www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 2 Issue 12 Dec,2013 Page No. 3472-3476

Color STIPs for the Live Feed

V. Lakshmi Chetana[1], D. Prasad[2],G.Krishna Chaitanya[3]

[1]Department of Computer Science & Engineering,
DVR & Dr.HS MIC College of Technology, Kanchikacherla.
mailtochetana@gmail.com

[2]Assoc. Prof. , Department of Computer Science & Engineering,
DVR & Dr.HS MIC College of Technology, Kanchikacherla.
dprasad.vza@gmail.com
[3]Asst.Prof.,,Department of Computer Science &Engineering,
NRI Institute of Technology
chaitu.gkchaitanya@gmail.com

ABSTRACT:*Interest point detection is the vital aspect of computer vision and image processing. Interest point detection is very important for image retrieval and object categorization systems from which local descriptors are calculated for image matching schemes. Earlier approaches largely ignored color aspect as interest point calculation are largely based on luminance. However the use of color increases the distinctiveness of interest points. Subsequently an approach that uses saliency-based feature selection aided with a principle component analysis-based scale selection method was developed that happens to be a light-invariant interest points detection system. This paper introduces color interest points for sparse image representation. In the domain of video-indexing framework, interest points provides more useful information when compared to static images. Since human perception system is naturally influenced by motion of objects, we propose to extend the above approach for dynamic video streams using Space-Time Interest Points (STIP) method. The method includes the process of calculating the interest points in 3D domain (i.e., for feature extraction, this means that the main 2D concepts for images are extended to 3D). STIP renders moving objects in a live feed and characterizes the specific changes in the movement of these objects. A practical implementation of the proposed system validates our claim to support live video feeds and further it can be used in domains such as Motion Tracking, Entity Detection and Naming applications that have abundance importance.*

KEYWORDS : Interest Points, Image Retrieval, STIP, LoG, Object Categorization, Image Object Detection, Event Detection, Motion Analysis.

1.Introduction

Interest point detection is a vital research area in the field of computer vision and image processing. In particular image retrieval and object categorization heavily depend on interest point detection from which local image descriptors are calculated for image and object matching[1]. Majority of interest point extraction algorithms are purely based on luminance.These methods ignore the important information contained in the color channels. And hence interest point extraction based on luminance is not more effective. So color interest points are introduced.

Color plays a very important role in the pre-attentive stages of feature detection. Color provides extra information which allows the distinctiveness between various reasons of color variations, such as change due to shadows, light source reflections and object reflectance variations.[2] The detection and classification of local structures(i.e.edges, corners & T-junctions) in color images is important for many applications such as image segmentation, image matching, object recognition, visual tracking in the fields of image processing and computer vision[3],[4],[5]. Earlier approach of feature detectors uses the dense sampling representation of image in which the redundant information is also considered while extracting the features. In this paper we use a sparse representation of the image. The current trend in object recognition is increasing the number of points [6] by applying several detectors or by combining them [7][8] or making the interest point distribution as dense as possible[9].While such a dense sampling approach provides accurate object recognition, they basically change the task of discarding the non-discriminative points to the classifier. The extreme growth of image and video data sets, clustering and offline training of features become less feasible[10]. By reducing the number of features and working with expected number of sparse features, larger image data sets can be processed in less time Now the goal is to reduce the number of interest points extracted for obtaining better image retrieval or object recognition results.

Recent work has aimed to find unique features, i.e., by

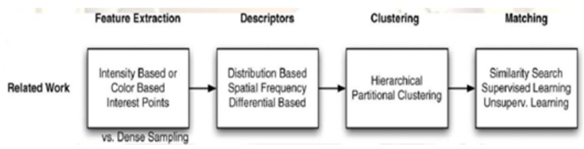

Figure 1: Main steps of Image Retrieval and Object Categorization

performing an evaluation of all features within the dataset or per image class and choosing the most frequent ones[11].For this, the best option is to use color to increase the distinctiveness of interest points[12][13]. This may provide selective search for robust features reducing the total number of interest points used for image retrieval.

