precision and recall are all above 90% except video sequence 1. Video 1 has some head-shaped objects and these objects are not totally static, so the false positive is relatively high. The false negatives are produced because sometimes the boundary pixels of foreground objects are similar with background which results to the computed gradients are nearly equal to zero. That means the edge information is missing. Thirdly, we compare the two HoG human head algorithms with and without motion comparison of HoG algorithms with and without motion and appearance filtering. Some head-like objects can be eliminated from detected results generated solely by HoG algorithm. Detected results and their corresponding edge gradient information

in Fig. 6. HoG detection results. First row shows original image patch and second row shows corresponding gradient features. One can see the more prominent the difference between foreground and background, the more abundant the gradient information is.

B. Object Tracking In this part we show the robust performance of our proposedtracking algorithm. finally mistakes the background subregion as the object.Other experiments and tests show that the proposed algorithm is available in all kinds of scenes from real surveillance 2836 Fig. 7. Frames 1, 25, 50, 75 are from one of the PETS sequences. a) The color signature based method has lost track of the man since low resolution and the clustering algorithm reduces the information used by tracking. b)Our tracking algorithm maintains a robust focus on the object throughout the sequence with the support of SURF points. systems. Tracking results shown in Fig. 8 demonstrate that it is robust and general on these surveillance sequences.

Meanwhile, we compare it with the color signature feature proposed by Zhao et al[13].In Fig. 7, we show the tracking result on one sequence comes from PETS video data. Since the color of the tracked human's head, body and legs is similar with the background, color signature based method cannot catch the object's motion

VI. CONCLUSION

For retrieving clear human head or face image from real surveillance systems, we proposed a novel approach to this problem. To the best of our knowledge, this is the first work using HoG, the SURF feature and EMD together in visual tracking of surveillance system. HoG feature combined with motion and appearance information can reduce false positives significantly. SURF and EMD based tracking has robust tracking performance. GPU acceleration design and coarseto-fine tracking strategy makes real-time processing of video streams possible.