

Semantically Effective Visual Concept Illustration for Images

K. K. Thyagarajan and G. Nagarajan, *Member, IACSIT*

Abstract—The field of analyzing and understanding high-dimension real world data in a knowledgeable form is known as computer vision. Image is such a high dimension real world data which is enormously available in meaningless way. Using Ontologies in this computer vision field will enable a high end Ontology knowledge based image analysis and understanding. The main objective of this paper is to develop a frame work that creates a Bag of Visual Words for sports events using low level features of images and to develop ontology to provide semantic meaning to the images. Such semantically available images can be easily retrieved using a semantic image retrieval system. We used the concepts of enhanced Scale Invariant Feature Transform (SIFT) for feature extraction and Support Vector Machine based classification model to provide semantic to the images for action classification of images

Index Terms—Computer vision, ontology, bag of visual words, SIFT, SVM and knowledge-based component.

I. INTRODUCTION

In computer vision, image representation and analysis is one of the broad research areas. We need certain techniques through which a system can recognize an image without any human interpretation. This paper explains the state of art technique used to analyze and recognize sport event images through Ontology. Using Ontology one can represent a domain through concepts [1] and its relationship in an organized way.

In this work the low level features of images are analyzed through the concept of SIFT. Through this SIFT concept the feature of the selected key point is identified. Such identified key points are clustered and grouped through the semi-supervised learning technique, Support Vector Machine. Such classified image key points are used to create a test bed called as Bag of Visual Words, which is a kind of dictionary where the image features are indexed hierarchy.

This paper is organized in such a way that , the list of related work in this area has been discussed, then in next session the overall framework of the proposed idea has been given, and then the concept of SIFT, SVM, Bag of visual word and Sports event Ontology are briefed with appropriate results. The final implementation and its future contribution are explained in the conclusion part.

II. RELATED WORK

In [2], the author concentrates on visual or textual query

Manuscript received August 2, 2013; revised October 13, 2013.

K. K. Thyagarajan is with RMD Engineering College, Kavaraipettai, India (e-mail: kkthyagarajan@yahoo.com).

G. Nagarajan is with Research Scholar of Sathyabama University, Chennai, India (e-mail: nagarajanme@yahoo.co.in).

analysis. Here in our work we are fusing both formats of queries as the user can give textual keyword or image as an input query.

In [3] the author takes into account the manual annotation of the images to determine the action of the still image of the players, in our work we use the low level features of the image for action classification for that we use SIFT[4] and SVM classifier.

In [5] the author uses Olympic game ontology by fusing both textual and visual information but for ontology the author uses only the anthropological structure of the Olympic game event, where in our work we also interpolate the image feature in those Ontology creations.

In [6], [7] the author uses Gait feature to identify the human action in the video frames.

III. PROPOSED FRAMEWORK

Fig. 1 shows the pictorial representation of the proposed framework, where from the given sports image the key points are identified using the concept of SIFT. From the key points the features are identified. Those estimated features are clustered to create a Bag of visual words dictionary from which the semantically similar visual words are clustered as one set.

Ontology is created for the events cricket, croquet, marathon, polo, rowing, Mountain climbing, sailing, ski, tennis and volley ball. Then the clustered visual word sets are mapped to that Ontology [8]-[10].

A. SIFT

Searching a data from a very large scale database is not a complicated task if the data is just a literal or numerical data, since all those data have only one representation. If we have large collection of images and to search an image from it by giving either image or text query as input is a very complicated job as an image can be represented in many different ways with its different features. In this paper we have used SIFT [11]-[14] for image feature analysis as it is proved to provide better matching result [15] even there exist few many papers as an extension of SIFT such as PCA-SIFT and ASIFT [16] still this is proved to be efficient to retrieve invariant and affine image when compared to the training image.

Scale Invariant Feature Transform analysis consists of four steps that is scale space extrema detection, key point localization, Orientation assignment and Key point description. Here we use SIFT to determine the key point and its corresponding orientation matrix so as to match with the trained image in Knowledge Intelligent phase.

