ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization *Vol. 5, Special Issue 2, March 2016*

National Conference on Future Technologies in Power, Control and Communication Systems (NFTPCOS-16)

on 10, 11 and 12th March 2016

Organised by

Dept. of EEE, College of Engineering Perumon, Kollam, Kerala – 691601, India

Design of Novel Algorithm for Spatio-Temporal Point Detection in 2D Color Space

Anasmon.A.K. [1], Aneesh.A.S [2]

Assistant Professor, Dept. of ECE, Sivaji College of Engineering and Technology, Tamilnadu, India[1]

PG Student [Applied Electronics], Sivaji College of Engineering and Technology, Tamilnadu, India [2]

ABSTRACT: High-level approaches for unconstrained human activity recognition aim concepts and may build-on low-level building blocks which typically consider generic video representations based on local photometric features. Local image features or interest points provide compact and abstract representations of patterns in an image. Interest point detection in still images is studied using computer vision. In the spatiotemporal domain, it is still unclear which features indicate useful interest points. It is proposed to extend the notion of spatial interest points into the spatio-temporal domain and show how the resulting features often reflect interesting events that can be used for a compact representation of video data as well as for its interpretation. To detect spatio-temporal events, a new concept is evolved on the idea of Gabor interest point operators and it detects local structures in space-time where the image values have significant local variations in both space and time. Then estimate the spatio-temporal extents of the detected events and compute their scale-invariant spatio-temporal descriptors. These descriptors are used to classify the events and construct video representation in terms of labeled space-time points. Such image points are frequently denoted as "interest points" and are attractive due to their high information contents. Highly successful applications of interest point detectors have been presented for image indexing, stereo matching, optic flow estimation and tracking, and recognition.

KEYWORDS: Bridgeless rectifier, discontinuous conduction mode(DCM),power factor correction, modified SEPIC rectifier, total harmonics distortion (THD).

I. INTRODUCTION

This work suggests playing a central role in video data that is abundantly available in archives and on the internet. Information about the presence of human activities is therefore valuable for video indexing, retrieval and security applications. However, these applications demand recognition systems to operate in unconstrained scenarios. For this reason, research has shifted from recognizing simple human actions under controlled conditions to more complex activities and events 'in the wild'. This requires the methods to be robust against disturbing effects of illumination, occlusion, viewpoint, camera motion, compression and frame rates. High-level approaches for unconstrained human activity recognition aim at modeling image sequences based on the detection of high level concepts and may build on low-level building blocks which typically consider generic video representations based on local photometric features.

For illustration purposes we have polled the detectors for the 55 strongest STIPs in the original 55-frame sequence, and show the detections on frame 48(Color online).complex, computationally expensive video processing operations, but may be superior to low-level approaches in terms of recognition rates. However, high-level approaches are sensitive to local geometric disturbances such as occlusion, which limits their applicability. Low-level approaches are conceptually simple, relatively easy to implement and potentially sparse and efficient. Due to the local nature of features on which low-level approaches are based, they are inherently robust against recording disturbances such as occlusion and clutter. Therefore, in this paper, we focus on low-level representations for recognizing human actions in video.

Low-level action recognition approaches are often based on spatio-temporal interest points (STIPs). Here, image sequences are represented by descriptors that are extracted locally around STIP detections. The descriptors are vector quantized based on a visual vocabulary, and subsequent learning and recognition operates on these quantized descriptors, comprising the well known bag-of-(spatio-temporal)-features framework. The formulations of spatio-temporal feature detectors and descriptors available in literature are based on single-channel intensity representations,

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization *Vol. 5, Special Issue 2, March 2016*

National Conference on Future Technologies in Power, Control and Communication Systems (NFTPCOS-16)

on 10, 11 and 12th March 2016

Organised by

Dept. of EEE, College of Engineering Perumon, Kollam, Kerala – 691601, India

of the video data. Due to the lack of photometric invariance of the intensity channel, current approaches are consequently sensitive to disturbing illumination conditions such as shadows and highlights. More importantly, discriminative information is ignored by discarding chromaticity from the representation.

In the spatial (non-temporal) domain, color descriptors outperform intensity descriptors in a variety of image matching and object recognition tasks. The reason for this improved balance between photometric invariance and discriminative power is illustrated by an estimate of the joint distribution of spatial intensity and color partial derivatives, being the image features based on which descriptors are formed. Thus, information is lost when either intensity or chromatic representations are considered in isolation.

