International Journal of Mechanical Engineering (IJME)
ISSN 2319-2240
Vol. 2, Issue 3, July 2013, 45-54

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

SECURE AUTHENTICATION SYSTEM USING VIDEO SURVEILLANCE

S. TAMILSELVAN & G. NITHYA

Arulmigu Meenakshi Amman College of Engineering, Namandi, Thiruvannamalai, Tamil Nadu, India

ABSTRACT

A Biometric person recognition for secure access to restricted data/services using PC with internet connection. To study, an application PC to be used as a biometric capturing device that captures the video and recognition can be performed during a standard web session. The main contribution of this novel proposal is, making comparison of portrait. Centroid context algorithm is used for selecting an random movements from an video and stored it in a database and make them to compare with video. To better characterize a portrait in a sequence, triangulate it into triangular meshes, which we extract the features: skeleton feature and centroid feature. Skeleton feature and centroid context feature working together makes human movement analysis a very efficient and accurate process. Depth first search(DFS) scheme is used to extract the skeletal feature of a portrait from triangulation result, from skeletal feature result, centroid context feature is extracted, which is a finer representation that can characterize the shape of a whole movements. For efficient and accurate process, generate a set of key portrait from a movement sequence. The ordered key portrait sequence is represented by string. For arbitrary matching action, string matching algorithm is used for implementing the concept.

KEYWORDS: Human Movement Analysis, String Matching, Triangulation

INTRODUCTION

In an era of information technology, mobile phones and PC are more and more widely used worldwide, not only for basic communications, but also as a tool to deal with personal affairs and process information acquired anywhere at any time. These scenarios, however, require extremely high security level for personal information and privacy protection through individual identification against un-authorized use in case of theft or fraudulent use in a networked society. Currently, the most adopted method is the verification of Personal Identification Number (PIN), which is problematic and might not be secured enough to meet this requirement. As is illustrated in a survey (Clarke & Furnell, 2005), many mobile phone users consider the PIN to be inconvenient as a password that is complicated enough and easily forgotten and very few users change their PIN regularly for higher security As a result, it is preferred to apply biometrics for the security of mobile phones and improve reliability of wireless services.

And also, Authentication is an area which has grown over the last decades, and will continue to grow in the future. It is used in many places today and being authenticated has become a daily habit for most people. Examples of this are PIN code to your banking card, password to get access to a computer and passport used at border control. We identify friends and family by their face, voice, how they walk, etc. As we realize there are different ways in which a user can be authenticated, but all these methods can be categorized into one of three classes. The first is something you know (e.g., a password), the second is something you have (e.g., a token) and the third is something you are (e.g., a biometric property).

The organization of this paper is as follows, Section II explains about the Video Surveillance process. Section III reviews some existing biometrics methods. Section IV presents the proposed Video Surveillance System for secure authentication. Section V discusses the experimental result.

VIDEO SURVEILLANCE

In this context of video surveillance, most of the emphasis is devoted to techniques capable of execution in real time on standard computing platforms and with low-cost off-the-shelf cameras. One important goal of these systems is *human-behavior analysis and object matching* by, making comparison of portrait. Centroid context algorithm is used for selecting an random movements from an video and stored it in a database and make them to compare with video. First, we apply background subtraction to extract body portrait from video sequences and then derive their boundaries by contour tracing. Next, a triangulation technique is used to divide a portrait into different triangular meshes. From the triangulation result, which we extract, the important features: skeleton feature and centroid feature. Skeleton feature and centroid context feature working together makes human movement analysis a very efficient and accurate process.

