IJDCST @November Issue- V-1, I-7, SW-50
ISSN-2320-7884 (Online)
ISSN-2321-0257 (Print)

Color Interest Points for Dynamic Stream Object Categorization

[1] Talari Keerthi , [2] CH. Chandra Mohan
[1] Student, PVPSIT, KANURU, VIJAYAWADA, KRISHNA DIST.
[2] Assistant Prof, PVPSIT,KANURU,VIJAYAWADA,KRISHNA DIST.

Abstract: Detection of interest points for subsequent processing is one of the vital aspects of computer vision. Object categorization of images heavily relies on interest point detection from which local image descriptors are computed for image matching. Since interest points are based on luminance, previous approaches largely ignored the color aspect. Later an approach that uses saliency-based feature selection optimized by a principal component analysis-based scale selection method is developed. It is used to reduce the sensitivity to varying imaging conditions, and thus it is a light-invariant interest point's detection system. Use of color increases the distinctiveness of interest points. In the context of object recognition, the human perception system is naturally attracted by differences between parts of images and by motion or moving objects. Therefore, in the video indexing framework, interest points provide more useful information when compared to static images. So we propose to extend the above approach for dynamic video streams using Space-Time Interest Points (STIPs) that uses an algorithm for scale adaption of spatio-temporal interest points. STIP detects moving objects in videos and characterizes some specific changes in the movement of these objects. A practical implementation of the proposed system validates our claim to support dynamic streams and further it can be used in applications such as Motion Tracking, Entity Detection and Naming applications.

I. INTRODUCTION

The recognition of texture and object categories is one of the most challenging problems in computer vision. Representation, detection and learning are the main issues that need to be tackled in designing a visual system for recognizing object categories. Interest point detection is an important research area in the field of image processing and computer vision. Image retrieval and object categorization heavily rely on interest point detection from which local image descriptors are computed for image and object matching. Color plays an important role in the pre-attentive stage in which features are detected as it is one of the elementary stimulus features. It is customary to define texture as a visual pattern characterized by the repetition of a few basic primitives. There is broad agreement on the issue of representation: object categories are represented as collection of features, each part has a distinctive appearance and spatial position.

The current 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. 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.

A stable number of features lead to a more predictable workload for such tasks. Recent work has aimed to find distinctive features by performing an evaluation of all features within the data set or per image class and choosing the most frequent ones. This approach requires an additional calculation step with an inherent demand on memory and processing time dependent on the number of features. This alternative may therefore provide selective search for robust features reducing the total number of interest points used for image retrieval. We propose color interest points to obtain a sparse image representation. Hence, we reduce the sensitivity to imaging conditions, light-invariant interest points are proposed. For color boosted points, the aim is to exploit color statistics derived from the occurrence probability of colors. Color boosted points are obtained through saliency-based feature selection. The use of color information allows extracting repeatable and scale-invariant interest points.

Color derivatives were taken to form the basis of a color saliency boosting function to equal the information content and saliency of a given color occurrence. Our aim is to select interest points based on color discriminative and invariant properties derived from local neighborhoods. Our focus is on color models that have useful perceptual and invariant properties to achieve a reduction in the number of interest points. A method of selecting a scale associated with the computed interest points while maintaining the properties of the color space used and to steer the characteristic scale by the saliency of the surrounded structure.

II. RELATED WORK

The main steps of image retrieval and object categorization are outlined.

Common Pipeline for Image Retrieval and Object Categorization:

Feature extraction is carried out with either global or local features. Global features lack robustness against occlusions and cluttering and provide a fast and efficient way of image representation. The local features are either intensity- or color-based interest points. Dense sampling of local features has been used as it provides good performance. Descriptors represent the local image information around the interest points. This can be categorized in to three classes:

- Distribution of certain local properties of the image
  Ex: Scale-invariant feature transform
- Spatial frequency
  Ex: wavelets
- Other differentials
  Ex: local jets

Efficient ways to calculate these descriptors exist previously calculated results can be used.

