AASCIT Communications
Volume 2, Issue 3
April 30, 2015 online
ISSN: 2375-3803

AASCIT American Association for Science and Technology

Human Action Recognition System

S. Maheswari Department of CSE, Manonmaniam Sundaranar University, Tirunelveli, India

P. Arockia Jansi Rani Department of CSE, Manonmaniam Sundaranar University, Tirunelveli, India

Keywords

Human Silhouette, Image Averaging, Template Matching, Correlation

Human action recognition has evoked considerable interest in the various research areas and applications due to its potential use in proactive computing. The objective of this work is to recognize various human actions like run, jump, walk etc. Moving Object detection and tracking is the first step for action recognition. The algorithm first makes use of the statistical background model and background subtraction method to extract the human action silhouettes. After extracting the silhouttes action recognition is done using template matching algorithm. Template matching algorithm employs correlation measure to find the similarity between the template and the given input.

Introduction

The action recognition is an automated analysis of on-going events and their context from video data. Proactive computing is a technology that pro-actively anticipats people's necessity in situations such as health-care or life-care and takes appropriate actions on their behalf. A system capable of recognizing various human actions has many important applications such as automated surveillance systems, human computer interaction, smart home health-care systems and control free gaming systems etc.

The rest of the article is organized as follows: Section II discusses the related work found from the literature. The proposed methodology is described in section III. Frame extraction, ROI Extraction, morphological processing, Template matching, Template Selection have been discussed in detail in this section. Section IV provides the experimental results and finally Section V concludes the work.

Related Work

Antonios Oikonomopoulos et.al (2006), focused on the problem of human action recognition using spatiotemporal events that are localized at points that are salient both in space and time. The spatiotemporal points are detected by measuring the variations in the information content of pixel neighborhoods not only in space but also in time. The classification scheme uses Relevance Vector Machines and on the chamfer distance measure. The classification results are presented for two different types of classifiers, displaying the efficiency for the representation in discriminating actions of different motion classes.

Marko Heikkila (2006), presented a texture-based method for modeling the background and detecting moving objects from a video sequence. Each pixel is modeled as a group of adaptive local binary pattern histograms that are calculated over a circular region around the pixel.

Salem Saleh Al-amri (2010), discussed about the study of segmentation techniques using threshold methods such as Mean method, P-tile method, Histogram Dependent Technique (HDT), Edge Maximization Technique (EMT) and Visual Technique for object detection. The performance comparison of these methods are reported in this paper.

M. I. Khalil (2010), presented a study of applying the template matching approach for character image recognition. A new license plate recognition system based on moving window matching algorithm has been implemented. The distance measure (squared Euclidean distance) technique has been used for measuring the similarities between the moving window and the plate image.

Olivier Barnich and Marc Van Droogenbroeck (2011), proposed a technique for motion detection that incorporates several innovative mechanisms. This technique stores, a set of values for each pixel taken in the past at the same location or in the neighborhood. It then compares this set to the current pixel value in order to determine whether that pixel belongs to the background, and adapts the model by choosing randomly which values to substitute from the background model.

Keigo Takahara (2011), proposed a robust background modeling technique. Firstly, a background model is established according to the temporal sequence of the frames. Secondly, the moving objects are detected based on the difference between the current frame and the background model. Object detection helps to identify pixels in the video frame that cannot be adequately explained by the background model, and outputs them as a binary candidate foreground mask. Finally, the background model is updated periodically to adapt the variety of the monitoring scene.

Rupali S. Rakibe et. al (2013) presented an algorithm for detecting moving objects from a static background scene to detect moving object based on background subtraction and contour projection analysis is combined with the shape analysis to remove the effect of shadow; the moving human bodies are accurately and reliably detected.

Proposed Human Action Recognition System

The proposed Human Action Recognition System is described in this section.

In this system, the input video is split into frames and then it is pre-processed to improve the brightness of the image. In any Action Recognition System, a pre-processing step is carried out to remove the noise. ROI extraction is carried out to extract the human silhouette. The extracted human silhouette is noisy so morphological processing is done to remove the artifacts or noise from the silhouette. Then using template selection templates are chosen for each action. Template matching is done using correlation for recognizing action. This process is described in figure.1.

