NEW TECHNIQUES IMPROVING ACCURACY OF COMPUTER VISION TECHNOLOGIES

[1]Muthu Dayalan
Researcher & Software Professional, Chennai & Tamil Nadu

Abstract- ***The computer vision technology has continued to evolve through advancement in researches. This has continually widened the computer vision and imaging processing to extended applications. This paper analyzes the technological advancement in these aspects. Machine vision system is among the emergent technologies, which digitizes the video images or information from a camera, processes the information to inspect the concerned object, and communicates with the system or the motion control[1]. 3D and virtual environments help to mimic the real life situations. The abnormality detection for the video parsing helps to solve the inter-class variability, viewpoint dependence and the object articulation, which makes the practice of object detection quite difficult. A video surveillance where foreground pixels are detected through the background subtraction. The curvature self-similarity visual recognition is a feature descriptor, which increase dimensionality such as the self-similarity and the co-occurrence statistics. It extends beyond the current common approximation of the objects by the use of straight lines. The result on the self-similarity curvature proves that there is provision of the accurate information on the straight lines, while the performance has improved based on the feature selection algorithm.***

Index Terms-virtual environments, vision system, 3D, video parsing, virtual recognition, self-similarity

INTRODUCTION

As the computer vision technology continues to evolve, the sector continually advances and the application diversifies. The sector is considered accommodative to wide range of applications, which has facilitated extended researches for improvement. The computer vision and the imaging processing have wide application, ranging from the anomalies in the medical imaging to the autonomous vehicles. Generally, computer imaging is involves the automatic extraction, analysis and understanding of relevant information from a single or a sequence of images.[2] The automatic visual understanding is developed based on theory and algorithms. By simple definition, computer vision is expressed as the science and technology of systems, which see and understand. Currently, it is offering solution to challenges in data acquisition and exploitations, in defense, security, training, robotics, medical marketplace among others. The end-to-end computer vision capabilities include aspects such as high-level visual analysis, enhanced sensation, and front-end video processing. The extraction of relevant information from the acquired images, under the real world application, and in real time calls for utilization of the right set of tools.[3] This has always existed as a challenge, which has led to the continued research to devise new techniques for improving the accuracy of the computer vision. Based on this background, this paper presents a comprehensive analysis of the new and emergent techniques, developed for improving the accuracy of the computer vision techniques.

MACHINE VISION SYSTEM

The computer vision is a vital component of the machine vision system. Through technology, the vision computer digitizes the video images or information from a camera, processes the information to inspect the concerned object, and communicates with the system or the motion control. The images and information is afterwards displayed as a result, for the human viewing. The process of extraction of the visual information requires execution off many occupational operations for a short period, hence the application of the parallel processing, to ensure the required speed. The encoders and position sensors could also be used to send input to the vision computers, and control the lighting and exposure of the camera under inspection. Within the specified time and cost, the vision computer should extract the required information from the required image. The general computer could be applied for the simple positioning and inspection tasks.[4] However, vision computer is required for high sped inspection and measurements of the objects. A critical aspect is that the computer vision is required to be flexible, since there are usually the changes in the problem requirements, during the development. The development tools are considered vital for the faster experimentation and prototyping.

3-D AND VIRTUAL ENVIRONMENTS

Under the exploration of the environment in which the human being operate, there is a requirement of sophisticated virtual environment tools, which could be applied in the simulation of multiple scenarios, as well as improve the situational understanding. Therefore, the computing power and the ability of learning within the 3-D virtual environments helps in mimic of the real life situations. Various groups such as the emergency response teams, the civilian workforce, the military personnel, as well as other categories could assume the realistic situation of assessing their

[1] McWalter, Kathy. "Machine Vision Boosts PCB Assembly Processes." *Assembly Automation* 13.2 (1993): 32.

[2] Bilal, Sara, et al. "Hidden Markov Model for Human to Computer Interaction: A Study on Human Hand Gesture Recognition." *The Artificial Intelligence Review* 40.4 (2013): 495-516.

[3] Chen, Chen, Roozbeh Jafari, and Nasser Kehtarnavaz. "A Survey of Depth and Inertial Sensor Fusion for Human Action Recognition." *Multimedia Tools and Applications* 76.3 (2017): 4405-25.