Clustering for signature generation or vocabulary estimation assigns the descriptors into a subset of categories. Due to the excessive memory and runtime

requirements of hierarchical clustering, partitioned clustering is the method of choice in creating feature signatures. Image descriptors are compared with previously learnt and stored models. Classification approaches need feature selection to discard irrelevant and redundant information. It is shown that a powerful matching step can successfully discard irrelevant information and better performance is gained. Clustering of a global dictionary takes several days for current benchmark image databases. The use of color provides selective search reducing the total number of interest points used for image retrieval. An extension of the Harris corner detector is proposed by Mikolajczyk and Schmid. The main idea is to carry out corner and blob detection on different scales. The precision of the scale estimation using either Laplacian of Gaussian methods depends on the choice of the scale sampling rate. Maximally stable extremum regions (MSERs) are obtained by a watershedlike algorithm. The algorithm is very efficient in runtime, performance, and detection rate and is extended to color. Extract scale- and illumination-invariant blobs through color by an adapted illumination model and a modification of the LoG. It is efficiently approximated by multiplying the LoG functions' output per channel but is of limited robustness. The most successful color features are based on the color Harris detector and successfully used in many examples. In the scenario of the image retrival they apply the fixed scale detector on gradually downsized images and use all the detections extracted. A color Gaussian pyramid is used to leads to multiple ambiguous features and the inability to match images at different scales. The method is independent of the color space used. The derivatives of the invariants are incorporated in the Harris second moment matrix. It uses fixed scales for matching of images under varying lighting. A photometric quasi-invariant HIS color space providing a corner detector with better noise stability characteristics compared to existing photometric invariants and a color boosting hypothesis for defining salient colors. Our contribution is to extend this approach by incorporating a scale selection strategy to detect color interest points.

III. APPROACH

An object model consists of a number of parts. Each part has an appearance, relative scale and can be occluded or not. The shape is represented by the mutual position of the parts. Entire model is generative and probabilistic shape and occlusions are all modeled by probability density functions. The process of learning an object category is one of first detecting regions and their scales and then estimating the parameters of the above densities from these regions.

Recognition is performed on a query image by again first detecting regions and their scales and then evaluating the regions in a Bayesian manner. Features are found using the detector of Kadir and Brady. This method finds regions that are salient over both location and scale. Each point on the image a histogram is made of the intensities in a circular region of radius (scale). The entropy of this histogram is then calculated and the local maxima are candidate scales for the region. The *N* regions with highest saliency over the image provide the features for learning and recognition. Good example illustrating the saliency principle is that of a bright circle on a dark background. The scale is too small

then only the white circle is seen and there is no extrema in entropy. In practice this method gives stable identification of features over a variety of sizes and copes well with intra-class variability. The measure is designed to be invariant to scaling, although experimental tests show that this is not entirely the case due to aliasing and other effects.

The feature detector identifies regions of interest on each image. Once the regions are identified, they are cropped from the image and rescaled to the size of a small pixel patch. Each patch exists in a 121 dimensional space. We must somehow reduce the dimensionality of each patch whilst retaining its distinctiveness. A 121-dimensional Gaussian is unmanageable from a numerical point of view and also the number of parameters involved is too many to be estimated. In the learning stage, we collect the patches from all images and perform PCA on them. Patch's appearance is then a vector of the coordinates within the first principal components. This gives a good reconstruction of the original patch whilst using a moderate number of parameters per part.

Learning is carried out using the expectation maximization (EM) algorithm which iteratively converges from some random initial value of θ to a maximum. The scale information from each feature allows us to learn the model shape in a scale-invariant space. Learning complex models such as these has certain difficulties. Surprisingly, we assume given the complexity of the search space, the algorithm is remarkable consistent in its convergence. Initial conditions were chosen randomly within a sensible range and convergence usually occurred within 50-100 EM iterations. Estimating this from foreground data proved inaccurate so the parameters were estimated from a set of background images and not updated within the EM iteration.

