Novel Intent based Dimension Reduction and Visual Features Semi-Supervised Learning for Automatic Visual Media Retrieval

Subramanyam kunisetti[1]† and Suban Ravichandran[2]††

[1]Research Scholar, Department of CSE,Annamalai University,Annamalainagar,Chidambaram,Tamilnadu. India
ORCID:0000-0003-3241-9888
[2]Associate Professor, Department of IT, Annamalai University, Annamalainagar,Chidambaram,Tamilnadu. India

Abstract

Sharing of online videos via internet is an emerging and important concept in different types of applications like surveillance and video mobile search in different web related applications. So there is need to manage personalized web video retrieval system necessary to explore relevant videos and it helps to peoples who are searching for efficient video relates to specific big data content. To evaluate this process, attributes/features with reduction of dimensionality are computed from videos to explore discriminative aspects of scene in video based on shape, histogram, and texture, annotation of object, co-ordination, color and contour data. Dimensionality reduction is mainly depends on extraction of feature and selection of feature in multi labeled data retrieval from multimedia related data. Many of the researchers are implemented different techniques/approaches to reduce dimensionality based on visual features of video data. But all the techniques have disadvantages and advantages in reduction of dimensionality with advanced features in video retrieval. In this research, we present a Novel Intent based Dimension Reduction Semi-Supervised Learning Approach (NIDRSLA) that examine the reduction of dimensionality with explore exact and fast video retrieval based on different visual features. For dimensionality reduction, NIDRSLA learns the matrix of projection by increasing the dependence between enlarged data and projected space features. Proposed approach also addressed the aforementioned issue (i.e. Segmentation of video with frame selection using low level features and high level features) with efficient object annotation for video representation. Experiments performed on synthetic data set, it demonstrate the efficiency of proposed approach with traditional state-of-the-art video retrieval methodologies.

***Keywords*:** *Intent based video retrieval, visual features, Dimensionality reduction, annotation of object, classification, multi labeled data learning, selection of feature and extraction of feature.*

1.Introduction

In recent improvements, digital media have been increased with abundance of data relates to visual around our present outside web environment. Present human being generates millions of visual data to improve/increase strategic data sources relate to digital media. Maintenance and organizing of visual data is an impressive concept to owing to large amount of digital data because of different tasks appeared i.e. visual search data, reverse search data, retrieval of data, indexing and captioning have been aggressive concept from over decade past years. Because of intervention in human growing on different domains like sensing of remote, forensic, search of user historical data, recommendations of different user's data, and some other domains like video processing in satellite, search of multimedia recommendations, privacy in some digital media systems and demand on video. Based on visual search video retrieval there is an active decade is necessary for large volumes of image descriptors present in video to get fast and accurate video retrieval, most popular video related descriptors like SURF, SIFT, BRISK, FREAK and etc. All these descriptors are matches with image frames in video and explore video based on combined hypothesis selection from different video patches. Scale invariant feature transform is the robust when compare to all descriptors used in video patch retrieval from video data

Manuscript received June 5, 2022
Manuscript revised June 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.6.32

sources. Basic representation of video is it stores significant data about main issues appeared in the world which are successfully managed and presented in time variant manner. Content based video retrieval is the efficient approach for retrieving similar videos from data sources. It is the extension to content based image retrieval, in that video contains specific features like story, scene, shot, frame selection, each frame consists with images.

Reduction of dimensionality with visual feature matching is the crucial concept in video retrieval, previously, a review of different sources related to noise (dimension) reduction can be identified. Based on standards of machine learning approaches related to supervised, semi supervised and unsupervised settings. Many of the noise reduction tolerant versions classifiers have been proposed i.e. discriminative analysis, logistic regression, k-nearest neighbor, different boosting calculations, support vector machine, deep neural networks and other general classification related frameworks are performed with similar parameter sequences. Based on these approaches a little survey have been conducted to evaluate noise or dimension reduction in non standard manner where noise magnitude to be increased parallel. Non standard settings consists i) semi-supervised learning, it refers to some situation like where if few of the labeled video data available then noise on those few labels is most prevalent, and where data jointly inferred from unlabeled video data; ii) multi-labeled learning, where video data consists different domains which may not belongs to specific domain; iii) high dimensional data, where abundance of features are represented with noise labeled data then dimensionality curse problem may appear, in that situation reduction of dimensionality (RD) is useful based on different visual features at pre-processing stage. So that a Novel Intent based Dimension Reduction Semi-Supervised Learning Approach (NIDRSLA) that examine the reduction of dimensionality with explore exact and fast video retrieval based on different visual features. For dimensionality reduction, NIDRSLA learns the matrix of projection by increasing the dependence between enlarged data and projected space features. Proposed approach also addressed the aforementioned issue (i.e. Segmentation of video with frame selection using low level features and high level features) with efficient object annotation for video representation.