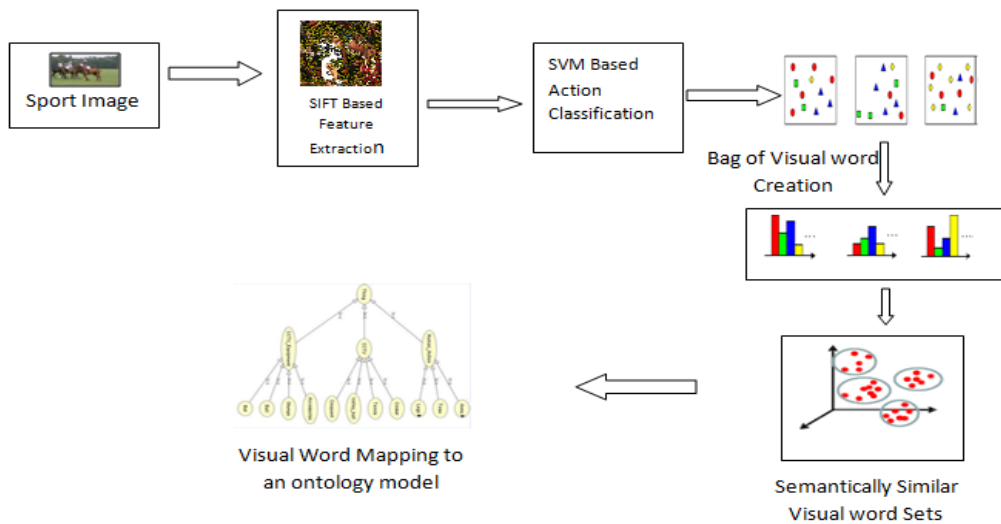


Fig. 1. Proposed framework.

1) Scale space representation

To determine the invariance feature, first the image is scaled using multi-scale signal representation concept. This is done by smoothing the image in different filter coefficient. The better filter suggested for smoothing is Gaussian filter as this- is one of those filters which would not introduce any new spurious structure while filtering the image. A Gaussian function [10] coefficient for a 2D data such as image is given by:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (1)$$

where σ represent the scaling factor of Gaussian filter.

Now for an 2D image $I(x, y)$ the scale space is determined by following (2)

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (2)$$

where the Gaussian filter value is convolute to the every single pixel value $I(x, y)$.

To determine the key point in the image, an n number of Gaussian filtered image is taken. Where n should be multiple of 5 as the images are all grouped to 5 octaves. Normally an octave is said to be a interval between two different musical pitch which are half of its frequency, similarly here octave is an interval between two same images but with different filter coefficients. With this set of image the difference between the Gaussian filtered image is determined which is called as DoG. The main objective of this step is to find the interesting points for the given image, by finding the local maxima of the image with some function. Thus from the DOG the scale's are normalized by applying Laplacian. From the Laplacian response, the peak points are determined which are considered as key points. Fig. 2 shows the Laplacian response plot and the maxima [17] extraction points.

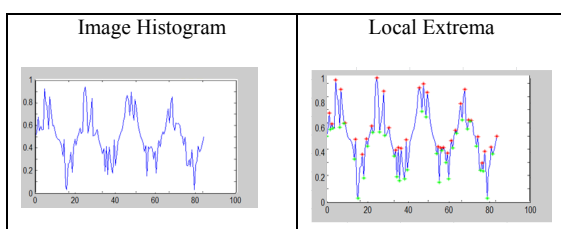


Fig. 2. Image extrema.



Fig. 3. Key points.

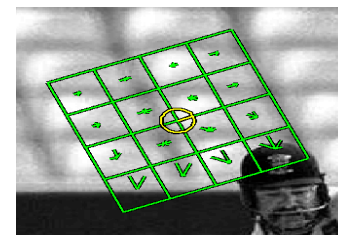


Fig. 4. Key point orientation.

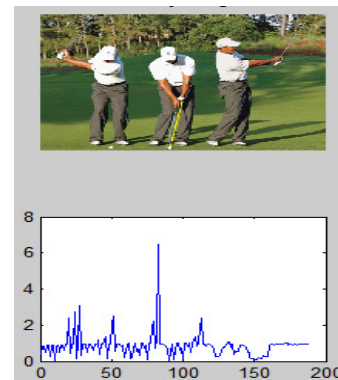


Fig. 5. Key point orientation histogram.