Recognition proceeds by first detecting features and then evaluating these features using the learnt model. By calculating the likelihood ratio and comparing it to a threshold, the presence or absence of the object within the image may be determined. As in learning efficient search techniques are used since large mean around 2-3 seconds are taken per image.

IV. COMPONENTS OF REPRESENTATION

We first discuss scale- and affine-invariant local regions and the descriptors of their appearance, and then describe different image signatures and similarity measures suitable for comparing them. We use two complementary local region detector types to extract salient image structures:

- The *Harris-Laplace* detector
  It responds to corner-like regions
- The *Laplacian* detector
  It extracts blob-like regions

These two detectors are invariant to scale transformations alone as shown in the fig.1. We can either use rotationally invariant descriptors to achieve rotation invariance. The dominant gradient orientation is computed as the average of all gradient orientations in the region. We obtain affine-invariant versions of the Harris-Laplace and Laplacian detectors through the use of an *affine adaptation* procedure. Normalization leaves a rotational ambiguity that can be eliminated either by using rotation-invariant descriptors or by finding the dominant gradient orientation. The normalized

circular patches obtained by the detectors described in the previous section serve as domains of support for computing appearance-based descriptors. We use three different descriptors:

a. SIFT

It has been shown to outperform a set of existing descriptors

b. STIP

It based on *STIP images* used for matching range data

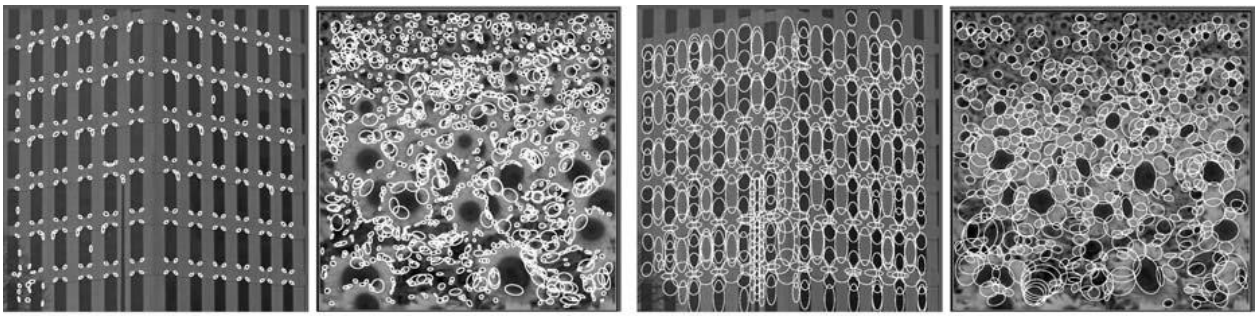

Figure 1: Illustration of affine Harris and Laplacian regions on two natural images

c. RIFT

It is a rotation-invariant version of SIFT

After detecting salient local regions and computing their descriptors, we need to represent their distributions in the training and test images. Method for doing this is to cluster the set of descriptors found in each image to form its *signature. Earth Mover's Distance* (EMD) has shown to be very suitable for measuring the similarity between image signatures. An alternative to image signatures is to obtain a global *texton vocabulary* by clustering descriptors from a special training set and then to represent each image in the database as a histogram of texton labels.