Till now we discussed the way how the interest points are calculated in spatial domain. In this paper ,we propose color space –time interest points(STIP)(i.e., detecting the color interest points in spatio temporal domain). Space-time interest points are the points which are relevant both in space and time. STIP detectors are the extensions of 2D interest point detectors that incorporate temporal information.

2. Common Pipeline for Image Retrieval and Object Categorization

Fig. 1 illustrates the common pipeline for image retrieval and object categorization

Feature extraction is carried out with either global or local features. In general, global features lack robustness against occlusions and cluttering and provide a fast and efficient way of image representation. Local features are either intensity or color -based interest points. Earlier, dense sampling of local features has been used as it provides good performance, but the only problem is requires more number of interest points for feature extraction.[20].

Descriptors represent the local image information around the interest points. They can be categorized into three classes: They describe the distribution of certain local properties of the image [e.g., scale- invariant feature transform (SIFT)], spatial frequency (e.g., wavelets), or other differentials (e.g., local jets) [13] .For every feature extracted a descriptor is computed. A disadvantage is that the runtime increases with their number. Efficient ways to calculate these descriptors exist, e.g., for features with overlapping areas of support, previously calculated results can be used. [20]

Clustering for signature generation, feature generalization,or vocabulary estimation assigns the descriptors into a subset of categories. There are hierarchical and partitional approaches to clustering. Due to the excessive memory and runtime requirements of hierarchical clustering [14], partitional clustering such as k-means is the method of choice in creating feature signatures.

Matching summarizes the classification of images. Image descriptors are compared with previously learnt and stored models. This is computed by a similarity search or by building a model based on supervised or unsupervised learning techniques. Classification approaches need feature selection to discard irrelevant and redundant information [15] – [17]. It is shown that a powerful matching step can successfully discard irrelevant information, and better performance is gained [9]. Training and clustering are the most time-consuming steps in state-of-the-art recognition frameworks.[20]The important part in image retrieval and object categorization is feature extraction. So this paper mainly focuses on extracting the features in images. We extend the idea of interest points into spatio-temporal domain i.e., to the dynamic feed (video).The next sections describe about the color interest points both in spatial and spatio-temporal domain.

3. Color Interest points in Spatial Domain

Interest point is a recent terminology in computer vision that refers to the detection of interest points for subsequent processing .An interest point is a point in the image which in general can be categorized as

- Has a well defined position in the image space
- Has a clear mathematically well-defined definition.
- It is stable under local and global perturbations in the image domain.
- The notion of interest point has an attribute of scale.

Earlier approaches used intensity based interest point detectors. In this method, the given image is converted into gray-scale image and then processing is performed which is a time consuming and not accurate. Dense sampling approach is used to find out the interest points which results in, more number of interest points are needed to identify the object in the image and low intensity points are ignored.

To overcome these problems, color interest points are introduced. Color plays a very important role in the pre-attentive stages of feature detection. Color provides extra information which allows the distinctiveness between various reasons of color variations, such as change due to shadows, light source reflections and object reflectance variations. Harris –Laplacian detector is used to find out the spatial interest points. It is defined as follows.

The Harris-Laplace interest point detector locates the interest points in two steps. A multi-scale point detection is performed in the first step, the second step is an iterative selection of the scale and the localization. The scale-adapted second moment matrix is defined as

$$M(x,\sigma) = \sigma_D^2 g(\sigma_I)\begin{bmatrix} I_x^2(x,\sigma_D) & I_xI_y(x,\sigma_D) \\ I_xI_y(x,\sigma_D) & I_y^2(x,\sigma_D)\end{bmatrix}$$

Where σ_I is the integration scale , σ_D is the differentiation scale and a I_x and I_y are the derivatives computed in x and y direction.M is the auto-correlation matrix.