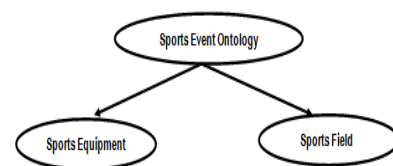


Fig. 6. Sports event ontology- level 1

2) Key point detection

Once we determine all the key points the number of key points would be numerous as shown in Fig. 3, if we

process this much of key points for a single image the computation time would be more and for comparing N images from different domains would be enormous. So we can eliminate some of the key points those have low contrast. This can be calculated by excluding the key points with local minima of less than 0.03 or any other threshold as fixed. Also we can eliminate the key points near the edges by determining the edges. SIFT uses Hessian Matrix for edge detection.

3) Orientation detection

For all the selected key points, the orientations are determined by using image pixel gradient. It is the direction where the intensity of a pixel changes from low to high. With all these gradient details a histogram is generated as shown in Fig. 5 and that direction whose histogram value is maximum is taken as the concern key points orientation. Figure 4 shows the orientation of a single key point.

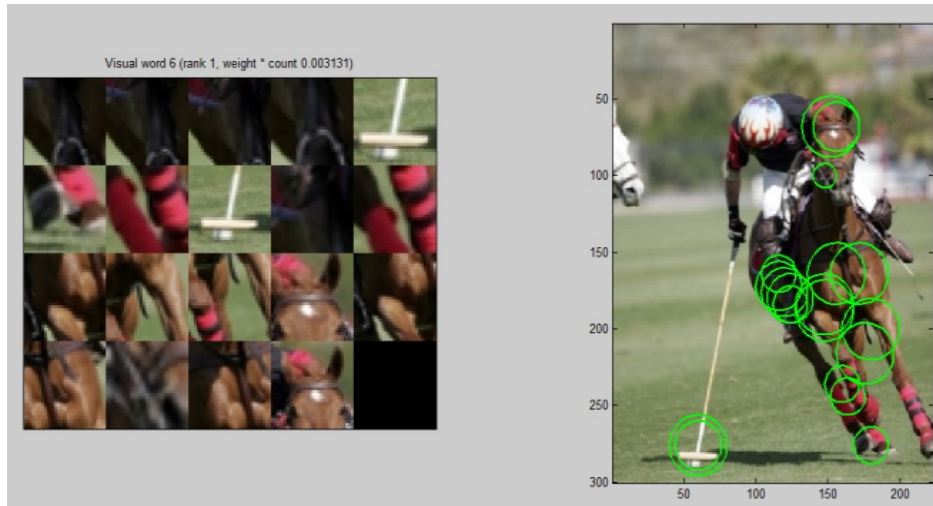


Fig. 8. Visual word creation using modified SIFT.