V. RESULT ANALYSIS

Experiments were carried out as follows: each dataset was split randomly into two separate sets of equal size. The decision was a simple object present/absent one, except for the cars dataset where multiple instances of the object were to be found. A limited amount of preprocessing was performed on some of the datasets. The spotted cat dataset was only 100 images originally and another 100 were added by reflecting the original images making 200 in total. There were two phases of experiments. Datasets with scale variability were normalized so that the objects were of uniform size. Algorithm was then evaluated on the datasets and compared to other approaches. The algorithm was run on the datasets containing scale variation and the performance compared to the scale-normalized case. The only parameter that was adjusted at all in all the following experiments was the scale over which features were found. The face and motorbike datasets have tight shape models but some of the parts have a highly variable appearance. These parts any feature in that location will do

regardless of what it looks like. The majority of errors are a result of the object receiving insufficient coverage from the feature detector. One possibility is that the threshold is imposed on *N* many features on the object are removed. The feature detector seems to perform badly when the object is much darker than the background. The clustering of salient points into features within the feature detector. A recall-precision curve1 (RPC) and a table comparing the algorithm to previous approaches to object class recognition as shown in the fig.2.

Dataset	Ours	Others	Ref.
Motorbikes	92.5	84	[17]
Faces	96.4	94	[19]
Airplanes	90.2	68	[17]
Cars(Side)	88.5	79	[1]

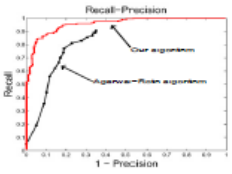

Figure 2: Comparison to other methods

The diagram on the right shows the RPC for and our algorithm on the cars dataset and on left the table gives ROC equal error rates on a number of datasets. The fig.3 discuss about the 6 typical face models. The top left figure shows the shape model. Ellipses represent the variance of each part and the probability of each part being present is shown just to the right of the mean. Top right figure shows 10 patches closest to the mean of the appearance density for each part and the background density. Along with the determinant of the variance matrix, so as to give an idea as to the relative tightness of each distribution. The pink dots are features found on each image and the coloured circles indicate the features of the best hypothesis in the image.

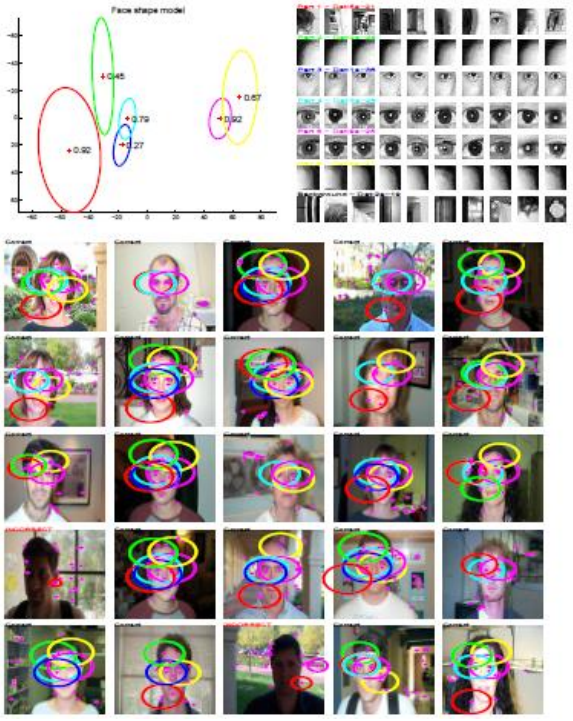

Figure 3: A typical face model with 6 parts

Size of the circles indicates the score of the hypothesis. Exactly the same algorithm settings are used for next consider example. As we consider the typical airplane with 6 parts.

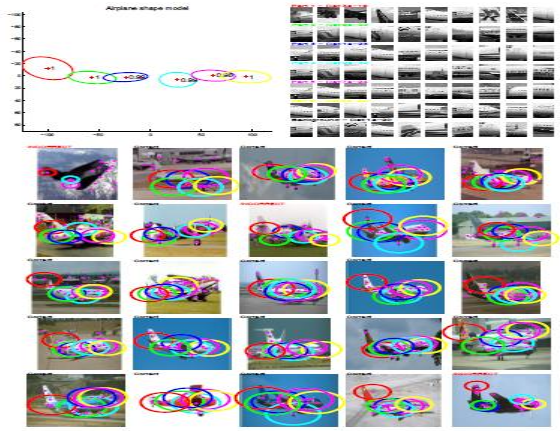

Figure 4: A typical airplane model with 6 parts

The table 1 can create a table for number of datasets which creates confusion in the identification.