For effective feature detection and extraction based on multi-channel differential representations in the spatio-temporal domain, it is thus a precondition that similar conclusions hold for the joint distribution of temporal intensity and color derivatives. This is verified by observing in which the joint distribution of temporal color and intensity derivatives is shown to strongly resemble the distribution of spatial derivatives, these applications demand recognition systems to operate in unconstrained scenarios. For this reason, research has shifted from recognizing simple human actions under controlled conditions to more complex activities and events 'in the wild'. This requires the methods to be robust against disturbing effects of illumination, occlusion, viewpoint, camera motion, compression and frame rates. High-level approaches for unconstrained human activityrecognition aim at modeling image sequences based on the detection of high level concepts and may build on low-level building blocks which typically consider generic video representations based on local photometric features.

II. RELATED WORKS

The recognition of realistic human actions in videos based on spatio-temporal interest points (STIPs). Existing STIP-based action recognition approaches operate on intensity representations of the image data. Because of this, these approaches are sensitive to disturbing photometric phenomena, such as shadows and highlights. In addition, valuable information is neglected by discarding chromaticity from the High-level approaches for unconstrained human activity recognition aim at modeling image sequences based on the detection of high level concepts , and may build on low level building blocks which typically consider generic video representations based on local photometric features High-level approaches are based on complex, computationally expensive video processing operations but may be superior to low-level approaches in terms of recognition rates. However, high-level approaches are sensitive to local geometric disturbances such as occlusion, which limits their applicability. Low-level approaches are conceptually simple, relatively easy to implement and potentially sparse and efficient. Due to the local nature of features on which low-level approaches are based, they are inherently robust against recording disturbances such as occlusion and clutter. Therefore, in this paper, we focus on low-level representations for recognizing human actions in video.

A. Harris STIP Detector

Harris STIPs are local maxima of the 3D Harris energy function based on the structure tensor. A multi-channel formulation of the structure of tensor has been developed in which prevents opposing color gradient directions to cancel each other out. Here, we incorporate multiple channels in the spatio-temporal structure tensor. The HOG3D descriptor is formulated as a discretized approximation of the full range of continuous directions of the 3D gradient in the video volume. That is, the unit spheres Centered at the gradient location is approximated by a regular n-sided polyhedron with congruent faces. Tracing the gradient vector along its direction up to intersection with any of the polyhedron faces identifies the dominant quantized direction.

B. Harris- Affine Detector

The Harris affine detector can identify similar regions between images that are related through affine transformations and have different illuminations. These affine-invariant detectors should be capable of identifying similar regions in images taken from different viewpoints that are related by a simple geometric transformation: scaling, rotation and shearing. These detected regions have been called both invariant and covariant. On one hand, the regions are detected invariant of the image transformation but the regions covariantly change with image transformation Do not dwell too much on these two naming conventions; the important thing to understand is that the design of these interest points will make them compatible across images taken from several viewpoints. Other detectors that are affine-invariant include Hessian affine region detector, Maximally stable extremely regions, Kadir–Brady saliency detector, edge-based regions (EBR) and intensity-extrema-based regions (IBR).

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization *Vol. 5, Special Issue 2, March 2016*

National Conference on Future Technologies in Power, Control and Communication Systems (NFTPCOS-16)

on 10, 11 and 12th March 2016

Organised by

Dept. of EEE, College of Engineering Perumon, Kollam, Kerala – 691601, India

The Harris affine detector algorithm iteratively discovers the second-moment matrix that transforms the anisotropic region into a normalized region in which the isotropic measure is sufficiently close to one. The algorithm uses this shape adaptation matrix, U, to transform the image into a normalized reference frame. In this normalized space, the interest points' parameters (spatial location, integration scale and differentiation scale) are refined using methods similar to the Harris-Laplace detector.

The second-moment matrix is computed in this normalized reference frame and should have an isotropic measure close to one at the final iteration. Harris affine region points tend to be small and numerous. Both the Harris-Affine detector and Hessian-Affine consistently identify double the number repeatable points as other affine detectors: ~1000 regions for an 800x640 image. Small regions are less likely to be occluded but have a smaller chance of overlapping neighboring regions. The Harris affine detector responds well to textured scenes in which there are a lot of corner-like parts. However, for some structured scenes, like buildings, the Harris-Affine detector performs very well. This is complementary to MSER that tends to do better with well structured (segmentable) scenes. Overall the Harris affine detector performs very well, but still behind MSER and Hessian-Affine in all cases but blurred images.