NIDRSLA be the efficient dimensionality reduction approach which can be used to represent different settings like visualization of video in pre-processing before performing classification. To improve better test results of NIDRSLA, we use some quantitative measurements like k-nearest neighbor used to represent low dimensional video data. In our implemented experiments, our proposed approach compare with different baseline approaches performed on labeled synthetic data. NIDRSLA is the better approach when compared to traditional RD approaches according to different metrics applied on statistical test analysis.

Main contribution of proposed approach is described as

i) Semi supervised noise tolerant dimension reduction approach based on maximization of dependences.

ii) Implement a novel framework to handle classification of video retrieval with dimensionality

iii) Comprehensive analysis of proposed approach explains effectiveness of NIDRSLA with existing approaches applied on different synthetic data sets in real world entity.

2. Review of Related work

In this section, we discuss about different literature study about present media related retrieval, these survey of techniques mainly discuss feature representation, retrieval of digital media related data.

Ladahke et al. give an audit of various recovery errands with their applications and early methodologies which prompt the advancement in the field as of now. The creators additionally outline the fundamental procedure for the CBIR assignments. In [4], Csurka et al., analyze classifiers the visual order which is one of the field's soonest works after which further developed techniques arose which prompted the current notoriety. After the appearance of profound learning and the presentation acquires that accompany it, a few strategies taking on profound learning methods arose. Both customary techniques and profound learning strategies enjoy their benefits and weaknesses. We center on current realities as to why profound learning (DL) is the one of the best alternative for the content-based recovery assignments. Deep learning (DL) is better at learning both nearby and worldwide elements whereas conventional techniques need separate modules to learn/distinguish neighborhood provisions like shapes, surface, shading, edges, objects, direction, and so forth and they perform inadequately in learning worldwide elements by affiliation. DL models can be intended to be scale, spatial, and shading invariant and can perform well in such cases subject to preparing different information. Conventional video preparing modules need separate preprocessing relating to their plan and the provisions from various modules of customary strategies must be totaled to a proper length and it is undeniably challenging to keep up with the relationship whereas profound learning models promptly yield highlight portrayals of fixed measurement.

The starter works in the field depended on obvious signs like shape, shading, surface, edge, and spatial components [19] after which include identification procedures, for example, SIFT, Speed up robust features (SURF), and their variations were utilized for better execution [18]. Afterward, these cleared the way for the utilization of nearby elements utilizing inadequate portrayals and a Bag of visual words (BoVW) [19]. Exhibitions of this load of strategies were outperformed by profound learning methods. The original work [17] utilizes auto-encoders to digest and get familiar with the portrayals of videos for content-based recovery assignments by the planning the input videos into the succinct 28-bit parallel codes. [3] proposes to utilize the visual portrayals got from the top layers for the prepared neural organization to be utilized as visual encoding for recovery.

In [29], the creators propose to utilize unaided profound learning techniques to semantically hash videos for the recovery by separating imperative data to work on the effectiveness of the visual hashing. [15], [12] present not many different strategies on hashing techniques for video recovery utilizing profound learning. [16] Proposes to utilize Image Net [5] pre-prepared models as component extractors for video recovery assignments and thinks about the presentation against other contemporary techniques. [14] Propose a profound positioning model to catch bury and intraclass contrasts to improve the insightful capacity of the model utilizing trio inspecting. [7] offers a far-reaching overview on the movement of the profound learning-based techniques utilized for the CBIR throughout the decade with an itemized investigation on the exhibition of best-in-class models. The vast majority of the chips away at video-related errands utilizing profound learning were totally overwhelmed by the 3D convolution neural organization (3D CNNs) [11] and its variations with optical stream data. 3D CNNs essentially, utilize commonly little 3D pieces to learn spatial and worldly provisions together by working on spatial components of casing and transient measurement which is across the casings. However they perform well on standard assignments, they experience the ill effects of computational intricacy, failure to learn explicit traits of items in the recordings as they learn spatial-worldly data firmly which confines them to be utilized in recovery applications.

Afterward, [6] presented LRCN, the use of repetitive organizations with the help of CNNs for video subtitle age errand, and J. Y. Ng et al. [16] concocted a design to utilize 2D CNNs and RNNs with optical stream for video characterization. The previously mentioned works are demonstrated by the potential and adequacy coupling of CNNs and RNNs are together for the video-related undertaking like the arrangement and inscribing. Later [10] proposed a technique utilizing the 2D CNNs and RNNs for video recovery by ascertaining the implanting misfortune between the two distinctive contributions to the organization. The variations of utilizing 2D CNNs with RNNs can overcome the computational intricacy of being lightweight in nature. RNNs are innately equipped for the learning transient groupings well and recordings are comprised by the activities of items under the center. The movement in the recordings is learned well by RNN and CNNs are demonstrated throughout the years to perform on video-related assignments. However numerous techniques addressed all through the paper can work with extraordinary exactness of recovery, they regularly experience the ill effects of the shortcoming as far as the season of recovery because of their substantial and complex nature which blocks them to utilized in a genuine world, down to earth applications. A few works offer to natural strategies for streamlining and adaptability of recovery assignments such as applying PCA for dimensionality decrease [14] of element portrayal for quicker examination, quicker ordering in recovery undertakings [23], applying diverse grouping techniques [20] for simpler coordinating, recovery, and estimate of client input inquiries and so forth [29]. These secluded enhancements can help in planning a recovery system that may work viably in reasonable application where exact and quick extraction results are essential core interest.