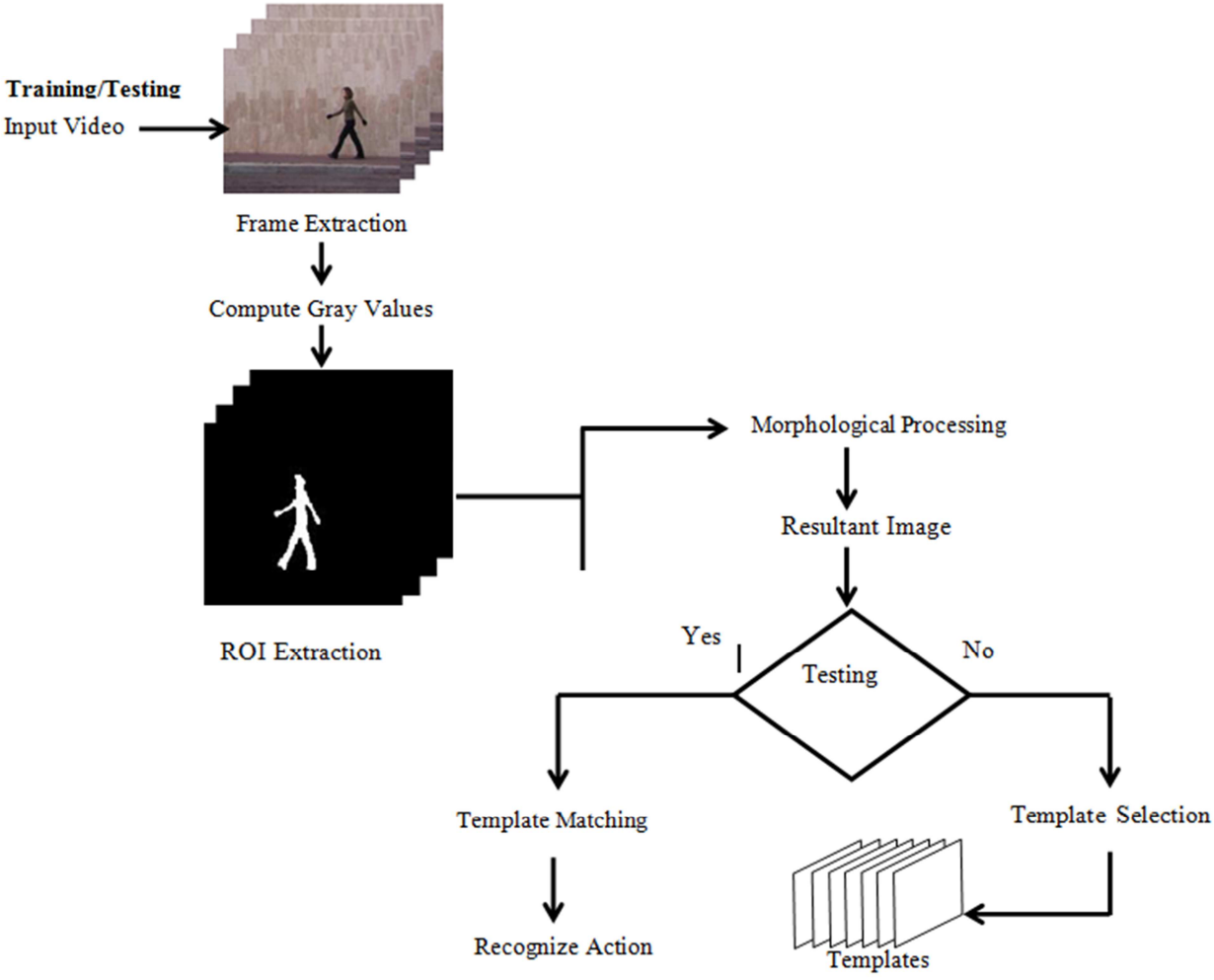

Figure 1. *Action Recognition System Architecture*

Input Selection

The input videos are taken from Weizmann Dataset. The input videos are recorded in a homogeneous background with a static camera. The input videos include walk, run, jump, jack, wave, and jump with run actions. The Weizmann Dataset is downloaded from the website www.wisdom.weizmann.ac.il/vision/SpaceTimeActions.The properties of the input videos are

Type: VLC media file (. avi)

Resolution: 180 x 144

Frame rate: 25 frames per second

Frame Extraction

The input video is split into frames. Frame rate refers to the number of frames that are projected or displayed per second. The luminance value for each pixel is extracted from the color image eliminating the respective hue and saturation value.

