iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 Issue: VIII Month of publication: August 2015

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 3 Issue VIII, August 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Face Annotation System for Images and Videos Search

Rabitha N[1], Sruthy Manmadhan[2]
[1,2]Dept of Computer Science & Engineering,
NSS College of Engineering, Palakkad, Kerala

Abstract— This paper combining a labeled facial recognition dataset with face images extracted from videos on YouTube and face images returned from using a search engine. The web search engine and the video search engine can be viewed as very weak alternative classifier which provides weak labels." Using the results from these two different types of search queries as forms of weak labels, a robust method for classification can be developed. This method is based on graphical models but also encorporates a probabilistic margin. Labeled data may also contain weak sources of information that may not necessarily be used to maximum effect. For example, facial image datasets for the tasks of performance driven facial animation, emotion recognition, and facial key-point or landmark prediction often contain alternative labels from the task at hand. In emotion recognition data, for example, emotion labels are often scarce. This may be because these images are extracted from a video, in which only a small segment depicts the emotion label. As a result, many images of the subject in the same setting using the same camera are unused. However, this data can be used to improve the ability of learning techniques to generalize to new and unseen individuals by explicitly modelling previously seen variations related to identity and expression. Once identity and expression variation are separated, simpler supervised approaches can work quite well to generalize to unseen subjects. In many cases in facial images, sources of information may be available that can be used to improve tasks. This includes weak labels which are provided during data gathering, such as the search query used to acquire data, as well as identity information in the case of many experimental image databases.
Keywords— Content Based Image Retrieval, SBFA, K-means algorithm, ULR

I. INTRODUCTION

In addition, the problem of machine recognition of human faces continues to attract researchers from disciplines such as image processing, pattern recognition, neural networks, computer vision, computer graphics, and psychology. The strong need for user-friendly systems that can secure our assets and protect our privacy without losing our identity in a sea of numbers is obvious. At present, one needs a PIN to get cash from an ATM, a password for a computer, a dozen others to access the internet, and so on. Although very reliable methods of biometric personal identification exist, for example, fingerprint analysis and retinal or iris scans, these methods rely on the cooperation of the participants, whereas a personal identification system based on analysis of frontal or profile images of the face is often effective without the participant's cooperation or knowledge. Commercial and law enforcement applications of FRT range from static, controlled-format photographs to uncontrolled video images, posing a wide range of technical challenges and requiring an equally wide range of techniques from image processing, analysis, understanding, and pattern recognition. One can broadly classify FRT systems into two groups depending on whether they make use of static images or of video. Within these groups, significant differences exist, depending on the specific application. The differences are in terms of image quality, amount of background clutter (posing challenges to segmentation algorithms), variability of the images of a particular individual that must be recognized, availability of a well-defined recognition or matching criterion, and the nature, type, and amount of input from a user.

A general statement of the problem of machine recognition of faces can be formulated as follows: given still or video images of a scene, identify or verify one or more persons in the scene using a stored database of faces. Available collateral information such as race, age, gender, facial expression, or speech may be used in narrowing the search (enhancing recognition). The solution to the problem involves segmentation of faces (face detection) from cluttered scenes, feature extraction from the face regions, recognition, or verification. In identification problems, the input to the system is an unknown face, and the system reports back the determined identity from a database of known individuals, whereas in verification problems, the system needs to confirm or reject the claimed identity of the input face.

Face perception is an important part of the capability of human perception system and is a routine task for humans, while building a similar computer system is still an on-going research area. The earliest work on face recognition can be traced back at least to the 1950s in psychology and to the 1960s in the engineering literature. Some of the earliest studies include work on facial expression of emotions by Darwin. But research on automatic machine recognition of faces really started in the 1970s and after the seminal work of Kanade. Over the past 30 years extensive research has been conducted by psychophysicists, neuroscientists, and engineers on various aspects of face recognition by humans and machines. Psychophysicists and neuroscientists have been concerned with issues such as whether face perception is a dedicated process (this issue is still being debated in the psychology community and whether it is done holistically or by local feature analysis. Many of the hypotheses and theories put forward by researchers in these disciplines have been based on rather small sets of images. Nevertheless, many of the findings have important consequences for engineers who design algorithms and systems for machine recognition of human faces.