A scale-space representation with the Harris function for pre-selected scales is built at first. The interest points are extracted at each level of the scale-space representation by detecting the local maximum in the 8-neighborhood of a point x by means of the equation

$$R = \det M - k(trace\ M)^2$$

where det M = $\lambda_1 * \lambda_2$ and trace M = $\lambda_1 + \lambda_2$ and k is a constant. The typical value of the constant k is equal to 0.04.A threshold is used to reject the maxima of small cornerness, as they are less stable under variations in imaging conditions. After a multi-scale Harris corner detection, the characteristic scale is determined. The selection of the characteristic scale is based on scale selection as proposed in [18]-[19].

4. Color Interest points in Spatio –Temporal domain

Space-Time Interest Points(STIPs) are the interest points which are interested both in spatial and temporal domains. These points are especially interesting because the focus information initially contained in thousands of pixels on a few specific points which can be related to spatio-temporal events in the sequence. The STIPs appear in articulated motions only i.e..like walking running, jumping, jogging boxing,waving etc..STIP detectors are extensions of 2D interest points detectors that incorporate temporal information. STIP are based on spatio-temporal corners. Constant motion of a corner does not produce any STIPs. Spatio-temporal corners are located in region that exhibits a high variation of image intensity in all three directions(x,y,t). STIPs detection is performed by using the second moment matrix defined for a pixel(x,y) at time *t* having intensity *I*(x,y,*t*) by

$$M(x,y,t) = \begin{pmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} & \frac{\partial^2 I}{\partial x \partial t} \\ \frac{\partial^2 I}{\partial x \partial y} & \frac{\partial^2 I}{\partial y^2} & \frac{\partial^2 I}{\partial y \partial t} \\ \frac{\partial^2 I}{\partial x \partial t} & \frac{\partial^2 I}{\partial y \partial t} & \frac{\partial^2 I}{\partial t^2} \end{pmatrix}$$

STIPs are identified from local maxima of a cornerness function computed for all pixels across spatial and temporal scales.The extension of the Harris corner function called saliency function is defined as

$$R(x,y,t) = \det(M(x,y,t) - k(trace\, M(x,y,t))^3$$

where k is a parameter empirically adjusted. STIP correspond to high values of the above mentioned saliency function.[21]

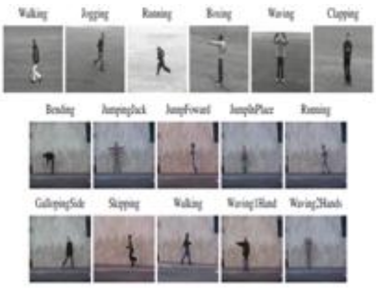

Figure 2 : Different articulated motions

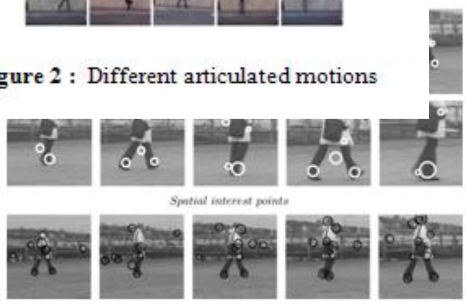

Figure 3 : Top : Spatio - temporal interest points. Bottom : Spatial interest points

5.Parameters for evaluating STIPs performance

The evaluation of spatio - temporal features mainly depend on various factors like

- **Gaussian blur :** By increasing the Gaussian blur per channel we calculate the number of interest points obtained.
- **Noise :** Random values are added to the video and then we observe the number of IPs .
- **Change of lightening :** The videos are darkened and lightened by changing the lightness of the colors to simulate increasing and decreasing lighting conditions.
- **Scale and Rotation :** To evaluate the videos invariance to scaling and rotation, these values are adjusted at different levels and observed .
- **Frames per Second :** To decrease the demand for storage space,surveillance videos are often handled with very few fps.[22]

In this paper , the work is compared with already existing spatio-temporal interest point algorithms like Cuboid, Harris 3D and Hessian 3D. From Table 1 Harris 3D and Hessian 3D are invariant to scale. Hessian 3D uses dense sampling approach where as the remaining uses sparse data for detection of interest points.The existing space-time interest point detectors are all luminance based and this paper provides color space-time interest points.There are many parameters for evaluation , but in this paper the complete results are compared using the contrast parameter.