[4] Li, Yao, et al. "Mining Mid-Level Visual Patterns with Deep CNN Activations." *International Journal of Computer Vision* 121.3 (2017): 344-64.

skills and knowledge before the actual execution.[5] The emergence and adoption of the robotic systems, which automatically creates the 3-D models and maps under complex environment, help in supporting mission critical uses. More importantly, the understanding accorded by the 3-D mapping projection capabilities is more accurate.

ABNORMALITY DETECTION FOR VIDEO PARSING

The inter-class variability, viewpoint dependence and the object articulation makes the practice of object detection quite difficult. The various ways in which an object behaves abnormally makes the process of detecting abnormal objects even harder[6]. The common trend has been the use of local image patches where they are categorized individually. However, a more advanced and effective tool is the application of the video parsing novel approach. Under this technique, a set of object hypothesis are established, for parsing the concerned video frames. In this way, the hypothesis explains the foreground as well as establishing the hypothesis themselves. In this case, an abnormal hypothesis is considered to the one which necessary to explain the foreground, but it does not have a match at the training examples. The abnormal objects are discovered through parsing the scene, and inference of all the required object hypotheses without a prior knowledge of what to look for.

Figure 1: The video parsing approach inference process

A good example of video parsing application is the video surveillance. In this case, a stationery camera is used to record a video. The foreground pixels are detected through the background subtraction. This helps in the calculation of the flow features, which expresses the objects' local behavior[7]. The critical aspect of this computer vision technology is that it follows a two-stage approach. The first step involves the generation of a shortlist of the object hypotheses. These are created through evaluation of a little appearance-based classifier, at the foreground of the candidate object hypothesis. In order to have a retention of a moderate number off hypotheses, the classifier is set to have a low precision and a low recall. The order of 10 to 100 of hypotheses is maintained[8]. Relevance of the hypotheses is a critical factor here. The following figure illustrates the optical flow of features, which are computed in the pre-processing.

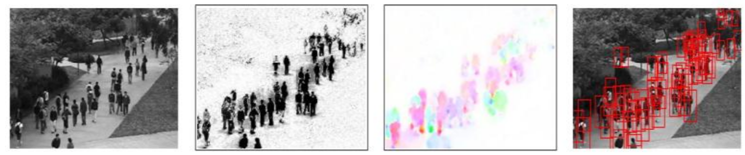

Figure 2: Pre-processing of the video parsing method

The second stage involves the parsing of the scene through selection of subset hypotheses, from the list demonstrating the whole foreground. This is additionally explained through the use of object model. The non-parametric object model is applied for the sampling of the normal objects obtained from the training video. The nonparametric object model is shown in the table below.

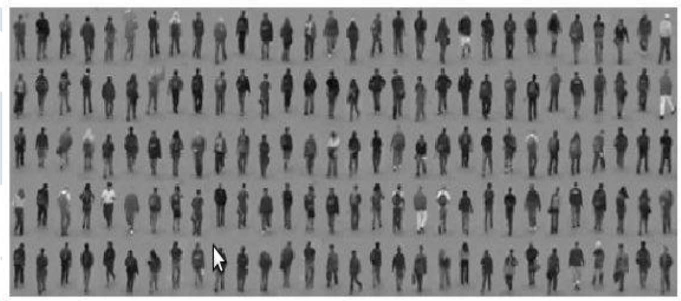

Figure 3: The non-parametric object model

Since the activation or deactivation of the hypotheses, and their object model explanation are mutually competing, then they have to be solved simultaneously. However, as suggested by researchers, the hypotheses selection problem is combinatorial in nature[9]. This makes the underlying process of optimization quite difficult, which is alternatively solved. The inference process exploits the explaining-away property. Therefore, if the same section of the foreground explains two basic hypotheses, then the performance is reduced. The case is that one hypotheses activation decreases the chances of the other hypotheses being active. The superfluous hypotheses are deactivated by this property, to retain those properties, which are important in explanation of the scene.