II. RELATED WORKS

A. Clustering Algorithm With Possibility Model

Berg et al. [1] presents the combination of a possibility model with a clustering algorithm. This combination is to present the relationship between the facial images and the names in their captions for the facial images and the detected names in the same document. The simply clustering method is used to captioned new images and automatically link name. For improving the performance clustering process are combine with possibility model .combining this two methods there are accurate labeled set of faces. The result of this work shows that by analyzing language carefully can produce much better clustering and also learn a natural language classifier to determine who is pictured from text alone. This method works on particular data set it further more improve for free text on webpage using simple image representation and context model.

B. Graph Based Approach

Ozkan and Duygulu [2] proposed a graph-based model for finding the densest sub-graph as the most related result. The proposed method is to associate names and faces for querying people in large news photo collection. In most cases the number of same faces of queried person will be large so the faces are more similar to each other. They proposed the graph based method to find the similar subset with possible set of faces with query person name. Similarity are represent by SIFT describers. Then apply a greedy graph algorithm. Guillaumin et al.[3] introduced a modification to incorporate the constraint that a face is only depicted once in an image. There are two scenarios of naming persons in database for finding face of person and assigning name to all faces. The text based result is not greatly improved. To improve a resent graph based approach introduce the constraints when optimizing the objective function generative models have previously been proposed to solve the multi-person naming task .by comparing generative and graph based methods the most significant method is graph based method .in future extends the graph based method to multi-person naming Guillaumin et al. [3] proposed to iteratively update the assignment based on a minimum cost matching algorithm. In their follow-up work Guillaumin et al. [4], they further improve the annotation performance by using distance metric learning techniques to gain more distinguish feature in low dimension space.

C. Query Expansion

T. Mensink and J.J. Verbeek [5], by using ideas from query expansion the performance of name-based scheme can be further improved. In this paper they are interested to finding images of people on the web and more clearly labeled the new images. The text base initial results are not perfect. The performances are depending on the assumptions. To improve such poor performance proposed “query expansion”. They applied this idea on early proposed method on which filter the initial result set. Using Gaussian mixture modelling and logistic discriminant model the query expansion is improving the performance in both of method. The research suggest the model learned from caption based supervision

D. Purify Web Facial Images

This aims to correct noisy web facial images for face recognition applications [6], [7]. These works are proposed as a simple pre-processing step in the whole system without adopting sophisticated techniques. T.L. Berg, A.C. Berg et al. [6] applied a modified k means clustering approach for cleaning up the noisy web facial images. Zhao et al. [10] propose system that can learn and recognize face by combining weakly labeled text, image and video. Consistency learning proposed to create face model for popular person .the text images on the web as a weak signal of relevance and learn consistent face model from large and noisy training sets. Effective and accurate face detection and tracking is applied. Lastly key faces are selected by clustering to get compact and robust representation. The effective ness is increase due to represent key face and removes duplicate key face. They used the unsupervised machine learning techniques and propose a graph-based label refinement algorithm to optimize the label quality over the whole retrieval database. Z. Wu, Q. Ke, J. Sun, and H.-Y. Shum [8] mainly addressed the face retrieval problem, by using local and global features which propose an effective image representation. Future Work is to design a supervised learning algorithm to automate this process to further improve the visual word vocabulary for face. This system is highly scalable, and they plan by using a computer cluster to apply on a web-scale image database.

E. Retrieval Based Face Annotation

D. Wang, S.C.H. Hoi, Y. He. And J. Zhu [9] the WLRLCC algorithm is focused on learning more features for the top retrieved facial images for each query. By weak label regularized local coordinate coding. Retrieval based face annotation is used in mining massive web facial images for automatic face annotation .there are two challenges first is how effectively retrieve most of similar facial images. Second is how to effectively perform annotation. They proposed weak label regularised local coordinate coding (WLRLCC) technique. They also proposed the optimization algorithm i.e.WLRLCC algorithm .This algorithm boosts the performance of the retrieval based face annotation approach on a large scale web facial image D. Wang, S.C.H. Hoi, and Y. He et al. [10] this proposed system investigated a unifying learning scheme by combining both transductive and inductive learning technique to mine web facial images for face annotation. They proposed Weak label Laplacian support vector machine (WL-LapSVM) algorithm by adopting WLRLCC algorithm.