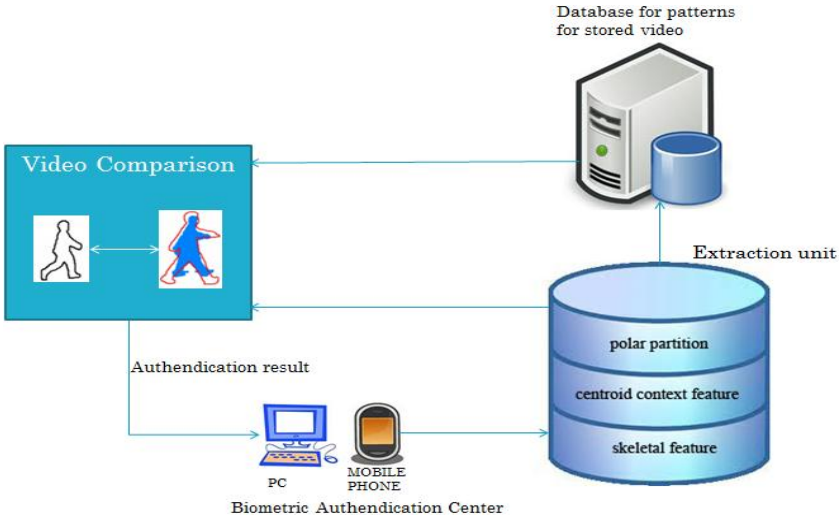


Figure 1: Architecture Diagram of Video Surveillance

The skeleton feature is used for coarse search, while the Centroid context, being a finer feature, is used to classify posture more accurately. To extract the skeleton feature, a graph search method is used that builds a spanning tree from the triangulation result. The spanning tree corresponds to the skeletal structure of the analyzed body posture. The proposed skeleton extraction method is much simpler than silhouette-based techniques. In addition to the skeletal structure, the tree provides important information for segmenting a portrait into different body parts. Based on the result of body movement segmentation, a new posture descriptor is extracted, namely, the centroid context descriptor, which is a finer representation that can characterize the shape of a whole movements. This descriptor utilizes a polar labeling scheme to label every triangular mesh with a unique number. Then, for each body part, a feature vector, i.e., the centroid context is constructed by recording all related features of the centroid of each triangular mesh according to this unique number. Then compare the different portrait more accurately by measuring the distance between their Centroid contexts.

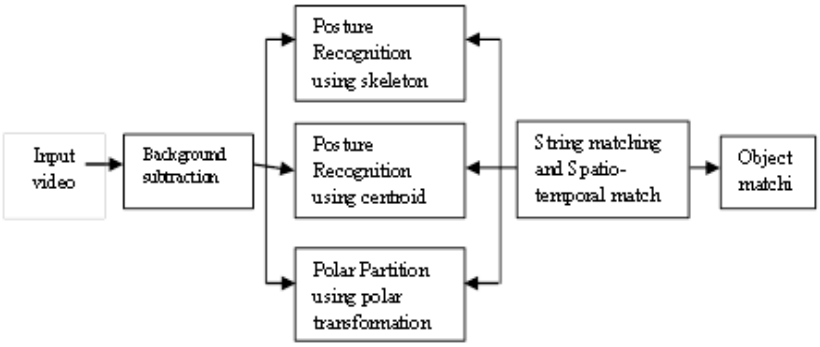


Figure 2: Flowchart of Video Surveillance

RELATED WORK AND CONTRIBUTIONS

We now briefly introduce some of the popular biometric modalities.

Face

Humans have a remarkable ability to recognize fellow beings based on facial appearance. So, face is a natural human trait for automated biometric recognition. Face recognition systems typically utilize the spatial relationship among the locations of facial features such as eyes, nose, lips, chin, and the global appearance of a face [9]. Recognition is affected by changes in Lighting, The person's hair, Age and also, If the person wear glasses.

Fingerprint

Fingerprint-based recognition has been the longest serving, most successful and popular method for person identification. Fingerprints consist of a regular texture pattern composed of ridges and valleys[9]. These ridges are characterized by several landmark points, known as minutiae, which are mostly in the form of ridge endings and ridge bifurcations. The drawbacks of fingerprint are it can make mistakes with the dryness or dirty of the finger's skin, as well as with the age (is not appropriate with children, because the size of their fingerprint changes quickly). For some people it is very intrusive, because is still related to criminal identification.

Iris

The iris is the colored annular ring that surrounds the pupil. Iris images acquired under infrared illumination consist of complex texture pattern with numerous individual attributes, iris comparison is done based on unique features e.g. stripes, pits, and furrows, which allow for highly reliable personal identification [9]. The iris is a protected internal organ whose texture is stable and distinctive, even among identical twins (similar to fingerprints), and extremely difficult to surgically spoof. However, relatively high sensor cost, along with relatively large failure to enroll (FTE) rate reported in some studies, and lack of legacy iris databases may limit its usage in some large-scale government applications.