3. Basic Preliminaries

In this section, we discuss about basic representation of feature representation/extraction used in proposed approach

a) Description

In $\{a_i\}_{i=1}^{m}$, m be the different data dimensional data points $a_i \in \square^{D}$, let us assume that consisted labeled video data points arranged which contains labeled data with unlabeled data vector representations i.e. $k+v=m$, Let A be the $m*D$ dimensional matrix with different dimensional points represented as vectors.

Let us C be the no. of class labels and $\gamma_i^K \in \{1,0\}^c$ be the data labeled vector with different data points $a_i, i=1,2,....,k$, data point elements are explored from $\gamma_i^K \in \{1,0\}^c$ for class labels equal to 1 for unlabeled class label represented as 0. Let us $\gamma^K \in \{1,0\}^{1*c}$ be the labeled matrix with labeled class data $\gamma_i^K, i=1,2.....k$ and $\gamma^V \in \{1,0\}^{v*c}$ be the associated matrix formation unlabeled video data.

Projection matrix with linear regression objective functions should be learned as $Q \in \square^{D*d}$, lower dimensional representation vector space is $z \in \square^{d}, z = Q^T a$

Where d<D and Q^T transpose of vector representation of matrix Q. In our implementation, we assume potentially noise labeled matrix Y^V . Label propagation of NIDRSLA; introduce soft label representation $F \in \square^{m*C}$ for associated matrix $Y = \begin{pmatrix} Y^K \\ Y^V \end{pmatrix}$, F_i represents random data point a_i which is belongs to C class label.

b) K-nearest neighbor label Propagation

Labeled propagation of similar data with same class label evaluated using neighborhood graph, formulate the labeled propagation for different labels to reduce noise with respect to dimensionality based on feature extractions. Procedure to evaluate dimensionality reduction is described as

First construct neighborhood graph based on adjacent matrix which is described as

$$W_{i,j} = \exp(-\sigma^2 \| a_i - a_j \|^2)$$

Where $\| a_i - a_j \|$ be the Euclidian distance metric between input data points ai &aj and σ be the parameter and this distance also used for k-NN evaluation is describes as

$$W_{i,j} = \begin{cases} 0 & otherwise \\ 1 & a_i(a_j 's)k-NN / a_j(a_j 's)k-NN \end{cases}$$

Based on above formations, evaluate dimensionality reduction as described as

$$F(t+1) = I_\alpha TF(t) + (I - I_\alpha)Y$$

I_α be the n*n dimensional diagonal matrix with high parameters $_\alpha$ $0 \le \alpha_1 < 1$ (parameter value present in between) for low dimension representation in video retrieval.

4. NIDRSLA Implementation

a) Description: In this work, a course of action-driven design for recuperating near videos from a helpful information base is proposed. In fig 1 the step by step portrayal of the proposed recuperation structure was shown. The secret NIDRSLA model means to learn channel part by making an undeniably remarkable depiction of data in each and every layer.

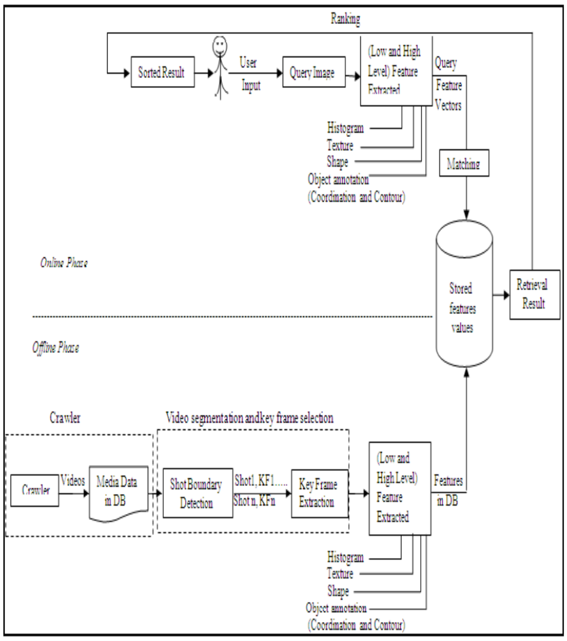

Figure 1 Implementation procedure is done by the proposed approach.