F. Search Based Face Annotation

Dayong Wang, Steven C.H. Hoi et al. [11] Propose an effective unsupervised label refinement for refining the web facial images. For improving the performance they also propose optimization algorithm to solve large-scale learning effectively i.e. clustering based approximation the propose system improve the performance of search based face annotation scheme. The work are different form all previous work by two things. To solve general content based face annotation problem using search based where face image as query image. They unsupervised label refinement algorithm which enhanced new label matrix. This work also related recent work of the WIRLCC method [9].The unified learning scheme [10].Adopted locality sensitive hashing [12]. Adopted unsupervised face alignment technique [16].extract the GIST features [14].

III.PROPOSED SYSTEM

A. Face Annotation In Images

Investigate and implement a promising search based face annotation scheme by mining large amount of weakly labeled facial images freely available on the WWW. A novel method is proposed to enhance the label quality via a graph-based and low-rank learning approach. An efficient clustering-based approximation algorithm for large-scale label refinement problem is defined. Conducted an extensive set of experiments, in which encouraging results were obtained.

The system flow of the planned framework of search-based face annotation that consists of the subsequent steps:

Facial image data collection;

Face detection and facial feature extraction;

High-dimensional facial feature indexing;

Learning to refine weakly labeled data;

Similar face retrieval

Face annotation by majority voting on the similar faces with the refined labels.

The first step is that the information assortment of facial pictures, during which crawled a group of facial pictures from the WWW by an existing internet program (i.e., Google) in line with a reputation list that contains the names of persons to be collected. The process output of this retrieval process, shall get a group of facial pictures, every of them is related to some human names. The second step is to pre-process internet facial pictures to extract face-related data, together with face detection and alignment, facial region extraction, and facial feature illustration. The GIST feature extraction to represent the extracted faces. As a result, every face is diagrammatic by a d-dimensional feature vector.

The third step is to index the extracted options of the faces by applying some economical high-dimensional assortment technique to facilitate the task of comparable face retrieval within the succeeding step. In this approach, adopt the neighbourhood sensitive hashing (LSH), an awfully well-liked and effective high-dimensional assortment technique. The process of face annotation could take a look at part for the given question facial image for annotation, initial conduct the same face retrieval method to look for a set of most similar faces (typically prime K similar face examples) from the antecedently indexed facial information. The set of prime K similar face examples retrieved from the information, subsequent step is to annotate the facial image with a label (or a set of labels) by using a majority choice approach that mixes the set of labels related to these prime K similar face examples. Focus attention on one key step of the higher than framework, i.e., the unsupervised learning method to refine labels of the weak labeled facial pictures.

1) Clustering Based Approximation Algorithm: The proposed system uses clustering-based approximation algorithms to speed up the solutions for large-scale problems. Clustering strategy could be applied in two different levels: The "image-level," which can be used to directly separate all the n facial images into a set of clusters. The other is on "name-level," which can be used to First separate the m names into a set of clusters, then to further split the retrieval database into different subsets according to the name-label clusters.

i. K-MEANS Clustering Algorithm: K-MEANS is a prototype-based, simple partitioned clustering technique which attempts to find a user-specified k number of clusters. These clusters are represented by their centroids. A cluster centroid is typically the mean of the points in the cluster. This algorithm is simple to implement and run, relatively fast, easy to adapt, and common in practice. The algorithm consist of two phases: the first phase is to define k centroids, one for each cluster. The next phase is to take each point belonging to the given data set and associate it to nearest centroid.

The k-means algorithm works as follows:

1. Select initial centroid of the k clusters. Repeat steps b through c until the cluster membership stabilized.
2. Generate a new partition by assigning each data to its closest cluster centroid.
3. Compute new cluster centroid for each cluster.

The most widely used convergence criteria for the k-means algorithm is minimizing the SSE.

$$SSE = \sum_{j=1}^{k} \sum_{x_i \in c_j} \|x_i - \mu_j\|^2 \; Where \; \mu_j = \frac{1}{n_j} \sum_{x_i \in c_j} x_i$$

Denotes the mean of cluster cj and nj denotes the no. of instances in cj. The k-means algorithm always converges to a local minimum. The particular local minimum found depends on the starting cluster centroids. The k-means algorithm updates cluster centroids till local minimum is found. Before the k-means algorithm converges, distance and centroid calculations are done while loops are executed a number of times, say l, where the positive integer l is known as the number of k-means iterations. The precise value of l varies depending on the initial starting cluster centroids even on the same dataset. So the computational complexity of the algorithm is O(nkl), where n is the total number of objects in the dataset, k is the required number of clusters and l is the number of iterations. The time complexity for the high dimensional data set is O(nmkl) where m is the number of dimensions.