Hand Geometry

It is claimed that individuals can be discriminated based on the shape of their hands. Person identification using hand geometry utilizes low resolution (~20 ppi) hand images to extract a number of geometrical features such as finger length, width, thickness, perimeter, and finger area[9]. The hand geometry systems have large physical size, so they cannot be easily embedded in existing security systems. It is very expensive and for some people it is very intrusive, because is still related to criminal identification.

Voice

Speech or voice-based recognition systems identify a person based on their spoken words. The generation of human voice involves a combination of behavioral and physiological features[9]. The physiological component of voice generation depends on the shape and size of vocal tracts, lips, nasal cavities, and mouth. The movement of lips, jaws, tongue, velum, and larynx constitute the behavioral component of voice which can vary over time due to person's age and medical condition (e.g., common cold). The main drawback of voice recognition is, an illness such as a cold can change a person's voice, making absolute identification difficult or impossible and also it provides low accuracy.

Signature

Signature is a behavioral biometric modality that is used in daily business transactions (e.g., credit card purchase). Dynamic signatures help in acquiring the shape, speed, acceleration, pen pressure, order and speed of strokes, during the

actual act of signing[9]. As a result, individuals who do not sign their names in a consistent manner may have difficulty enrolling and verifying in signature verification. Error rate: 1 in 50.

DNA

The DNA is an acronym for deoxyribonucleic acid which is present in nucleus of every cell in human body and therefore a highly stable biometric identifier that represents physiological characteristic. The DNA structure of every human is unique, except from identical twins, and is composed of genes that determine physical characteristics (like eye or hair color). Human DNA samples can be acquired from a wide variety of sources; from hair, finger nails, saliva and blood samples[9]. The DNA matching is quite popular for forensic and law enforcement applications. However, it requires tangible samples and cannot yet be done in real time. DNA matching process is expensive, time consuming and therefore not yet suitable for large scale biometrics applications for civilian usage.

The recognition accuracy of individual biometric traits outlined above may not be adequate to meet the requirements of some high security applications. The low individuality or uniqueness and lack of adequate quality of individual biometric traits for some users in the target population can also pose problems in large scale applications. This has led to an increased interest in using video for the task of biometric recognition. Not only does video provide more information, but also is more suitable for recognizing humans in general surveillance scenarios. Other than the multitude of still frames, video makes it possible to characterize biometrics based on inherent dynamics like gait which is not possible with still images.

SECURE AUTHENTICATION SYSTEM USING VIDEO SURVEILLANCE

File Formats

Most videos files have at least two types of file formats. First there is the container, and then the codec which is used inside the container. The container is what describes the whole structure of the file, and specifies which codec's are being used.

The following is a list of some of the more common types of container formats:

- AVI FORMAT(.avi)
- MPEG FILE FORMAT
- WMV FILE FORMAT
- MP4 FORMAT (.mp4)
- QUICKTIME FORMAT (.mov)
- Mpg FORMAT (.mpg)
- 3GP FILE EXTENSION (.3gp)

Extraction Unit

Background Subtraction

It uses a model of background variation that is a bimodal distribution constructed from order statistics of background values during a training period, obtaining robust background model even if there are moving foreground objects in the field of view, such as walking people, moving cars, etc. It uses a two stage method based on excluding moving pixels from background model computation[8]. In the first stage, a pixel wise median filter over time is applied to

several seconds of video (typically 20-40 seconds) to distinguish moving pixels from stationary pixels (however, our experiments showed that 100 frames ≈ 3.3 seconds are typically enough for the training period, if not too many moving objects are present). In the second stage, only those stationary pixels are processed to construct the initial background model