Despite its direct number juggling, NIDRSLA is one of the most stunning resources in vision systems. Overall there are 3 kinds of layers are there in NIDRSLA i.e., there are convolutional layer, pooling layer and totally related layer. The resulting layer is generally treat's as a special layer and at the data layer model gets data tests. Each and every convolutional layer generates feature maps by convolving with data feature maps. The model feature maps made by the convolutional layers is expected to down by the pooling layer, which are regularly drilled by finding closed by the maxima in an area. Moreover translational invariance is given by the pooling layer, and the it diminishes the number of neurons to get ready in class layers. In totally related layer each and every neuron has undeniably slower affiliation is stood out from the convolutional layer. In the piece of NIDRSLA after totally related layer is known as the classifier part and before related layer is known as component extractor part. The de-followed depiction design is used to shown after subsections.

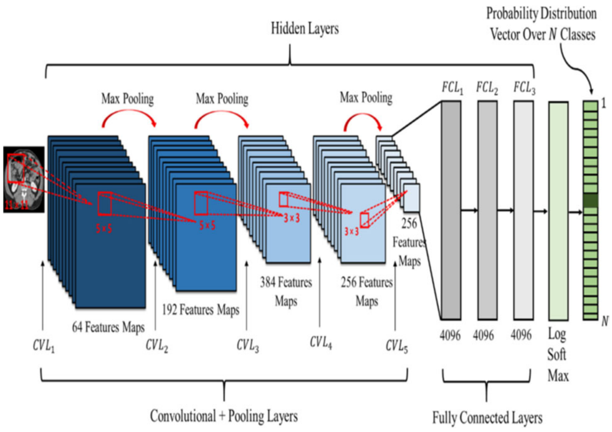

Figure 2 Reduction of Noise/ dimension in pre-processing of proposed approach

In fig 2 outline models are utilized for getting ready involved totally 8 layers are there in that 5 layers are convolutional layers and remaining 3 layers are totally related layers. CVL and FCL is also known as convolutional layers and totally related layers, where the layers number is also known as addendum for e.g., the first convolutional layer is also known as CVL 1. The result of the 3rd totally related layer was supported to which is having delicate limit of 24 results, which generate probability scattering of each and every class mark. Thusly the size of the probabilities vector is 1 x 24, where the class of datasets are gotten to identifies the each and every vector part. The gray scale in time 224 x 224 estimates as wellsprings of information and not in the slightest degree like the model introduced in CNN use pieces which is having lower number. The first convolutional layer(CVL 1) channels the video data with size of 11 x 11 for 64 parts with 4 pixels walk comparable. The walk is the division between focal points having open field of neurons in neighborhood in the map. The results of first convolutional layers (CVL 1) is non-linearity is urged and after that moved to compressing neighboring neurons by the help of spatial max-pooling layer. The changed straight unit nonlinearity is used for the results if all convolutional layers and totally related layers. This framework with changed straight unit has no ability to arrange couples of times faster than the relative tanh unit moreover it allows to move with vanishing point issues.

a) Dependence Maximization based Dimensionality Reduction

Main motivation behind this module i.e. maximization of dependence, it describes the relation between features and explored label with respect to input object. We want maximize dependences between similarity in feature and similarity in label, common measure dependences based on Hilbert Schmidt independence selection [HSIS] which is described as

$$HSIS(A,B) = \frac{1}{(m-1)^2} tr(KHLH)$$

tr denotes the traceable matrix, $H \in \mathbb{R}^{m*m}$ is given by $H_{i,j} = \delta_{i,j} - n^{-1}$ where $\delta_{i,j} = 1$ then i=j if $\delta_{i,j} = 0$ otherwise, L be the kernel matrix over feature space evaluated over labeled space.

Basic pseudo code used for dimensionality reduction with different visual features described in algorithm

```
Basic pseudo code representation of NIDRSLA

Input: $A:m*D$ matrix Y:m*C matrix labels, hyper related parameters l, Id, &d
Initialize the labeled data set with different parameters described as $Y_{ic}^{V} = 0, c = 1,2,....,C$
Construct adjacent matrix based on above equations
Normalize the adjacency matrix with evaluated settings i.e. $W = D^{1/2} W D^{1/2}$.
Evaluate stochastic matrix using $T = D^{-1} W$, where $d_{ij} = \sum_{l=1}^{m} W_{ij}$.
Solve the linear system evaluation based on above equations is $(I - I_{\alpha} T)F = (I - I_{\alpha})Y$.
Evaluate F.
Design the approximate matrix formation $Q^T H F F^T HA$
Construct projection related matrix formation using $Q^T H F F^T HA$ then $q \in \mathbb{R}^{D*d}$.
Output: $Q : \mathbb{R}^{D} \to \mathbb{R}^{d}$
```

Algorithm 1. Pseudo code of NIDRSLA.

Let propagation matrix $q \in \mathbb{R}^{D*d}$ & described function $\varphi : \mathbb{R}^{D} \to \mathbb{R}^{d}, \varphi(a) = q^T a$, then kernel function is evaluated as

$$\kappa(a_i, b_i) = \left\langle \varphi(a_i), \varphi(a_j) \right\rangle = \left\langle q^T a_i, q^T a_j \right\rangle = q^T a_i a_j^T q$$

Here, $\{a_i\}_{i=1}^{m}$ be the given kernel matrix with approximate parameter sequence $L = Aq^T Q A^T$.