Harris-Affine and Hessian-Affine detectors are less accurate than others: their repeatability score increases as the overlap threshold is increased. The detected affine-invariant regions may still differ in their rotation and illumination. Any descriptor that uses these regions must account for the invariance when using the regions for matching or other comparisons.LoG is an acronym standing for Laplacian of Gaussian, DoG is an acronym standing for difference of Gaussians (DoG is an approximation of LoG), and DoH is an acronym standing for determinant of the Hessian. These scale-invariant interest points are all extracted by detecting scale-space extrema of scale-normalized differential expressions, i.e., points in scale-space where the corresponding scale-normalized differential expressions assume local extrema with respect to both space and scale. The interest points obtained from the multi-scale Harris operator with automatic scale selection are invariant to translations, rotations and uniform rescaling in the spatial domain. The images that constitute the input to a computer vision system are, however, also subject to perspective distortions. To obtain an interest point operator that is more robust to perspective transformations, a natural approach is to devise a feature detector that is invariant to affine transformations.

Harris STIP detection algorithm can only be very effective for gray-scale images. It has very poor real-time performance because of the large amount of calculation. The Harris STIP detector is variant to noise and can cause loss and error in interpreting the action. Existing STIP-based action recognition approaches operate on intensity representations of the image data. Because of this, these approaches are sensitive to disturbing photometric phenomena, such as shadows and highlights. In addition, valuable information is neglected by discarding chromaticity from the photometric representation.

The trend in object recognition is toward increasing the number of points, applying several detectors or combining them, or making the interest point distribution as dense as possible. While such a dense sampling approach provides accurate object recognition, they basically shift the task of discarding the non discriminative points to the classifier. With the explosive growth of image and video data sets, clustering and offline training of features become less feasible. By reducing the number of features and working with a predictable number of sparse features, larger image data sets can be processed in less time.

Additionally, a stable number of features leads to a more predictable workload for such tasks. color interest points to obtain a sparse image representation. To reduce the sensitivity to imaging conditions, light-invariant interest points are proposed. To obtain light-invariant points, the quasi-invariant derivatives of the HSI color space are used. For color boosted points, the aim is to exploit color statistics derived from the occurrence probability of colors. This way, color boosted points are obtained through saliency-based feature selection.

C. Local Descriptor

Local features are used for object category recognition and classification. In this paper, the comparison is done for various descriptors, different interest regions and for different matching approaches. The descriptors on real images are evaluated with different geometric and photometric transformations and for different scene types are compared. The comparison of descriptors in this context requires a different evaluation setup. It is unclear how to select a representative set of images for an object category and how to prepare the ground truth since there is no linear transformation relating images within a category. Several descriptors and detectors have been added to the comparison and the data set contains a larger variety of scenes types and transformations. Many different techniques for describing local image regions have been developed. The simplest descriptor is a vector of image pixels. Cross-correlation can

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization *Vol. 5, Special Issue 2, March 2016*

National Conference on Future Technologies in Power, Control and Communication Systems (NFTPCOS-16)

on 10, 11 and 12th March 2016

Organised by

Dept. of EEE, College of Engineering Perumon, Kollam, Kerala – 691601, India

then be used to compute a similarity score between two descriptors. The Gabor transform overcomes these problems, but a large number of Gabor filters is required to capture small changes in frequency and orientation. Gabor filters and wavelets are frequently explored in the context of texture classification. The detectors provide the regions which are used to compute the descriptors. If not stated otherwise, the detection scale determines the size of the region. In this evaluation, they had used five detectors. Harris-Affine differs from Harris-Laplace by the affine adaptation, which is applied to Harris-Laplace regions.

III. SYSTEM OVERVIEW

Color STIPs are multichannel reformulations of STIP detectors and descriptors, for which we consider a number of chromatic and invariant representations derived from the opponent color space. The color STIPs with Gabor array is used to overcome the drawback of the existing Harris-STIP detector. The main merits of the color STIP detectors are, Descriptor dimensionality is reduced by allocating opposing gradient directions to the same orientation bin. This approach is insensitive to disturbing photometric phenomena, such as shadows and highlights.

Low-level action recognition approaches are often based on spatio-temporal interest points (STIPs). Here, image sequences are represented by descriptors that are extracted locally around STIP detections. The descriptors are vector quantized based on a visual vocabulary, and subsequent learning and recognition operates on these quantized descriptors, comprising the well known bag-of-(spatio-temporal)-features framework. The block diagram represents the architecture of the color spatio-temporal interest point detector.