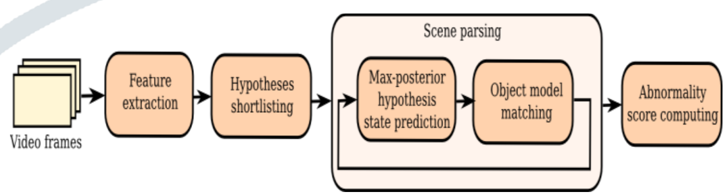

Figure 4: the abnormality detection parsing method

[5] Essa, Irfan A. "Computers Seeing People." *AI Magazine* 20.2 (1999): 69.

[6] Models." *Multimedia Tools and Applications*76.11 (2017): 13585-614.

[7] D'acunto, M., et al. "A Methodological Approach for Combining Super-Resolution and Pattern-Recognition to Image Identification." *Pattern Recognition and Image Analysis* 24.2 (2014): 209-17.

[8] Thomas, Nitha, Joshin John Mathew, and Alex James. "Threshold Logic Based Low-Level Vision Sparse Object Features." *International Journal of Intelligent Computing and Cybernetics* 9.4 (2016): 314-24.

[9] Yao, Angela, Juergen Gall, and Luc Van Gool. "Coupled Action Recognition and Pose Estimation from Multiple Views." *International Journal of Computer Vision* 100.1 (2012): 16-37.

It is very critical to conduct an evaluation of the parsing approach based on the UCSD dataset[10]. This is a vital benchmark for the abnormality detection, which was considered and proposed by Mahadevan et al. In this case, there is no staging in the dataset abnormalities, but rather consists of unusual objects such as the walkaways or the unusual behavior such as the cycling people. The training and test video are quite different, where the training video consists of the normal objects, while the test videos are made up of both normal as well as the abnormal instances.

CURVATURE SELF-SIMILARITY VISUAL RECOGNITION

As illustrated by PPPPP, the detection of the category level object needs perfect informative object representation. This demand has led to the discovery of the feature descriptors, for the increasing dimensionality such as the self-similarity and the co-occurrence statistics. One of the new techniques in this respect is the object representation, based on the curvature self-similarity. This technology extends beyond the current common approximation of the objects by the use of straight lines. The motivation behind this aspect is the adoption of the gradient histogram by the vision community. There is a resultant object boundaries approximation of the straight line, since there approximates the local regions by a histogram, over the discrete set of contained discrete set of the edge orientation.

It is quite difficult to distinguish between a smooth from the cornered one, or from a set of varying oriented lines, based particularly on the oriented gradients histograms[11]. The following features illustrate this.

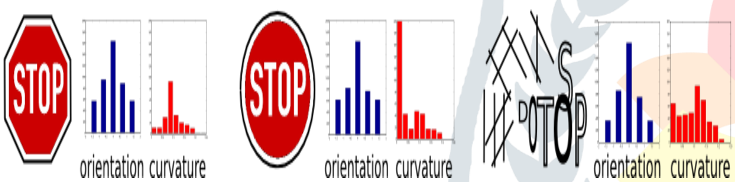

Figure 5: Smooth and cornered curves

The approximation of the contour curvature co-occurrence could provide more information regarding the boundaries of an object, hence decreasing the level of distinctiveness[12]. The computation of the co-occurrence between the boundaries, which are discriminatively, curved, the self-similarity descriptor curvature id developed, whose object description is more accurate as well as detailed. However, it is important to note that when the discriminability is improved, the high dimensionality turns to be a critical aspect. This is due to the absence of the dimensionality curse and globalization ability. More critically, in the presence of the little amount of training data, even the sophisticated and high-end algorithms cannot manage to suppress the superfluous dimensions like the high dimensional data[13]. As a result, it important to have a feature selection, by the use of the present day informational features, and more importantly, the curvature self-similarity. Therefore, the recommended technology in this case is the SVMs embedded feature selection technique, which minimizes the complexity and enhances the generalization capabilities of the objects models.