Based on these conditions, to get irrelevant label reduction and dimensionality is also reduced using optimized kernel function i.e. $\psi(Q) = tr\left(HAQ^{T}QA^{T}HFF^{T}\right) = tr\left(Q^{T}A^{T}HFF^{T}HAQ\right)$. Based on semantic data relations of $Q^{T}HFF^{T}HA$. Hence reduction of dimensionality with direct propagations with optimized label which is described as

$$\arg_{q}\max\psi(Q) = \arg\max tr(Q^{T}(A^{T}HFF^{T}HA)Q)w.r.t(Q \in \square^{D*d}, QQ^{T} = T)$$

It describes the optimized maximal solving of dimensionality reduction with preferable operations. Based on this procedure, we compute the minimal reduction of dimensionality with semantic features of video data.

5. Experimental Evaluation

The evaluation study of proposed implementation with comparison to traditional existing video retrieval approaches based on input query. To do this work efficiently we use synthetic data set performed or evaluated. Proposed implementation done in JAVA platform with Netbeans user interface to accumulate different results performed on different data sets. NIDRSLA performed on 5 different video retrieval data sources and perform different parametric with classification results.

After evaluating optimal solution of the query relates to video content then precision, recall and F-measure, time metrics are generated for performance evaluation

$$precision = \frac{No.of\ relevant\ videos\ retreived}{Total\ no.of\ videos\ retrived}$$

$$recall = \frac{No.of\ relevant\ videoss\ retreived}{Total\ no.of\ relevant\ videos\ in\ datasource}$$

$$F\ Score = 2\frac{precision * recall}{precision + recall}$$

These are the important factors to evaluate the traditional approaches with that of the performance of proposed approaches. Remember is the portion of the appropriate videos which has been retrieved with proposed calculation.

Performance metrics of proposed approach with comparison to traditional approaches a light, cogent and end-to-end content based retrieval (LCEECR) (1) novel feature selection-based approach (NFSA)(2) two-level label recovery mechanism (3) (TLRM) described as follows:

Table 1. Comparison accuracy values

Databases	NIDRSLA	LCEECR	NFSA	TLRM
Data set 1	**0.96**	0.71	0.81	0.72
Data set 2	**0.95**	0.62	0.79	0.68
Data set 3	**0.92**	0.58	0.70	0.64
Data set 4	**0.94**	0.74	0.62	0.58
Data set 5	**0.93**	0.64	0.77	0.64

Table 1 shows that when compare to the existing techniques our proposed accuracy values gives better results.

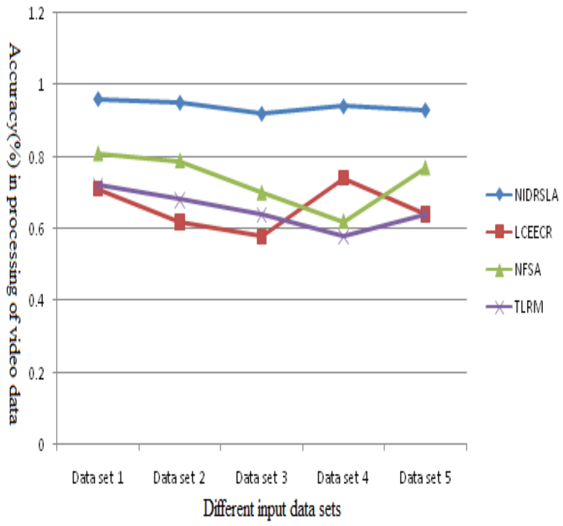

Figure 3 Performance evaluation with respect to accuracy.

As shown in figure 3, it show the performance evaluation of accuracy in retrieval videos from overall video related data, if we increase data sets values then accuracy of proposed approach gives better and accurate when compare to existing approaches in terms of video retrieval from video data sources. Whenever we increase the data sets values then existing approaches give less results in term of accuracy with comparison to NIDRSLA, as shown in table 1, NIDRSLA gives highest accuracy when different data sets applied on semantic way.

Table 2 describes the values relates to precision which contain different attribute relations, video retrieval.

Data sets	NIDRSLA	LCEECR	NFSA	TLRM
Data set 1	0.72	0.42	0.39	0.47
Data set 2	0.78	0.35	0.53	0.52
Data set 3	0.71	0.38	0.47	0.48
Data set 4	0.68	0.45	0.43	0.38
Data set 5	0.74	0.38	0.44	0.45

Table 2 Precision values

As shown in table 2, the precision values i.e. highest matching values relates to video describes better precision in retrieval of video from large video data sources.