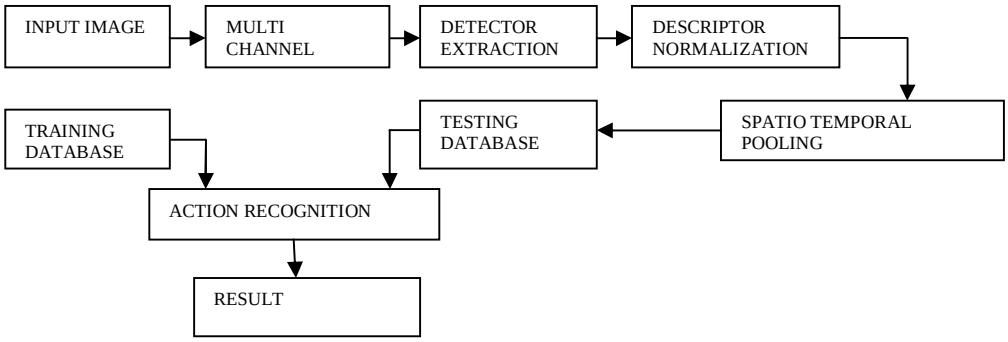

Fig.1. Block diagram of Color Spatio-Temporal Interest Point Detector

The input provided to the system is video streams. The videos are arrangement of images in frames. The frame is composed of picture elements just like a chess board. Each horizontal set of picture elements is known as a line. The picture elements in a line are transmitted as sine signals where a pair of dots, one dark and one light can be represented by a single sine. The product of the number of lines and the number of maximum sine signals per line is known as the total resolution of the frame. The higher the resolution the more faithful the displayed image is to the original image. But higher resolution introduces technical problems and extra cost. So a compromise should be reached in system designs both for satisfactory image quality and affordable price. In moving picture (TV) the number of frames scanned per second is known as the frame rate. The higher the frame rate, the better the sense of motion.. To increase the sense of motion it is customary to scan the very same frame in two consecutive phases. Performance is evaluated in a leave-one-out cross validation scheme, in which the flipped version of the considered test video is removed from the training set. The input image frames are used for the action recognition purpose.

IV. SIMULATION RESULTS

The performance of this detection technique is analyzed by giving a video as the input. The action recognition is performed in MATLAB 2013a. Figure 2.indicates the input video frames. The video is a collection of images. The input image is chosen and video is read by the code. Performance is evaluated in a leave-one-out cross validation scheme, in which the flipped version of the considered test video is removed from the training set.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization *Vol. 5, Special Issue 2, March 2016*

National Conference on Future Technologies in Power, Control and Communication Systems (NFTPCOS-16)

on 10, 11 and 12th March 2016

Organised by

Dept. of EEE, College of Engineering Perumon, Kollam, Kerala – 691601, India

Fig.2. Input video stream

The input image frames are used for the action recognition purpose. The input images frames may have the actions such as walking, jumping, pushups or any other actions. Videos are represented in a variety of color spaces exhibiting different levels of photometric invariance. Dataset consists of 30 videos taken from television series, movies and lab recordings where each video is artificially distorted by applying different types of photometric and geometric transformations. Every transformation type is associated to a challenge, in which the distortion is applied in increasingly severe steps. For evaluating the performance here, the videos from the television series up to the first occurring shot boundary are considered. Here the full set of challenges such as blur, compression, darken, lighten, median filter, noise, sampling rate and scaling and rotation are considered.

The multichannel filter, which is perfectly suitable for real time implementation, is used to remove impulsive noise and other impairment from color TV signals. There are various sources that can generate impulsive noise, such as man-made phenomena, car ignition systems, industrial machines in the vicinity of the receiver, switching transients in power lines, and various unprotected electric switches. In addition, natural phenomena also generate impulsive noise. Impulsive noise is frequently encountered during the transmission of TV signals through ultra-high frequency, very-high frequency, terrestrial microwave links, and FM satellite links. It is therefore important to develop a digital signal processing technique that can remove such image impairment in real-time and thus, guarantee the quality of service delivered to the consumer. With the advent of the all-digital TV system, such filters can lead to systems with accurate image reproduction fidelity despite any unforeseen transmission developments.