Shadow Detection

Let $B(i, j)$ be the background image formed by temporal median filtering, and $I(i, j)$ be an image of the video sequence. For each pixel (i, j) belonging to the foreground, consider a $(2N + 1) \times (2N + 1)$ template T_{ij} such that $T_{ij}(n, m) = I(i + n, j + m)$, for $-N \leq n \leq N$, $-N \leq m \leq N$ (i.e. T_{ij} corresponds to a neighborhood of pixel (i, j)). Then, the NCC between template T_{ij} and image B at pixel (i, j) is given by:

$$NCC(i,j) = \frac{ER(i,j)}{E_B(i,j)E_{T_{ij}}}, \tag{4}$$

where

$$\begin{aligned} ER(i,j) &= \sum_{n=-N}^{N}\sum_{m=-N}^{N} B(i+n, j+m)T_{ij}(n,m), \\ E_B(i,j) &= \sqrt{\sum_{n=-N}^{N}\sum_{m=-N}^{N} B(i+n, j+m)^2}, \text{ and} \\ E_{T_{ij}} &= \sqrt{\sum_{n=-N}^{N}\sum_{m=-N}^{N} T_{ij}(n,m)^2}. \end{aligned} \tag{5}$$

For a pixel (i, j) in a shadowed region, the NCC in a neigh- boring region T_{ij} should be large (close to one), and the energy $E_{T_{ij}}$ of this region should be lower than the energy $E_B(i, j)$ of the corresponding region in the background image. Thus, a pixel (i, j) is pre-classified as shadow if:

$$NCC(i, j) \geq L_{\text{ncc}} \text{ and } E_{T_{ij}} < E_B(i, j), \tag{6}$$

Where L_{ncc} is a fixed threshold. If L_{ncc} is low, several foreground pixels corresponding to moving objects may be misclassified as shadows. On the other hand, selecting a larger value for L_{ncc} results in less false positives, but pixels related to actual shadows may not be detected. In fact, the influence of the threshold L_{ncc} for shadow detection can be observed in Figure(3). This Figure illustrates the application of our shadow detector in the foreground image of Figure 3(c) using $N = 4$, for different thresholds L_{ncc}. Black pixels are foreground pixels, and gray pixels correspond to shadowed pixels according to Equation (6). Our experiments indicated that choosing $L_{\text{ncc}} = 0.95$ results in a good compromise between false positives and false negatives, and that $N = 4$ is a good neighborhood size.

Shadow Refinement

The NCC provides a good initial estimate about the location of shadowed pixels, by detecting pixels for which the surrounding neighborhood is approximately scaled with respect to the reference background. However, some background pixels related to valid moving objects may be wrongly classified as shadow pixels. To remove such false positives, a refinement stage is applied to all pixels that satisfy Equation (6).

$$\text{std}_R\left(\frac{I(i,j)}{B(i,j)}\right) < L_{\text{std}} \text{ and } L_{\text{low}} \leq \left(\frac{I(i,j)}{B(i,j)}\right) < 1, \tag{7}$$

The proposed refinement stage consists of verifying if the ratio $I(i, j)/B(i, j)$ in a neighborhood around each shadow pixel candidate is approximately constant, by computing the standard deviation of $I(i, j)/B(i, j)$ within this neighborhood. More specifically, we consider a region R with $(2M + 1) \times (2M + 1)$ pixels (we used $M = 1$ in all experiments) centered at each shadow pixel candidate (i, j), and classify it as a shadow pixel if:

Experimentally obtained values were $L_{\text{std}} = 0.05$ and $L_{\text{low}} = 0.5$ (however, we believe that further studies on the selection of L_{std} and L_{low} are needed).