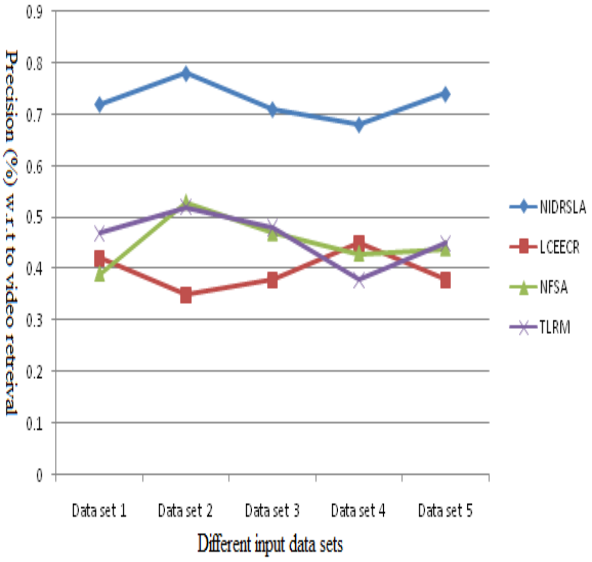

Figure 4. Performance evaluation of precision in video retrieval

Table 3 shows that recall values gives effective video of retrieval formation with the help of relevant and irrelevant videos from five different real time video sources.

Table 3 Video data processing is done by using recall values.

Data sets	NIDRSLA	LCEECR	NFSA	TLRM
Data set 1	0.81	0.54	0.61	0.71
Data set 2	0.75	0.48	0.50	0.64
Data set 3	0.69	0.56	0.49	0.72
Data set 4	0.81	0.60	0.52	0.49
Data set 5	0.79	0.52	0.51	0.41

Figure 5 describes the performance evaluation of recall with comparison to existing classification approaches, all those approaches having low matching rate of true negatives and false negatives with associated resistance and retrieval video data from video data sources.

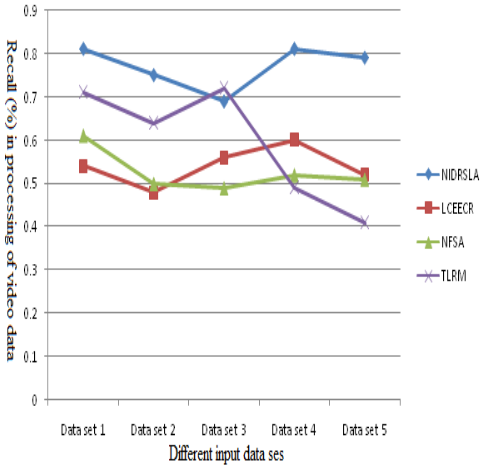

Figure 5 Performance evaluation of recall with processing video data.

Figure 5 shows that the retrieval formations with the help of five different query videos from five different video dataset sources in selection of video data with multi labeled attributes.

Table 4 shows the execution time values with respect to processing of different data sets to evaluate or retrieval of relevant matched videos from video data sources.

Table 4. Time values with respect to processing of overall data sets.

Data sets	NIDRS LA	LCEECR	NFSA	TLRM
Data set 1	8.1	15.4	16.1	12.6
Data set 2	9.4	14.8	15.0	16.7
Data set 3	13.9	18.6	17.9	17.2
Data set 4	11.1	19.8	18.2	19.5
Data set 5	12.9	21.35	25.1	24.1

Figure 6 describe the execution time performance of proposed approach and other traditional approaches to evaluate overall process of selection of relevant videos with respect to avoid errors in selection of relevant video data with different attribute relations.

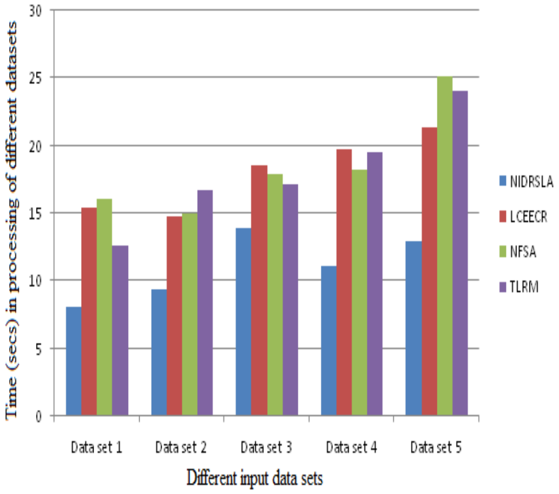

Figure 6. Performance evaluation of time with respect to processing of data sets.

The recovery of all five different sets are done by using five different videos of querying and for each and every dataset we have to get perfection, measurable principles are identified. The evaluation between the demonstrates the efficiency of the suggested system, functions removal centered on multi-objective marketing techniques, , suggested comprehensive methods gives best results when compare to current systems. Finally our proposed approach gives best and efficient results with respect to the processing and exploring video data from data sources.