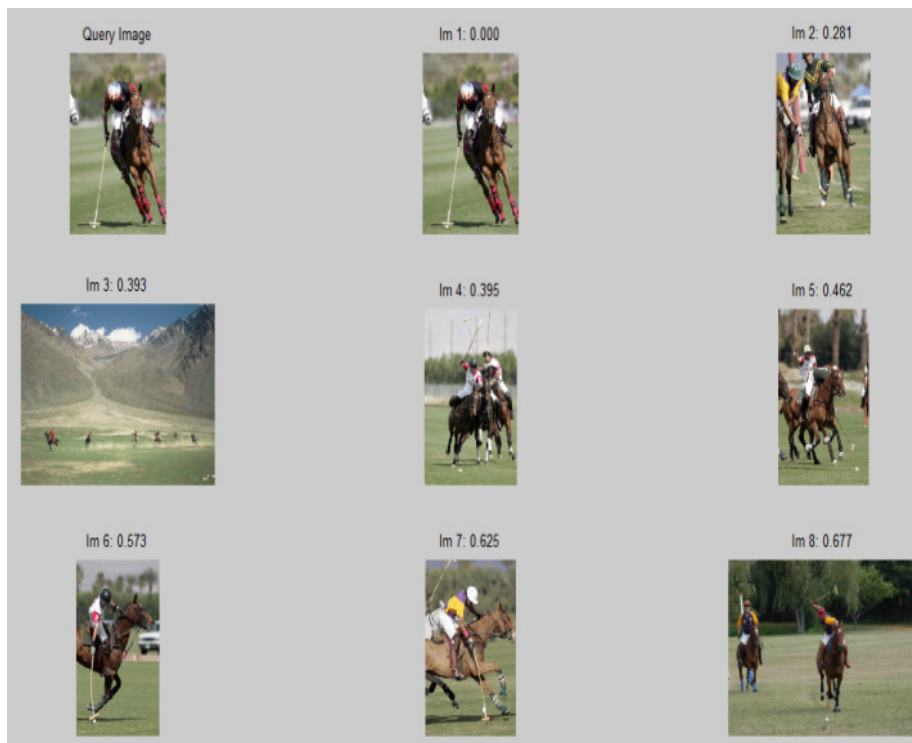


Fig. 9. Relevant image classification.

B. SVM

Support Vector Machine classifier [18]-[20] is a kind of unsupervised Learning algorithm. That studies its environment from its past history of input and output data sets. Such kind of data sets are said to be a training data sets and the prediction depends upon the hypothesis function used. In SVM, Kernal functions are used as such hypothesis function. For image classification, more general linear kernel

[21]-[23] can be used.

Let (X_i, Y_i) be the attributes in our training example Assume X_i be the histogram value of the Key points identified, and it belong to a hypothesis space R, Then Y_i is -1 (negative images) if it does not belong to R or +1 (positive images) if it belongs. The kernel function can be represented as

$$f(X_i) = w X_i + b \quad (3)$$

where w , is the normal of a line used to classify the categories of images and b constant value called bias.

C. Bag of Visual Words

The Bag of visual word concept is used to determine the low level features used to describe an image. To determine the features of an image, mostly the interesting key points on the images are determined. The concept used to determine the key points are Blob detector, Corners detector and SIFT detector.

The key point values are clustered using any of the supervised learning techniques such as K-Mean clustering or SVM kind of clustering technique. Such clustered key points histogram values are used to index the images in this bag of visual word concept.

D. Sports Ontology

Ontology is an organized way of representing the entities and their relationship in a domain. This ontology can be represented using OWL the Web Ontology Language. At first few decades after introducing the concept of Semantic web technologies Resource Description Framework, RDF [24], [25] was used to define the ontology of a domain, where the hyperlink in a html page is replaced by the <subject, object, predicate> triplets

Then in second decade of ontology the concept of RDF Schema (RDFS) was used which introduces the concept of creating associate relationship between two entities. Then by 2004 [26] W3C recommend OWL which is the combination of RDF, RDFS and some of cardinal expression which is missing out in RDFS which also help us to build more complex domain ontology.

The sports events considered for the creation of sports ontology are cricket, croquet, marathon, polo, rowing, Mountain climbing, sailing, ski, and tennis and volley ball. The sports event ontology is designed by classifying the classes in two different classes at the top level which is shown in Fig. 6 and Fig. 7 shows the overall ontology created.

The overall Sports Event ontology created is using the following classes and sub classes:

The Sub-classes of sports field class are: Grass field, Water field, Road field, Indoor stadium, and Rocky field and Snow field

The Sub-classes of sports equipment class are:

Ball, Bat, Boat, Hoop, Mallet, Mast, Net, Oar, Pole, Racquet, Rope, Stump and Skis

For each sports image, the visual feature of each individual object such as ball, bat, pole, Net and so on are identified using Bag of Visual word as shown in Figure 8. Such identified features can be mapped with the Sports event OWL's sub-classes.

IV. RESULT AND IMPLEMENTATION

The above discussed work has been implemented using Matlab 7.10 under Windows operating system. Figure.8 shows the visual word creation using modified SIFT. The procedures to create the visual words are: after extracting the SIFT descriptors, these vectors would be the input of a clustering algorithm. Here we used the K-means algorithm

for its rapidity. Euclidean distance is chose to measure the degree of dissimilarity between two vectors.

For our implementation we have taken seven out of ten sports event as discussed in earlier section. So, for seven sports events 50 image per each event so totally 350 images are taken as test images. Out of 350 images 241 images are classified as per our proposed framework.