Dataset	Total size of dataset	Object width	Face model	Airplane model
Faces	435	300	96.4	32
Airplanes	800	300	63	90.2

Table 1: A confusion table for a number of datasets

The diagonal shows the ROC equal error rates on test data across four categories that the algorithm's parameters were kept *exactly* the same. The performance above can be improved dramatically (airplanes to 94.0% and faces to 96.8%) if feature scale is adjusted on a per-dataset basis. The table shows the performance of the algorithm across the four datasets. The table shows the confusion between models which is usually at the level of chance. The performance of the scale-invariant models on unscaled images to that of the scale-variant models on the pre-scaled data. There is a significant improvement in performance with the scale invariant model. Due to the feature detector performing badly on small images and in the pre-scaled case. This dataset was tested against background road scenes than the background images.

VI. CONCLUSION

Interest point detection is an important research area in the field of image processing and computer vision. Its usage can be found in the facial recognition, motion detection, license plate detection applications. To reduce the sensitivity to imaging conditions, light-invariant interest points are proposed. Feature selection takes place at the very first step of feature extraction and is carried out independently per feature. Prior approaches using HSI, PCA combination schemes concentrated on implementing interest points classification for object categorization in colored images. We propose to extend the approach for dynamic video streams using Space-Time Interest Points (STIPs) that uses an algorithm for scale adaption of spatio-temporal interest points.

A practical implementation of the proposed system validates our claim to support dynamic streams and further it can be used in applications such as Motion Tracking, Entity Detection and Naming applications

VII. REFERENCES

[1] Julian Stöttinger, Allan Hanbury, Nicu Sebe and Theo Gevers . "Sparse Color Interest Points for Image Retrieval and Object Categorization". In IEEE Transactions on Image Processing, vol. 21, no. 5, may 2012.

[2] S. Agarwal and D. Roth. Learning a sparse representation for object detection. In *Proc. ECCV*, pages 113–130, 2002.

[3] Y. Amit and D. Geman. A computational model for visual selection. *Neural Computation*, 11(7):1691–1715, 1999.

[4] E. Borenstein. and S. Ullman. Class-specific, top-down segmentation. In *Proc. ECCV*, pages 109–124, 2002.

[5] M. Burl, M.Weber, and P. Perona. A probabilistic approach to object recognition using local photometry and global geometry. In *Proc.*
ECCV, pages 628–641, 1998.

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

[7] C. Harris and M. Stephens, "A combined corner and edge detection," in *Proc. 4th Alvey Vis. Conf.*, 1988, pp. 147–151.

[8] T. Kadir and M. Brady, "Saliency, scale and image description," *Int. J. Comput. Vis.*, vol. 45, no. 2, pp. 83–105, Nov. 2001.

[9] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *Int. J. Comput. Vis.*, vol. 60, no. 1, pp. 63–86, Oct. 2004.

About Author:

TALARI KEERTHI received the **B.Tech** (IT) Degree from VVIT, Nambur, and Guntur Dist. and affiliated to JNTU Kakinada University, India. She is currently pursuing M.Tech (CSE) Degree at the Dept of Computer Science Engineering, PVP Siddhartha Institute of Technology, Vijayawada

Chetla Chandra Mohan received his **B.Tech** (CSE) Degree from BVC Engineering College, odlaravu, East Godavari District and affiliated to JNTU Hyderabad University, India. He received the **M.Tech(**CSE)specialized in(Software Engineering) Degree from GIET Engineering College, Rajahmundry. He has more than 8 years teaching experience. He is currently working as an Assistant Professor in the Dept of Computer Science Engineering, at PVP Siddhartha Institute of Technology, Vijayawada and affiliated to JNTU Kakinada University, India. His interests are Software Engineering, Data Mining.