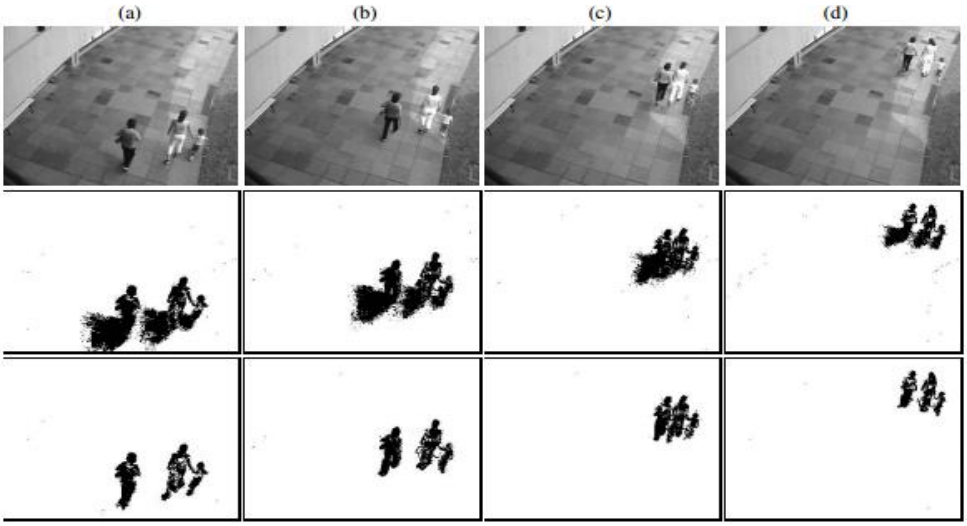


Figure 3: Top Row: Frames of Video Sequence. Second Row: Detected Foreground Objects Third Row: Foreground Objects with Shadow Removal

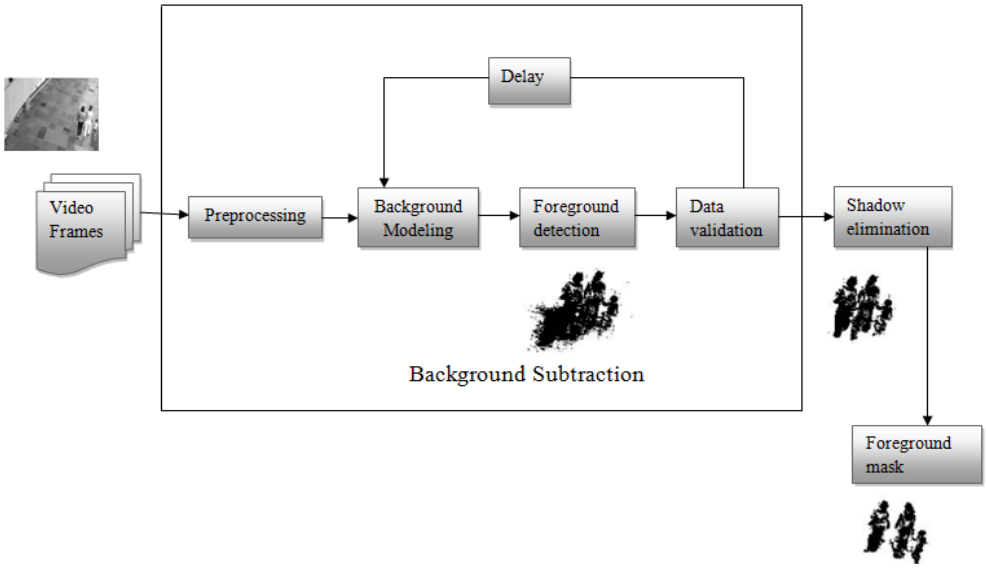


Figure 4: Block Dagram of Steps Involved in Background Subtraction and Shadow Elimination

The above figure shows the block diagram of background subtraction and shadow elimination, the video frames is taken as a input and it performs the preprocessing in that in order to remove the noise then it made to performs the background modeling to identify the foreground object and to detect them, then it validate the detected foreground object, if there is a delay in validating means then it again performs the background modeling process and then try to detect the foreground object. After foreground object detected it performs the shadow elimination to remove the shadow from the detected foreground object. Thus by performing these steps, we can obtain the foreground object from the video frames.

EXTRACTION PART

Triangulation-Based Skeleton Extraction

We presented a technique for triangulating a human body image into triangular meshes[4]. By connecting all the centroids of any two connected meshes, a graph can be formed. In this section, we use a depth-first search technique to find the skeleton that will be used for posture recognition. Figure 5 shows the block diagram of posture classification using skeletons.

In what follows, details of each block will be described. Assume that *P* is a binary posture. Using the above technique, is decomposed into a set of triangular meshes Ω_P , i.e.,$\Omega_P = \{T_i\}_{i=0,1,\ldots NTp-1.}$ Each triangular mesh in T_i in Ω_p has a centroid, and two meshes, T_i and T_j , are connected if they share one common edge. Then, based on this connectivity, *p* can be converted into an undirected graph G_p, where all centroids C_t in Ω_P are nodes on G_p ; and an edge exists between are connected. The degree of a node on the graph is defined by the number of edges connected to it. Then, we perform a graph search on G_p to extract the skeleton of *p[4]*.