ROI Extraction

It is the process of separating the foreground object from the adaptive background. As datasets are taken using a static camera, background subtraction techniques can be employed to compute foreground information. The objects are identified using image subtraction techniques.

$$F(u,v) = f(x,y) - p(x,y)$$

Where $p(x,y)$ is the background pixel, f(x, y) is the pixel in the input frame

The resultant foreground information is represented as a binary image which is called silhouette information. This image contains black background with white silhouette information. The silhouettes can be analysed to recognize the human actions. As silhouettes describe the outer contours of a person, it provides strong cues for action recognition.

Morphological Processing

The morphological processing is done to extract the features of interest in an image. Morphological processing is constructed with operations on set of pixels.

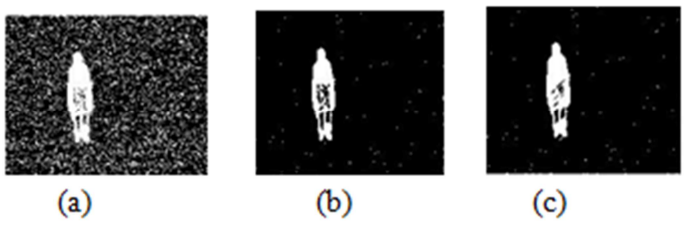

Figure 2. *(a)Noisy image (b) Noisy Image after Erosion (c)Minimization of Noise after erosion and closing.*

The resultant silhouette after ROI extraction is Noisy. In order to remove that noise Morphological processing is applied. The noisy image is the input of Morphological erosion. Morphological erosion is used for removing the unwanted portions other than the object of interest. Then Morphological closing is used for filling the holes in the object. Finally the silhouette of the object is extracted clearly.

Template Selection

Template Selection is based on image averaging. Image Averaging is obtained by finding the average of k images. K represents the number of images taken for Image averaging.

Figure 3. *Silhouettes of Walk Action*

The silhouette of five different persons performing walk action is shown. Image Averaging result is obtained from these five silhouettes for walk action. The resultant averaged image is set as the template for the respective action. Similarly Image Averaging is found out for all class of actions. Finally for each class of action the template is chosen.

Template Matching

Recognition is based on template matching. The action instances like run, jump, walk, jack, hand-waving and jump with run are considered for analysis. The actions are registered in terms of both time and frame rate. These action instances have different time and frame rate for different human actions. The similarity between the template and the resultant silhouette is measured using correlation measure.

$$r = \frac{\sum\sum(A_{mn} - \overline{A})\,(B_{mn} - \overline{B})}{\sqrt{(\sum\sum(A_{mn} - \overline{A})^2)(\sum\sum(B_{mn} - \overline{B})^2)}}$$

r is the correlation coefficient. $\bar{A}$ is the mean of A and $\bar{B}$ is the mean of B. Here A is the template and B is resultant silhouette of the input. The silhouette is matched with the predefined templates using correlation. Then the template having maximum correlation is recognized as the output class.