THE SVM SELECTION OF FEATURES

Three categories of feature selection could be applied. These are the embedded, wrappers and filters techniques. The embedded technique, in comparison to the wrappers and filters, involves the selection feature as a section of the process. The selection process involving the SMVs is a critical aspect in this research, since it involves SMV is considered as a classifier. The model parameter *w* is used for integrating the SMV learning process in to the feature selection. However, it does not reduce the dimensionality for the functions of the non-linear kernel. The embedded feature selection concept has been selected which incorporates the feature selection parameter denoted as θ, which is included directly in to the SMV classifiers. The training data is the {(xi,yi)}where 1≤i≤N. The training data labels are denoted by yi∈{−1,+1}. By the use of these aspects, the concerned optimization problem could be developed and expressed in the following ways:

Min $\frac{1}{2}\|w\|_2^2 + C\sum \mathbb{C}i$

This is subject to:

Yi (wTψ (θ∗xi) +b) ≥1−ℂi ∧ ℂi ≥ 0 ∧ ‖θ‖1 ≤ θ0

The SVM kernel in this case is denoted by K(x,z):=ψ(x).ψ(z)K(x,z):=ψ(x).ψ(z)

The feature vector z is denoted by an unknown function indicated as ψ(x) which is typically a higher dimension space. The L1 norm is restrained by the constant θ0. To solve the problem, the training data is split in to two-the validation set and the training set. The second optimization process is applied for the parameter θ, by the application of the SMV algorithm.[14] The second optimization process, on the other hand is applied for minimizing the feature vector dimensionality, and improves the performance of the classification.

To represent the holistic self-similarity the curvature aspect is applied. The curvature data is first obtained throughout the image edge pixels. The distance accumulation technique is applied for the estimating the curvature. As a result, the level of a curve in the indicated 2D line segment. However, the complex objects could not be described solely by the self-similarity descriptor[15]. This is

[10] Sagonas, Christos, et al. "Robust Statistical Frontalization of Human and Animal Faces." *International Journal of Computer Vision* 122.2 (2017): 270-91.

[11] Zheng, Yu, and Xinbo Gao. "Indoor Scene Recognition Via Multi-Task Metric Multi-Kernel Learning from RGB-D Images." *Multimedia Tools and Applications* 76.3 (2017): 4427-43.

[12] Patil, Hemprasad, Ashwin Kothari, and Kishor Bhurchandi. "3-D Face Recognition: Features, Databases, Algorithms and Challenges." *The Artificial Intelligence Review* 44.3 (2015): 393-441.

[13] Uddin, Md Zia. "A Depth Video-Based Facial Expression Recognition System Utilizing Generalized Local Directional Deviation-Based Binary Pattern Feature Discriminant Analysis." *Multimedia Tools and Applications* 75.12 (2016): 6871-86.

[14] Neves, João, et al. "Biometric Recognition in Surveillance Scenarios: A Survey." *The Artificial Intelligence Review* 46.4 (2016): 515-41.

[15] Wang, Chong, and Kai-qi Huang. "VFM: Visual Feedback Model for Robust Object Recognition." *Journal of Computer Science and Technology* 30.2 (2015): 325-39.

because it is associated with several ambiguities. Therefore, to address the issue, a more precise descriptor is applied which adds the orientation information at the level of 360 degree orientation. A good example of the original images, alongside their self-similarity curvature is presented below. Though the curvature similarity may appear the same for all the object category, it appears varying for the other categories.

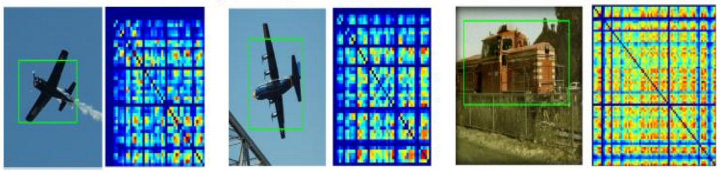

Figure 6: Self-similarity curvature

The result on the self-similarity curvature proves that there is provision of the accurate information on the straight lines, while the performance has improved based on the feature selection algorithm[16]. The powerful recognition is established through combination of the HOG and SVM components, which has led to discovery of powerful recognitions systems. Overall, this technique reduces the dimensionality involved in such high-dimensional aspects, which reduces the learning process complexity, and improve the performance.