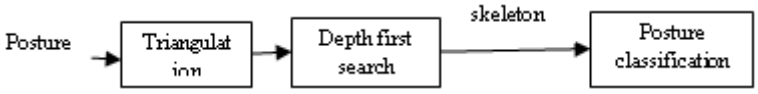


Figure 5: Flow Chart of Skeleton Based Posture Extraction

Triangulation-Based Simple Skeleton Extraction Algorithm (TSSE)

Input: a set of triangular meshes extracted from a human posture.

Output: the skeleton of S_P of P

Step 1: Construct a graph G_p from Ω_p according to the connectivity of nodes in Ω_p . In addition, get the centroid C_p from P.

Step 2: Find a node, H , whose degree is one and whose position is the highest among all nodes on Gp.

Step 3: Apply a depthfirst search to G_p to find its spanning tree.

Step 4: Get all leaf nodes L_i and branch nodes B_i from the tree. Let U be the union of H ,C_p , L_i , and $B_{i.}$

Step 5: Extract the skeleton S_p from U by linking any two nodes in U if they are connected through other nodes.

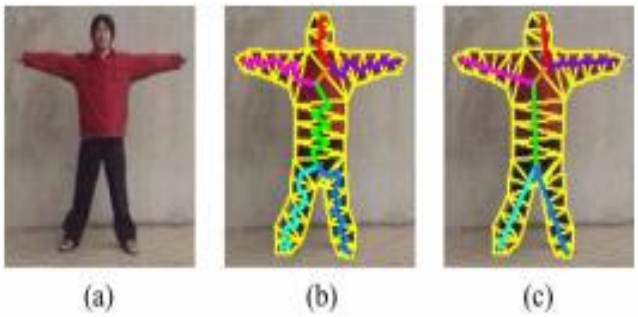


Figure 6: (a) is the Original Posture, and (b) is the Calculated Spanning Tree

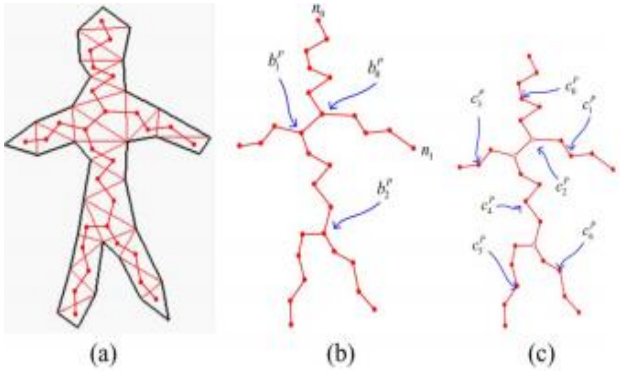
From (b), We can get the corresponding skeleton, shown in (c), using TSSE algorithm

Centroid Context-Based Description of Postures

The skeleton-based method provides a simple and efficient way to represent body postures. However, the skeleton is a coarse feature that can only be used in a coarse search process. To better characterize human postures and improve the accuracy of the recognition result, we use the centroid context[4]. A centroid context-based shape descriptor that can characterize the interior of a shape. In the literature, quite a number of global features have been proposed for shape recognition. For example, the moment descriptor has good invariance properties for trademark indexing but its calculation is very time-consuming. The Fourier descriptor is simple but easily affected by noise. Shape context is a good descriptor for shape recognition but it requires a set of dense feature correspondences. Therefore, we propose to use a new type of descriptor-"centroid context" to tackle the above problems. Basically, the descriptor can be used in the fine search process. Since the triangulation results of human postures vary, we calculate the distribution of every posture based on the relative

positions of the meshs' centroids. A descriptor of this form guarantees robustness and compactness. Assume all postures are normalized to a unit size[4]. Then, similar to the technique used in shape context, we project a sample onto a polar coordinate and label each mesh. Figure 7 shows a polar transform of a human posture. We use m to represent the number of shells used to quantize the radial axis and n to represent the number of sectors that we want to quantize in each shell. Given two postures, P and q, the distance between their centroids is measured by

$$d_{cc}(P,Q) = \frac{1}{2|V^P|}\sum_{i=0}^{|V^P|-1} w_i^P \min_{0\le j<|V^P|} C\left(c_i^P, c_j^Q\right) + \frac{1}{2|V^Q|}\sum_{j=0}^{|V^Q|-1} w_j^Q \min_{0\le i<|V^Q|} C\left(c_i^P, c_j^Q\right) \quad (9)$$



(a) (b) (c)

Figure 7: (a) Triangulation Result of a Posture (b) the Spanning Tree of (a); and (c) Centroids Derived from Different Parts (Determined by Removing all Branch Nodes)

Algorithm for Centroid Context Extraction

Input: the spanning tree of a posture .