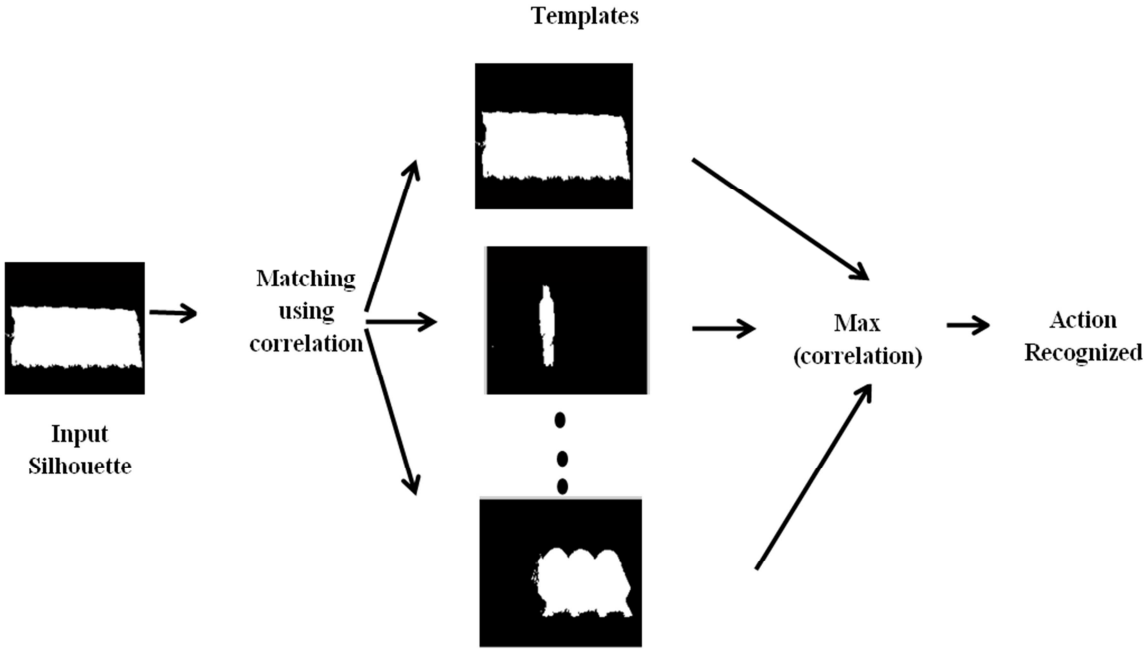

Figure 4. *Template Matching Process*

Experimental Results & Analysis

Weizmann Action Dataset is used for analysing the results. The performance of the system is analysed by testing different human actions. A common quantitative analysis is performed to assess the overall performance of recognition process. To analyse the performance precision, recall and F-measures are used.

Precision, also called as the positive predictive value is the fraction of retrieved action instances that are relevant. Precision is calculated by means of

$$precision = \frac{tp}{tp + fp}$$

Recall, also known as sensitivity is the fraction of relevant instances that are retrieved.

Recall is calculated by means of

$$Recall = \frac{tp}{tp + fn}$$

F-measure, combines precision and recall called as the harmonic mean of precision and recall. The balanced F-score is

$$F = 2\frac{precision.recall}{precision + recall}$$

Accuracy is the proportion of true matches (both true positive and true negative) in the total input actions.

$$\text{Accuracy} = \frac{nTP + nTN}{nTP + nTN + nFP + nFN}$$

The precision, recall, accuracy and F-measures obtained for the Weizmann dataset using template matching is shown in the Table1

Table 1. *Experimental results using template matching*

Actions	No.of Actions	Precision	Recall	Accuracy	F-measure
Run	10	0.70	0.70	0.89	0.70
Walk	10	0.89	0.80	0.95	0.84
Jump	9	1.00	1.00	1.00	1.00
Jack	9	1.00	1.00	1.00	1.00
Double sided wave	9	1.00	1.00	1.00	1.00
Jump with Run	9	0.70	0.78	0.91	0.74
Average		0.88	0.88	0.958	0.88

$$Accuracy(\%) = Average\ Accuracy\ X\ 100$$

The accuracy of the proposed work is 95.8%.The jump, jack and double sided wave actions produce 100% result. There may be a chance of misclassifying the actions run and walk. But here walk action produces 95% result and run action produces 89% result.

From the literature, we have collected various methods reported for action recognition and listed them year wise reference in the Table2.