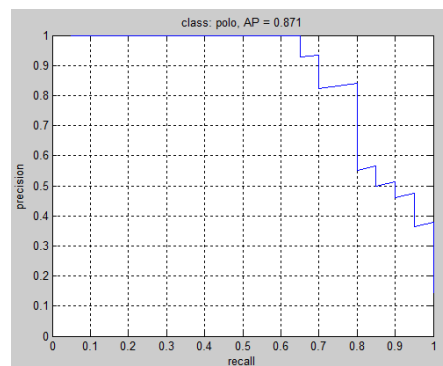


Fig. 10. Precision – recall – average precision recall.

Fig. 9 shows the implemented framework for the relevant image classification for the given query image.

To measure the classification performance quantitatively a Precision-Recall curve is computed and its average Precision-Recall.

Precision is the fraction of the images classified that are relevant to the user's query image. Recall is the fraction of the images classified that are relevant to the query that are successfully classified. From the Precision Recall plot the area under the precision-recall curve gives the Average Precision Recall. The AP provides an accuracy of 87.1% for the given Polo event image.

V. CONCLUSION

An image can be represented either by its low level features or by its high level features. To represent an image using both of these features is one of the challenging parts due to the semantic gap between them. In this paper a design of representing the sport image's low level feature semantic through modified SIFT is explained and the creation of this feature based bag of visual word dictionary is done through SVM. The creation of Sport event ontology is also illustrated.

REFERENCES

- [1] R. I. Minu and K. K. Thyagarajan, "Multi modal ontology search for semantic image," *ICTACT Journal on Image and Video Processing*, March 2012
- [2] J. H. Su *et al.*, "Ontology based semantic web image retrieval by utilizing textual and visual annotations," in *Proc. IEEE International Conference on Web Intelligence and Intelligent Agent Technology*, pp. 425-428, 2009.
- [3] B. P. Yao, A. Khosla, and F. F. Li, "Classifying actions and measuring action similarity by modeling the mutual context of object and human poses," in *Proc. 28th International Conference on Machine Learning*, pp. 327-333, 2011.
- [4] L. Zhang, F. Z. Lin, B. Zhang, "Support vector machine learning for image retrieval," in *Proc. of IEEE International Conference*, pp. 107-118, 2001.
- [5] S. Chimlek, K. Kesorn, P. P. Nga and S. Posland, "Semantically similar visual words discovery to facilitate visual invariance," in *Proc. of ICME IEEE conference*, pp. 1242-1247, 2010.
- [6] V. Thanikachalam and K. K. Thyagarajan, "Human action recognition based on motion and appearance," *International Journal of*