6. Conclusion

This paper has given an overview of the Intent based Dimension Reduction Semi-Supervised Learning Approach (NIDRSLA) that examine the reduction of dimensionality with explore exact and fast video retrieval based on different visual features. For dimensionality reduction, NIDRSLA learns the matrix of projection by increasing the dependence between enlarged data and projected space features. Proposed approach also addressed the aforementioned issue (i.e. Segmentation of video with frame selection using low level features and high level features) with efficient object annotation for video representation. We have provided novel criteria for content-based video retrieval from multiple video archives where the concerns all correspond to different video semantics and the objective is to find videos relevant to all concerns. This criterion can retrieve samples which are not quickly recovered by other multiple-query retrieval methods and any straight line secularization technique. We have provided theoretical outcomes on asymptotic non-convexity of proposed methodology i.e. k-NN with matrix propagation that show that the Pareto strategy is better than using straight line mixtures of position outcomes. To demonstrate the benefits of the suggested Pareto front side method by using this experimental datasets. Furthermore, we implement our proposed approach to explore label annotation based image from web image data sources in real time environment.

References

[1] Ambareesh Ravi, Amith Nandakumar,"A multimodal deep learning framework for scalable content based visual media retrieval", arXiv:2105.08665v1 [cs.LG] 18 May 2021.

[2] Wei Weng, Yan-Nan Chen, Chin-Ling Chen, Shun-Xiang Wu, Jing-Hua Liu, Non-sparse Label Specific Features Selection for Multi-label Classification, Neurocomputing (2019),doi: https://doi.org/10.1016/j.neucom.2019.10.016.

[3] Jianghong Ma, Tommy W.S. Chow, "Label-specific feature selection and two-level label recovery for multi-label classification with missing labels", https://doi.org/10.1016/j.neunet.2019.04.0110893-6080/© 2019 Elsevier Ltd. All rights reserved.

[4] Gong, C., Tao, D., Liu, W., Liu, L., & Yang, J. (2017). Label propagation via teachingto-learn and learning-to-teach. IEEE Transactions on Neural Networks and Learning Systems, 28(6), 1452–1465. http://dx.doi.org/10.1109/TNNLS.2016.2514360.

[5] Gong, C., Tao, D., Maybank, S. J., Liu, W., Kang, G., & Yang, J. (2016). Multimodal curriculum learning for semi-supervised image classification. IEEE Transactions on Image Processing, 25(7), 3249–3260. http://dx.doi.org/10.1109/TIP.2016.2563981.

[6] Huang, J., Qin, F., Zheng, X., Cheng, Z., Yuan, Z., & Zhang, W. (2018). Learning label-specific features for multi-label classification with missing labels. In 2018 IEEE fourth international conference on multimedia big data (BigMM) (pp. 1–5). http://dx.doi.org/10.1109/BigMM.2018.8499080.

[7] Zhang, Z., Li, F., Jia, L., Qin, J., Zhang, L., & Yan, S. (2018). Robust adaptive embedded label propagation with weight learning for inductive classification. IEEE Transactions on Neural Networks and Learning Systems, 29(8), 3388–3403. http://dx.doi.org/10.1109/TNNLS.2017.2727526.

[8] Zhang, Z., Zhang, Y., Li, F., Zhao, M., Zhang, L., & Yan, S. (2017). Discriminative sparse flexible manifold embedding with novel graph for robust visual representation and label propagation. Pattern Recognition, 61, 492–510. http: //dx.doi.org/10.1016/j.patcog.2016.07.042

[9] Zhang, R., Nie, F., & Li, X. (2018). Self-weighted supervised discriminative feature selection. IEEE Transactions on Neural Networks and Learning Systems, 29(8), 3913–3918. http://dx.doi.org/10.1109/TNNLS.2017.2740341.

[10] Zhang, Z., Jia, L., Zhao, M., Liu, G., Wang, M., & Yan, S. (2018). Kernelinduced label propagation by mapping for semi-supervised classification. IEEE Transactions on Big Data.

[11] Pang, T., Nie, F., Han, J., & Li, X. (2019). Efficient feature selection via l2,0- norm constrained sparse regression. IEEE Transactions on Knowledge and Data Engineering, 1. http://dx.doi.org/10.1109/TKDE.2018.2847685

[12] Liu, H., Li, X., & Zhang, S. (2017). Learning instance correlation functions for multilabel classification. IEEE Transactions on Cybernetics, 47(2), 499–510. http://dx.doi.org/10.1109/TCYB.2016.2519683.

[13] Ma, J., & Chow, T. W. (2018a). Robust non-negative sparse graph for semisupervised multi-label learning with missing labels. Information Sciences, 422(Supplement C), 336–351. http://dx.doi.org/10.1016/j.ins.2017.08.061.

[14] Ma, J., & Chow, T. W. S. (2018b). Topic-based algorithm for multilabel learning with missing labels. IEEE Transactions on Neural Networks and Learning Systems, 1–15. http://dx.doi.org/10.1109/TNNLS.2018.2874434.