Table 2. *Comparative Results on Existing methods and the Proposed Method*

	Technique	Year	Accuracy
1.	Proposed Human Action Recognition System	[2015]	95.8%
2.	Fusing Spatio Temporal Appearance	[2014]	93.5%
3.	Gabor Filters With Gradients And PLSA	[2008]	90.0%
4.	Spin With ST Features	[2008]	90.4%
5.	Harris3D With HOG3D	[2008]	84.3%
6.	3d-Sift	[2007]	82.6%
7.	Motion Context With Foreground Segmentation	[2008]	92.9%
8.	Chaotic Invariants With Silhouettes	[2007]	92.6%
9.	Shape Context With Gradients And PCA	[2007]	72.8%

The performance of the proposed method is compared with various existing methods listed in table 3.It shall be observed that the proposed method gives an accuracy of 95.8%,which is higher than that observed using the existing methods reported in this literature.

Conclusion

In this paper action is recognized using image averaging and correlation based template matching. The experiments on

Weizmann datasets show the performance of the proposed work in comparison with the existing approaches. Other datasets like KTH dataset, HOHa, UCF dataset can also be used. This system provides an efficient result in recognizing the actions. In future work, this approach may be extended for action recognition in crowded environments and may be applied in human interactive behaviour recognition.■

S. Maheswari

She is doing her Ph.D in Depatment of Computer Science and Engineering in Manonmaniam Sundaranar University, Tirunelveli, India. She received her B.E Degree in Computer Science and Engineering from Dr. G. U. Pope College of Engineering, Tuticorin, India in 2011 and M.E degree in Computer Science and Engineering in Manonmaniam Sundaranar University Tirunelveli, India in 2013.Her research interests include Video processing.
Email: jan20mahi91@yahoo.com

P. Arockia Jansi Rani

She received her B.E in Electronics and Communication Engineering from Government College of Engineering, Tirunelveli, Tamil Nadu, India in 1996 and M.E in Computer Science and Engineering from National Engineering College, Kovilpatti, Tamil Nadu, India in 2002. She has been with the Department of Computer Science and Engineering, Manonmaniam Sundaranar University as Assistant Professor since 2003. She has more than ten years of teaching and research experience. She completed her Ph. D in Computer Science and Engineering from Manonmaniam Sundaranar University, Tamil Nadu, India in 2012. Her research interests include Digital Image Processing, Neural Networks and Data Mining.
Email: jansi_msu@yahoo.co.in

References

[1] Haritaoglu, D. Harwood, and L. S. Davis, " Real-time surveillance of people and their activities", IEEE T-PAMI, 22:809–830, 2000

[2] W. Hu, T. Tan, L. Wang, and S. Maybank. "A survey on visual surveillance of object motion and behaviors", IEEE Transactions on Systems, Man and Cybernetics, 34:334–352, 2004.

[3] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes", In ICCV, 2005.

[4] Marko Heikkila and Matti Pietika inen, Senior Member, IEEE, "A Texture-Based Method for Modeling the Background and Detecting Moving Objects" IEEE Transactions On Pattern Analysis And Machine Intelligence,Vol. 28,No. 4,April 2006

[5] Z. Zhang, Y. Hu, S. Chan, and L.-T. Chia. "Motion context: A new representation for human action recognition" In ECCV, 2008.

[6] Senior. "An introduction to automatic video surveillance. In Protecting Privacy in Video Surveillance" Springer, 2009.

[7] Salem Saleh Al-amri, N.V. Kalyankar and Khamitkar S.D, "Image Segmentation by Using Thershod Techniques" Journal Of Computing, Volume 2, Issue 5, May 2010.

[8] M.I.Khalil, "Car Plate Recognition Using the Template Matching Method", International Journal of Computer Theory and Engineering, Vol. 2, No. 5, October, 2010.

[9] O. BARNICH and M. VAN DROOGENBROECK, "ViBe: A universal background subtraction algorithm for video sequences" IEEE Transactions on Image Processing, 20(6): 1709-1724, June 2011.

[10] Keigo Takahara, Takashi Toriu and Thi Thi Zin "Making Background Subtraction Robust to Various Illumination Changes" IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.3, March 2011

[11] Rupali S.Rakibe, Bharati D.Patil "Background Subtraction Algorithm Based Human Motion Detection" International Journal of Scientific and Research Publications, Volume 3, Issue 5, May 2013.