Detector	Scale Selection	Feature Set	Interest Points based on
Laptev(Harris 3D)	Yes	Sparse(rare)	Luminance
Geert Willems(Hessian 3D)	Yes	Dense	Luminance
Dollar(Cuboid 3D)	No	Sparse(rare)	Luminance
Proposed Method	No	Sparse(rare)	Color

Table 1 : Comparison among STIP detectors

6. Results

An analysis is done on the performance of STIPs at various contrasts. The evaluation is performed by observing the

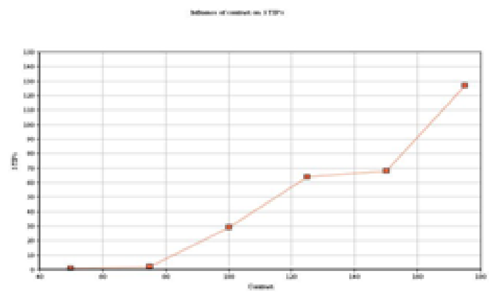

Figure 4 : Graph showing the no.of STIPs for different contrasts

variations of the number of STIPs compared with the initial state i.e during no contrast modification to slight variations of contrast. Fig 4 shows the results

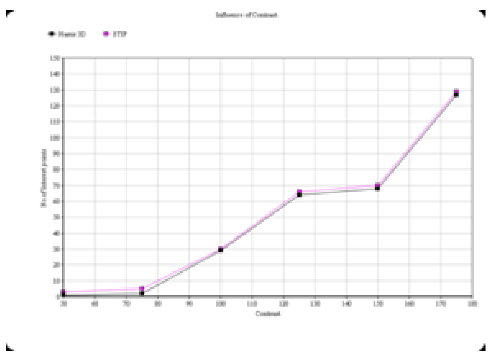

Figure 5: Influence of contrast on Harris 3D and proposed method

Fig.5. represents the graph for comparing the proposed method with the Harris 3D. The proposed method is better than Harris 3D at the contrast levels 120 to 150.

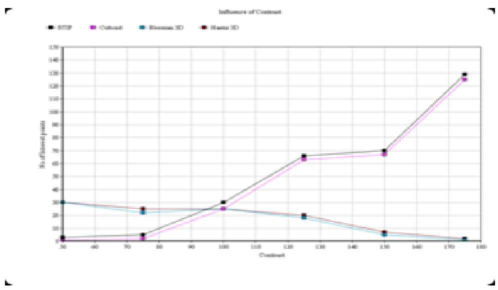

Figure 6 : Comparison of STIPs generated by various algorithms

Fig.6 shows the results obtained from various interest point detection algorithms and from the chart the proposed method suits to be best color STIP detector at various contrast levels.

7. Conclusion

In this paper, an approach has been proposed to extract the light invariant interest points based on color for static images. we have extended this interest points to dynamic content that is for videos which play a very important role for Object or Scene identification and retrieval and Motion Capturing. We have described an interest point detector that finds local image features in space –time characterized by a high variation of the image values in space and varying motion over time. The localization of the interest points of each frame is identified by the Harris Detector and Laplacian operator is applied to these points to find scale-invariant points in space-time. The combination of these two operators is combined as Space – Time Interest Points.

In future work, this application can be extended to the field of motion-based recognition and event detection. Generally the present motion capturing systems(Eg :CC cameras used in airports, shopping malls etc..;) are high in luminance content .But in gray level images or videos most of the important content is discarded due to their less saliency. So usage of this color STIP's can improve the detection or recognition or identification of objects.

References

[1] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *Proc. CVPR*, 2003, pp. II-264–II-271.

[2] Joost van de Weijer,"Color Features and Local Structure in Images".