[15] Ma, J., Tian, Z., Zhang, H., & Chow, T. W. (2017). Multi-label low-dimensional embedding with missing labels. Knowledge-Based Systems, 137(Supplement C), 65–82. http://dx.doi.org/10.1016/j.knosys.2017.09.005.

[16] Bandeira, A.S.;Mixon,D.G.;Recht, B.:Compressive classification and the rare eclipse problem. arXiv:1404.3203, 2014.

[17] Oymak, S.; Recht, B.:Near-optimal bounds for binary embeddings of arbitrary sets. arXiv:1512.04433, 2015.

[18] Li, M.; Rane, S.; Boufounos, P.: Quantized embeddings of scaleinvariant image features for mobile augmented reality, in IEEE Int. Workshop on Multimedia Signal Processing, Banff, Canada, 17–19 September 2012, 1–6.

[19] Jacques, L.: Smallwidth, lowdistortions: quasi-isometric embeddings with quantized sub-Gaussian randomprojections. arXiv:1504.06170, 2015.

[20] J. Liu, Y. Lin, , Y. Li, W. Weng, S. Wu, Online multi-label streaming feature selection based on neighborhood rough set, Pattern Recognition. 84 (2018) 273õ287.

[21] J. Lee, W. Seo, J. H. Park and D. W. Kim, Compact feature subset-based multi-label music categorization for mobile devices, Multimedia Tools & Applications. (2018) 1-15.

[22] L. Liu, L. Tang, L. He, S. Yao and W. Zhou, Predicting protein function via multi-label supervised topic model on gene ontology, Biotechnology & Biotechnological Equipment. 31(1) (2017) 1-9.

[23] Y. Lin, Q. Hu, J. Liu and D. Jie, Multi-label feature selection based on max-dependency and min-redundancy, Neurocomputing. 168 (2015) 92-103.

[24] W. Weng, Y. Lin, S. Wu, Y. Li and Y. Kang, Multi-label learning based on label specific features and local pairwise label correlation, Neurocomputing. 273 (2018) 385õ394.

[25] Shiv Ram Dubey. A decade survey of content based image retrieval using deep learning. arXiv preprint arXiv:2012.00641, 2020.

[26] Afshan Latif, Aqsa Rasheed, Umer Sajid, Jameel Ahmed, Nouman Ali, Naeem Iqbal Ratyal, Bushra Zafar, Saadat Hanif Dar, Muhammad Sajid, and Tehmina Khalil. Content-based image retrieval and feature extraction: a comprehensive review. Mathematical Problems in Engineering, 2019, 2019.

[27] Yumeng Liu and Aina Sui. Research on feature dimensionality reduction in content based public cultural video retrieval. In 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS),pages 718–722. IEEE, 2018.

[28] Subhadip Maji and Smarajit Bose. Cbir using features derived by deep learning. arXiv preprint arXiv:2002.07877, 2020.

[29] Antoine Miech, Jean-Baptiste Alayrac, Lucas Smaira, Ivan Laptev, Josef Sivic, and Andrew Zisserman. End-to-end learning of visual representations from uncurated instructional videos. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 9879–9889, 2020.

[30] Alexey Potapov, Innokentii Zhdanov, Oleg Scherbakov, Nikolai Skorobogatko, Hugo Latapie, and Enzo Fenoglio. Semantic image retrieval by uniting deep neural networks and cognitive architectures. In International Conference on Artificial General Intelligence, pages 196–206.Springer, 2018.

[31] Mohsen Ramezani and Farzin Yaghmaee. Retrieving human action by fusing the motion information of interest points. International Journal on Artificial Intelligence Tools, 27(03):1850008, 2018.

[32] Hiroki Tanioka. A fast content-based image retrieval method using deep visual features. In 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW), volume 5, pages 20–23. IEEE,2019.

[33]and Fang Huang. Cnn-vwii: An efficient approach for large-scale video retrieval by image queries. Pattern Recognition Letters, 123:82–88, 2019.

[34] Wengang Zhou, Houqiang Li, Jian Sun, and Qi Tian. Collaborative index embedding for image retrieval. IEEE transactions on pattern analysis and machine intelligence, 40(5):1154–1166, 2017.

[35] Lei Zhu, Jialie Shen, Liang Xie, and Zhiyong Cheng. Unsupervised visual hashing with semantic assistant for content-based image retrieval. IEEE Transactions on Knowledge and Data Engineering, 29(2):472–486, 2016.

[36] Mohammadreza Zolfaghari, Kamaljeet Singh, and Thomas Brox. Eco: Efficient convolutional network for online video understanding. In Proceedings of the European conference on computer vision (ECCV),pages 695–712, 2018.

Subramanyam Kunisetti received M.tech degrees in Computer Science from University of Hyderabad in 2009.currently working as Assistant Professor in R.V.R & J.C College of Engineering.ORCID ID is 0000-0003-3241-9888.