Output: the centroid context of *p.*

Step 1: Recursively trace T^p_{dfs} using a depth first search until the tree is empty. If a branch node (a node with two children) is found, collect all the visited nodes to a new path and remove these nodes from $T^p_{dfs.}$

Step 2: If a new path only contains two nodes, eliminate it; otherwise, find its path centroid $v^p_{i.}$

Step 3: find the centroid histogram of each path centroid.

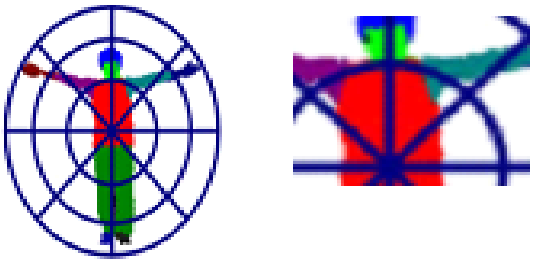
Step 4: Collect all the histograms $h_{v^p_{i(k)}}$ as the centroid context of *p.*

POLAR PARTITION

To provide the high level of security and to improve the effectiveness and accuracy, we introduce the polar transformation descriptor, this descriptor utilizes a polar labeling scheme to label every triangular mesh with a unique number. Then, for each body part, a feature vector, i.e., the centroid context is constructed by recording all related features of the centroid of each triangular mesh according to this unique number. From the constructed centroid feature by using polar transformation extract a small part and made them to compare the extracted part with portrait. We can then compare different postures more accurately by measuring the distance between their centroid contexts.

Each posture is assigned a semantic symbol so that each human movement can be converted and represented by a set of symbols. Based on this representation, we use a new spatio-temporal relationship match is used to measure similarity between two videos[6]. Our kernel function serves as a likelihood measurement between two sets of feature vectors extracted

from two videos containing human activities. An appropriate kernel function capturing characteristics of the activities is essential for classifying and detecting activity videos, which enables the correct recognition of the activities. The basic idea of our spatio-temporal relationship match is to evaluate the similarity between the structures of two sets of feature points. Given a set of spatio-temporal features extracted from a video, i.e. local video patches in a 3-D XYT space, our method calculates the spatial and temporal relationships satisfied by the feature points (e.g. point f1 is before f2 , and f1 is near f2). By comparing such relationships, the spatio-temporal relationship match measures "how many features two videos contain in common, and how many among them exhibit an identical relation". The major advantage of our match kernel is its efficient consideration of spatio-temporal structures among feature points. We represent the structure of the 3-D feature points as a set of pairwise relationships, and match them efficiently.



(a) Polar Transformation of Human Posture **Extracted Part from (a)**

Figure 8: Shows the Polar Transformation of Human Posture and Extracted Part

By applying the polar transformation for the human posture and extract a small part from the portrait and it is made to compare with portrait. This comparison is done by using spatio temporal relationship match it compares the structural similarity between the two video structures. This descriptor provides the higher security level by comparing the extracted part. After applying the string matching algorithm for comparing the distance between the portrait then, we introduce a novel spatio temporal match is used. This descriptor utilizes the polar transformation to extract a small part from the portrait and make them to compare with portrait. It is based on structural similarity[6] among the feature extracted from the video. By performing this comparison, we can better characterize the portrait and also we can improve the accuracy for video comparison.