[3] Cordelia Schmid, Roger Mohr and Christian Bauckhage ,"Evaluating of interest points detectors", published in "International Journal of Computer Vision 37, 2 (2000) 151--172" .

[4] Jianbo Shi and Carlo Tomasi,"Good features to track",research report submiited in December 2005.

[5] SILVANO DI ZENZO ,"A note on the gradient of a multi-image", COMPUTER VISION, GRAPHICS, AND IMAGE PROCESSING 33, 116-125 (1986)

[6] J. Zhang, M. Marszałek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A com-prehensive study," *Int. J. Comput. Vis.*, vol. 73, no. 2, pp. 213– 238, Jun. 2007.

[7] K. Mikolajczyk, B. Leibe, and B. Schiele, "Multiple object class de-tection with a generative model," in *Proc. CVPR*, 2006, pp. 26–36.

[8] J. Sivic, B. Russell, A. A. Efros, A. Zisserman, and B. Freeman, "Discovering objects and their location in images," in *Proc. ICCV*, 2005, pp. 370–377.

[9] T. Tuytelaars and C. Schmid, "Vector quantizing feature space with a regular lattice," in *Proc. ICCV*, 2007, pp. 1–8.

[10] G. Schindler, M. Brown, and R. Szeliski, "City-scale location recognition," in *Proc. CVPR*, 2007, pp. 1–7.

[11] P. Turcot and D. G. Lowe, "Better matching with fewer features: The selection of useful features in large database recognition problems," in *Proc. ICCV Workshop*, 2009, pp. 2109–2116.

[12] J. van de Weijer and T. Gevers, "Edge and corner detection by photo metric quasi-invariants," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 4, pp. 625–630, Apr. 2005.

[13] Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615– 1630, Oct. 2005.

[14] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264– 323, Sep. 1999

[15] R. Okada and S. Soatto, "Relevant feature selection for human pose estimation and localization in cluttered images," in *Proc. ECCV*, 2008, pp. 434–445.

[16] G. Dorko and C. Schmid, "Selection of scale-invariant parts for object class recognition," in *Proc. ICCV*, 2003, pp. 634–639.

[17] F. Jurie and B. Triggs, "Creating efficient codebooks for visual recognition," in *Proc. ICCV*, 2005, pp. 604–610.

[18] T. Lindeberg, "Feature detection with automatic scale selection," Inter-national Journal of Computer Vision, vol. 30, pp. 79–116, 1998.

[19] Martin Zukal,Petr Cika and Radim Burger,"Evaluation of interest point detectors for scenes with changing lightening conditions".

[20] Julian Stottinger,Allan Hanbury,Nicu Sebe,"Sparse Interest points for Image retrieval and Object Categorization"IEEE Transactions .Image Processing,Vol.21,No.5,May 2012.

[21] Alain Simac-Lejeune," Moving Object Analysis In Video Sequences Using Space-Time Interest Points" published in "VISIGRAPP 2012 7th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - VISAPP, Roma, Italy(2012)

[22] Julian Stottinger " Detection and Evaluation Methods for Local Image and Video Features" Technical report CVL-TR-4 ,Mar 2011.

Author Profile

Devarasetty Prasad, completed B.Tech in Computer Science Engineering.He obtained his M.Tech in Software Engineering.He is currently working as Associate Professor in the Dept of CSE in DVR & Dr.HS MIC College of Technology,Kanchikacherla,Andhra Pradesh. He has 10 years of experience in teaching.Presently, he is doing Ph.D under Andhra University,Visakapatnam in the domain of Software Engineering.His areas of interest in research are Software Engineering,Metrics,Open Source Software and Software maintainance.

Lakshmi Chetana Vemuri,persuing her M.Tech in Computer Science and Engineering in DVR & Dr.HS MIC College of Technology.She received her MCA from Hindu college under Acharya Nagarjuna University.Her areas of interest in research are Image Processing,Computer Vision, and Distributed Systems.