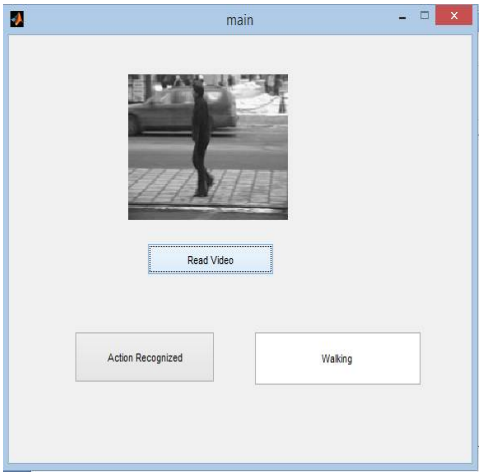

Fig.3. Action Recognition

Automatic text detection in video is an important task for efficient and accurate indexing and retrieval of multimedia data such as events identification, events boundary identification etc. The STIP feature points are to be extracted in order to analyze the action within the video easily. Figure 3 represents the action recognition. The computation of the feature vector of each movie relies on computing the SFA feature vectors of every spatio-temporal chunk in the movie. In our initial setting, the spatial dimensions were 10×10 px, and each frame of the chunk was converted into a 32-d.The success of an action recognition algorithm is dependent on how actions are represented and what classification method is used. There are many action representations, such as interest-point models, local motion

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

An ISO 3297: 2007 Certified Organization *Vol. 5, Special Issue 2, March 2016*

National Conference on Future Technologies in Power, Control and Communication Systems (NFTPCOS-16)

on 10, 11 and 12th March 2016

Organised by

Dept. of EEE, College of Engineering Perumon, Kollam, Kerala – 691601, India

models, and silhouette-based models, and many classifiers have been adopted, such as nearest-neighbor, SVM, boosting, and graphical-model-based classifiers. In order to describe actions, here a "bag of dense local features" approach with feature vectors either measuring the shape of a moving object's silhouette or its optical flow. Since such a representation is overly dense, we reduce its dimensionality by computing the empirical covariance matrix that captures the second-order statistics of the collection of feature vectors.

V. CONCLUSION

The proposed system incorporates chromatic representations in the spatio-temporal domain. The reformulation of STIP detection and description for multi-channel video representations is done accurately. Videos are represented in a variety of color spaces exhibiting different levels of photometric invariance. By this enhanced appearance modeling, the quality (robustness and discriminative power) of STIP detectors and descriptors for recognizing human activities in video is increased. A multi-channel formulation of the structure tensor has been developed which prevents opposing color gradient directions to cancel each other out. Here, multiple channels in the spatio-temporal structure tensor are incorporated. The Gabor STIP detector is based on a Gabor filtering procedure along the temporal axis. Invoking multiple channels is straightforward because the energy function is positive by formulation. Hence, no additional care has to be taken to account for conflicting response signs between channels.

REFERENCES

[1] M. Brown, G. Hua, and S. Winder, "Discriminative learning of local image descriptors," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 1, pp. 43–57, Jan. 2010.
[2] G. J. Burghouts and J. M. Geusebroek, "Performance evaluation of local colour invariants," Comput. Vis. Image Understand., vol. 113, no. 1, pp. 48–62, Jan. 2009.
[3] M. Chen and A. Hauptmann, "Mosift: Recognizing human actions in surveillance videos," Ph.D. dissertation, School Comput. Sci., Carnegie Mellon Univ. Pittsburgh, PA, USA, 2009.
[4] P. Dollàr, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio- temporal features," in Proc. 2nd Joint IEEE Int. Workshop VSPETS, Oct. 2005, pp. 65–72.
[5] I. Everts, J. C. van Gemert, and T. Gevers, "Per-patch descriptor selection using surface and scene attributes," in Proc. 12th ECCV, 2012, pp. 172–186.
[6] I. Everts, J. C. van Gemert, and T. Gevers, "Evaluation of color STIPs for human action recognition," in Proc. IEEE Conf. CVPR, Jun. 2013, pp. 2850–2857.
[7] A. Kläser, M. Marszalek, and C. Schmid, "A spatio-temporal descriptor based on 3d-gradients," in Proc. 19th BMVC, 2008, pp. 275–285.
[8] I. Laptev, "On space-time interest points," Int. J. Comput. Vis., vol. 64, nos. 2–3, pp. 107–123, 2005.
[9] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld, "Learning realistic human actions from movies," in Proc. IEEE Conf. CVPR, Jun. 2008, pp. 1–8.
[10] J. Liu, J. Luo, and M. Shah, "Recognizing realistic actions from videos 'in the wild'," in Proc. IEEE Int. Conf. CVPR, Jun. 2009, pp. 1996–2003.