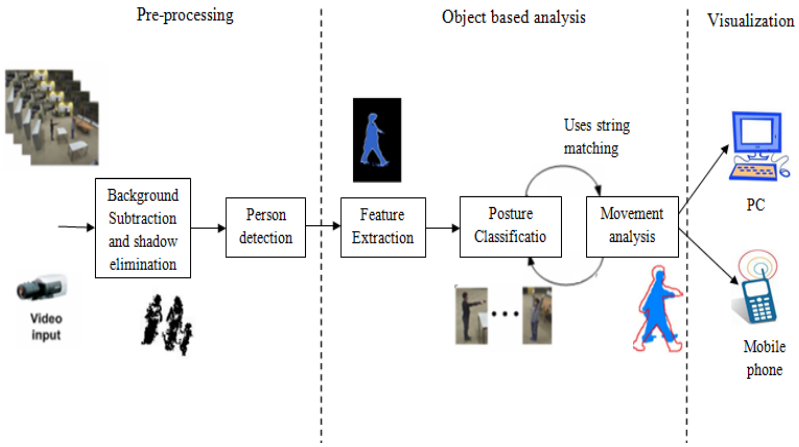


Figure 9: Block Diagram of Steps Involved in Extraction Unit and the Analysis of Movements

The above block diagram shows the method of object comparison, the basic and initial step is background subtraction and eliminating the shadow, then it performs the person detection, once the person is detected the feature extraction process is carried out. Based on the feature extracted the posture for the object is identified and classified. Finally the comparison is done based on string matching algorithm, it compares the distance between the movements from the two video. Once the comparison is done the result is made to visualize via PC or Mobilephone

CONCLUSIONS

We have presented a video object matching methodology, which is designed to detect, analyzes and compare the human activities from realistic video. Specifically, the method comprises the triangulation-based technique that extracts two important features, the skeleton feature and the centroid context feature, from a posture to derive more semantic meaning. The features form a finer descriptor that can describe a posture from the shape of the whole body or from body parts. Since the features complement each other, all human postures can be compared and classified very accurately and also to provide the high level of security, extract a small part from video and make them to compare with portrait. A novel string-based technique is used for recognizing human movements accurately. Even though events may have different scaling changes, they can still be recognized. Spatio-temporal relationship match is also used to measures a similarity between the structures of two set of feature points extracted from video and also measures "how many features two video contain in common, and how many among them exhibit in identical relation. We improve the object matching by first removing noises, errors removal and then performing the background subtraction and shadow elimination. In our approach the normalized cross correlation is applied to foreground pixels, and candidate shadow pixels are obtained. A refinement process is then applied to further improve shadow segmentation. This versatile technique has so far proven to be a unique and promising in the areas of biometrics. Thus this proposal is secure, highly usable and realistic approach to be implemented in PCs for user authentication.

REFERENCES

1. I.-C.Chang and C.-L.Huang, "The model-based human body motion analysis," Image Vis. Comput., vol. 18, no. 14, pp. 1067–1083, 2000.
2. R.Cucchiara, C.Grana, A. Prati, and R.Vezzani, "Probabilities posture classification for human-behavior analysis," IEEE Trans. Syst., Man, Cybern. A, vol. 35, no. 1, pp. 42–54, Jan. 2005.
3. S.Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," IEEE Trans. Pattern Recognit. Machine Intell., vol. 24, no. 4, pp. 509–522, Apr. 2002.
4. Jun-Wei Hsieh, Member, IEEE, Yung-Tai Hsu, Hong-Yuan Mark Liao, "Video Based Human Movement Analysis and Its Application to Surveillance Systems", IEEE Transactions on multimedia, vol. 10, no. 3, april 2008.
5. N.Werghi, "A discriminative 3D wavelet-based descriptors: Application to the recognition of human body postures,"Pattern Recognit Lett., vol. 26, no. 5, pp. 663–677, 2005.
6. M.S.Ryooand and J.K.Aggarwal,"Spatio-Temporal Relationship Match: Video Structure Comparison for Recognition of Complex Human Activities" Proceedings of the IEEE International Conference on Computer Vision (ICCV), Kyoto, Japan, Oct 2009.
7. Shuo Wang and Jing Liu P. R. China, "Biometrics on Mobile Phone", vol. 24, no. 4, pp. 509–522, Apr. 2005.
8. JulioCezar SilveiraJacques and RositoJungSoraia "Background Subtraction and Shadow Detection in Grayscale Video Sequences" IEEE Transactions on multimedia, vol. 10, no. 3, april 2009.
9. Anil K. Jain, Ajay Kumar," Biometrics of Next Generation: An Overview","second generation biometrics" springer, 2010.