- Advanced Information Science and Technology (JJAIST)*, vol. 7, no. 7, pp. 90-95, 2012.
- [7] V. Thanikachalam and K. K. Thyagarajan, "Human action recognition using accumulated motion and gradient of motion from video," in *Proc. of the International Conference on Computing Communication and Networking Technologies ICCCNT 2012*, pp. 26-28, 2012.
- [8] Y. G. Jiang and C. W. Ngo, "Ontology-Based visual word matching for near duplicate retrieval," in *Proc. IEEE International Conference on Multimedia & Expo*, pp. 1-4, 2008.
- [9] G. Nagarajan and K. K. Thyagarajan, "Multimodal anthropological based information retrieval system," *National journal of Computer Science*, Sathyabama University, pp. 39-45, 2012
- [10] "LSCOM lexicon definitions and annotations," DTO Challenge Workshop on Large Scale Concept Ontology for Multimedia, Columbia University ADVENT Technical Report #217-2006-3, 2006.
- [11] D. G. Lowe, "Distinctive image feature from scale invariant keypoints," *International Journal of Computer Vision*, pp. 1-3, 2004.
- [12] K. Mikolajczyk, B. Leibe, and B. Schiele, "Local features for object class recognition," in *Proc. the 10th IEEE International Conference on Computer Vision (ICCV '05)*, pp. 1792-1799, October 2005.
- [13] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615-1630, 2005.
- [14] J. Zhang, M. Marszałek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *International Journal of Computer Vision*, vol. 73, no. 2, pp. 213-238, 2007.
- [15] J. M. Morel and G. S. Yu, "ASIFT: A New framework for fully affine invariant image comparison," *Siam J Image Sciences*, vol. 2, no. 2, 2009.
- [16] S. Jonathan JB *et al.*, "Squint Svm for identification of relevant sections in web page for web search," in *Proc. 2009 Second International Conference on Intelligent Computation Technology and Automation*, pp. 2-4, 2009.
- [17] P. Quélhas, F. Monay, J. M. Odobez, D. G. Perez, and T. Tuytelaars, "A thousand words in a scene," *IEEE Transaction Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1575-1589, 2007.
- [18] J. X. Wu, "Efficient HIK SVM learning for image classification," *IEEE Transaction on Image Processing*, 2012, vol. 21, no. 10, pp. 4442-4453.
- [19] L. Zhang, F. Z. Lin, and B. Zhang, "Support vector machine learning for image retrieval," in *Proc. IEEE Int. Conf. on Image Processing*, vol. 2, pp. 721-724, 2001.
- [20] HaiyingGuna, SameerAntani, L. R. Long and G. R. Thoma, "Bridging the semantic gap using ranking SVM for image retrieval," in *Proc. of ISBI Conference*, pp. 354-357, 2009.
- [21] P. Ndjiki, O. Novychny, and T. Wiegand, "Merging MPEG 7 descriptors for image content analysis," in *Proc. IEEE conference*, vol. 3, 453-456, 2004.
- [22] C. C. Zhang and X. Chen *et al.*, "A multiple instance learning approach for content based image retrieval using one class support vector machine," in *Proc. IEEE conference*, pp. 1142-1145, 2005.
- [23] R. I. Minu and K. K. Thyagarajan, "Automatic image classification using SVM classifier," *CiiT International Journal of Data Mining and Knowledge Engineering*, vol. 3, no. 9, pp. 559 - 564, 2011.
- [24] L. Y. Yu, *Introduction to the Semantic Web and Semantic Web Services*, Chapman & Hall/CRC Taylor & Francis Group, 2007.
- [25] G. Nagarajan and K. K. Thyagarajan, "Linguistic conversion of syntactic to semantic web page," in *Proc. the Second International Conference on Advances in Computing and Information Technology (ACITY)*, July 13-15, 2012, pp. 207- 213.
- [26] M. Dean and G. Schreiber. (2004). OWL Web Ontology Language Reference. W3C Recommendation. [Online]. Available: <http://www.w3.org/TR/2004/REC-owl-ref-20040210/>.



K. K. Thyagarajan obtained his B.E., degree in Electrical and Electronics Engineering from PSG College of Technology (Madras University) and received his M.E., degree in Applied Electronics from Coimbatore Institute of Technology. He also possesses Post Graduate Diploma in Computer Applications from Bharathiar University. He obtained his Ph.D. degree in Information and Communication Engineering (Computer Science) from College of Engineering Guindy, Anna University. He has twenty five years of teaching experience. Now he is the Dean (Academic) of R.M.D. Engineering College. He has written 5 books in Computing including "Flash MX 2004" published by McGraw Hill (INDIA) and it has been recommended as text and reference book by universities and Polytechnics. He has published more than 50 papers in National and International Journals and Conferences. He is a grant recipient of DST & Tamil Nadu State Council for Science and Technology. His biography has been published in the 25th Anniversary Edition of Marquis Who's Who in the World. He has been invited as chairperson and delivered special lectures in many National and International conferences and workshops. His current interests are Multimedia Networks, Content Based Information Retrieval, Web services, Data Mining, e-learning, Image Processing.. He is reviewer for many International Journals and Conferences. He is a recognized supervisor for Ph.D by Anna University Chennai, MS University, JNTU and Sathyabama University, and now 11 students are doing Ph.D. under his supervision.



G. Nagarajan has received his diploma in electronic & communication engineering from directorate of technical education 1997. He has received his BE degree in electrical & electronic engineering from Manonmaniam Sundaranar University 2000. He received his ME degree in applied electronic engineering from Anna University 2005. He also obtain his ME degree in Computer Science Engineering at Sathyabama University 2007. . He is currently pursuing Ph.D. (part time) at Sathyabama University Chennai. He has published papers in International/National journals and conferences in the area of Web Ontology & Computer Vision.