REFERENCES

[1] Abbad, A., et al. "Post-Processing of Dimensionality Reduction Methods for Face Recognition." Pattern Recognition and Image Analysis 27.2 (2017): 266-75.

[2] Arunnehru, J., and M. Kalaiselvi Geetha. "Difference Intensity Distance Group Pattern for Recognizing Actions in Video using Support Vector Machines." Pattern Recognition and Image Analysis 26.4 (2016): 688-96.

[3] Bilal, Sara, et al. "Hidden Markov Model for Human to Computer Interaction: A Study on Human Hand Gesture Recognition." The Artificial Intelligence Review 40.4 (2013): 495-516.

[4] Carreira, João, Fuxin Li, and Cristian Sminchisescu. "Object Recognition by Sequential Figure-Ground Ranking." International Journal of Computer Vision 98.3 (2012): 243-62. Chen, Chen, Roozbeh Jafari, and Nasser Kehtarnavaz. "A Survey of Depth and Inertial Sensor Fusion for Human Action Recognition." Multimedia Tools and Applications 76.3 (2017): 4405-25. .

[5] Chen, Meng, et al. "Action Recognition using Lie Algebrized Gaussians Over Dense Local Spatio-Temporal Features." Multimedia Tools and Applications74.6 (2015): 2127-42.

[6] Chen, Meng, et al. "Modeling Spatio-Temporal Layout with Lie Algebrized Gaussians for Action Recognition." Multimedia Tools and Applications75.17 (2016): 10335-55.

[7] Chen, Pei, and David Suter. "An Analysis of Linear Subspace Approaches for Computer Vision and Pattern Recognition." International Journal of Computer Vision 68.1 (2006): 83-106.

[16] Palmero, Cristina, et al. "Multi-Modal RGB-Depth-Thermal Human Body Segmentation." *International Journal of Computer Vision* 118.2 (2016): 217-39.

[8] Chernyshov, V. G., and L. M. Mestetskii. "Mobile Computer Vision System for Hand-Based Identification." Pattern Recognition and Image Analysis 25.2 (2015): 209-14.

[9] D'acunto, M., et al. "A Methodological Approach for Combining Super-Resolution and Pattern-Recognition to Image Identification." Pattern Recognition and Image Analysis 24.2 (2014): 209-17.

[10] Essa, Irfan A. "Computers Seeing People." AI Magazine 20.2 (1999): 69.

[11] González-ortega, D., et al. "MLP-Based Face Recognition with Gabor Filters and PCA." Pattern Recognition and Image Analysis 23.1 (2013): 10-25.

[12] Gudigar, Anjan, et al. "Multiple Thresholding and Subspace Based Approach for Detection and Recognition of Traffic Sign." Multimedia Tools and Applications 76.5 (2017): 6973-91.

[13] Hayat, Munawar, Salman H. Khan, and Mohammed Bennamoun. "Empowering Simple Binary Classifiers for Image Set Based Face Recognition." International Journal of Computer Vision 123.3 (2017): 479-98.

[14] He, Ruhan, et al. "Integral Region-Based Covariance Tracking with Occlusion Detection." Multimedia Tools and Applications 74.6 (2015): 2157-78.

[15] Jiang, Hairong, Juan Wachs P., and Bradley Duerstock S. "Integrated Vision-Based System for Efficient, Semi-Automated Control of a Robotic Manipulator." International Journal of Intelligent Computing and Cybernetics 7.3 (2014): 253-66.

[16] Kazanskiy, N. L., and S. B. Popov. "Integrated Design Technology for Computer Vision Systems in Railway Transportation." Pattern Recognition and Image Analysis 25.2 (2015): 215-9.

[17] Khan, Babar, et al. "Bio-Inspired Approach to Invariant Recognition and Classification of Fabric Weave Patterns and Yarn Color." Assembly Automation 36.2 (2016): 152-8.

[18] Khan, Fahad Shahbaz, et al. "Coloring Action Recognition in Still Images." International Journal of Computer Vision 105.3 (2013): 205-21.

[19] Lee, Young-hae, Seungki Moon, and Hongchul Lee. "A New Approach for Automated Parts Recognition using Time Series Analysisand Neural Networks." Journal of Intelligent Manufacturing 8.3 (1997): 167-75.

[20] Li, Yao, et al. "Mining Mid-Level Visual Patterns with Deep CNN Activations." International Journal of Computer Vision 121.3 (2017): 344-64.

[21] Marques, Oge, Elan Barenholtz, and Vincent Charvillat. "Context Modeling in Computer Vision: Techniques, Implications, and Applications." Multimedia Tools and Applications 51.1 (2011): 303-39.

[22] McWalter, Kathy. "Machine Vision Boosts PCB Assembly Processes." Assembly Automation 13.2 (1993): 32.

[23] Neves, João, et al. "Biometric Recognition in Surveillance Scenarios: A Survey." The Artificial Intelligence Review 46.4 (2016): 515-41.

[24] Palmero, Cristina, et al. "Multi-Modal RGB-Depth-Thermal Human Body Segmentation." International Journal of Computer Vision 118.2 (2016): 217-39.

[25] Patil, Hemprasad, Ashwin Kothari, and Kishor Bhurchandi. "3-D Face Recognition: Features, Databases, Algorithms and Challenges." The Artificial Intelligence Review 44.3 (2015): 393-441.

[26] Rautaray, Siddharth S., and Anupam Agrawal. "Vision Based Hand Gesture Recognition for Human Computer Interaction: A Survey." The Artificial Intelligence Review 43.1 (2015): 1-54.

[27] Sagonas, Christos, et al. "Robust Statistical Frontalization of Human and Animal Faces." International Journal of Computer Vision 122.2 (2017): 270-91. ProQuest. Web. 11 Aug. 2017.

[28] Shamir, Lior, et al. "IICBU 2008: A Proposed Benchmark Suite for Biological Image Analysis." Medical and Biological Engineering and Computing 46.9 (2008): 943-7.

[29] Tang, Peng, et al. "Learning Extremely Shared Middle-Level Image Representation for Scene Classification." Knowledge and Information Systems 52.2 (2017): 509-30.

[30] Thomas, Nitha, Joshin John Mathew, and Alex James. "Threshold Logic Based Low-Level Vision Sparse Object Features." International Journal of Intelligent Computing and Cybernetics 9.4 (2016): 314-24.

[31] Uddin, Md Zia. "A Depth Video-Based Facial Expression Recognition System Utilizing Generalized Local Directional Deviation-Based Binary Pattern Feature Discriminant Analysis." Multimedia Tools and Applications 75.12 (2016): 6871-86.

[32] ---. "Human Activity Recognition using Segmented Body Part and Body Joint Features with Hidden Markov Models." Multimedia Tools and Applications76.11 (2017): 13585-614. ProQuest. Web. 11 Aug. 2017.

[33] Wang, Chong, and Kai-qi Huang. "VFM: Visual Feedback Model for Robust Object Recognition." Journal of Computer Science and Technology 30.2 (2015): 325-39.

[34] Wang, Hao, et al. "Exploring Hybrid Spatio-Temporal Convolutional Networks for Human Action Recognition." Multimedia Tools and Applications 76.13 (2017): 15065-81.

[35] Wei, Xue, Son Lam Phung, and Abdesselam Bouzerdoum. "Visual Descriptors for Scene Categorization: Experimental Evaluation." The Artificial Intelligence Review 45.3 (2016): 333-68.

[36] Wu, Hao, et al. "Recognition Improvement through Optimized Spatial Support Methodology." Multimedia Tools and Applications 75.10 (2016): 5603-18.

[37] Yang, Wenji, Lingfu Kong, and Mingyan Wang. "Hand Gesture Recognition using Saliency and Histogram Intersection Kernel Based Sparse Representation." Multimedia Tools and Applications 75.10 (2016): 6021-34.

[38] Yao, Angela, Juergen Gall, and Luc Van Gool. "Coupled Action Recognition and Pose Estimation from Multiple Views." International Journal of Computer Vision 100.1 (2012): 16-37.

[39] Zheng, Yu, and Xinbo Gao. "Indoor Scene Recognition Via Multi-Task Metric Multi-Kernel Learning from RGB-D Images." Multimedia Tools and Applications 76.3 (2017): 4427-43.