2) Locality Sensitive Hashing Algorithm: Index the extracted features of the faces by applying some efficient high-dimensional indexing technique to facilitate the task of similar face retrieval in the subsequent step. Indexing is done using LSH functions and by building several hash tables to increase the probability of collision for close points. At query time, the KNN search is performed by hashing the query point to one bucket per hash table and then to rank all discovered objects by their distance to the query point. The closest K points are returned as the final result.

B. Video Frame Extraction

Different Key frame extraction techniques is to

Compute the frame differences based on some criteria and then discard the frames whose difference with the adjacent frames is less than a certain threshold. Various low level features have been applied for this purpose including color histograms, frame correlations, edge histogram, etc.[15]

For instance, Pal and Leigh [16] used fuzzy geometrical and information measures to develop an algorithm to estimate the difference between two consecutive frames. The similarity between the frames was measured in terms of weighted distance in fuzzy feature space. Hanjalic et al. compared the difference in color histograms of consecutive frames with a threshold to obtain key frames.

Introduced by DeMenthon et al. [17]. The key frames were extracted by finding discontinuities on a trajectory curve, which represent video sequence.

In the Flexible Rectangles algorithm [18],the frame differences were used to build a ‘‘Content Development Curve’’ from a curve composed of a predefined number of rectangles through the use of an error minimization algorithm.

The Adaptive Sampling algorithm [19] extracted key frames by uniformly sampling the y-axis of the curve of cumulative frame differences .The resulting sampling on the x-axis represented the key frames.

The Shot Reconstruction Degree Interpolation [20] selected the key frames based on the ability of frames to reconstruct the original video shot using frame interpolation.

Ciocca and Schettini[21] extracted key frames by first finding the cumulative frame differences based on certain frame descriptors such as color histogram, histogram of edges and wavelets. Next, a curve of cumulative frame differences was sketched, and then the midpoints of two curvature points on this curve were selected as key Prof.Archana.V.Potnurwar et al. frames. A curvature point is a point on the curve where the angle changes are drastic. The frame difference based methods are intuitive and simple in nature. These properties make them suitable for many real-time and/or online applications. However, for extracting a particular key frame, these techniques only consider sufficient content change between the consecutive frames (or between current frame and last key frame). Therefore, a key frame that is extracted by these methods does not fully represent the portion of the video preceding it.

Some researchers used clustering for extracting key frames by treating video frames as points in the feature space.

The core idea behind such techniques is to cluster the frames based on some similarity measure and then select one key frame from each cluster. Yeung and Yeo [22] proposed a method to generate a pictorial summary of a video sequence which consists of a set of video posters, each representing a scene in the sequence. The key frames were extracted using a time-constrained clustering method which takes into account both visual properties and temporal locality of the shots. The video posters were generated by combining key frames based on their dominant scores assigned during the clustering phase. Zhuang proposed a technique for key frame extraction based on unsupervised clustering using a color histogram as the visual content. A node is added to a cluster only if the similarity measure between the frame and the cluster centroid is greater than a certain threshold.

Doulamis et al. [23] presented a technique for summarizing stereoscopic videos which used clustering of shots to reduce redundancy. The clustering was performed based on the multidimensional fuzzy classification of segment features extracted from stereoscopic frames.

A motion based clustering algorithm was introduced by Zhang et al. [24] in which the clustering was done based on the motion compensation error.

Furini et al. [25] proposed a summarization technique called ‘‘STIMO’’ (ST Ill and MOving Video Storyboard) based on the clustering of HSV color descriptors. Avilaet al. [26] presented a method ‘‘VSUMM’’ (Video Summarization) which extracted color features from the frames after pre-sampling the frames from video. After removal of useless frames, the rest of the frames are clustered based on the k-means clustering algorithm.

The main advantage of clustering based methods is that they generate less redundant summaries as compared to the consecutive frame difference based techniques. The problem with most of the clustering methods (less time constrained clustering) is that temporal information of the frames is not considered. In other words, the key frames are selected regardless of the temporal order of each frame. Therefore, the key frames may not preserve the temporal visual sequence of the video. The proposed system is a web application in that proposes an efficient clustering-based approximation algorithm to solve the problem of large scale label purification. It also includes in this the hit count of the particular image that how many times that image will be searched.

"Content-based" means that the search will analyse the actual contents of the image rather than the metadata such as keywords, tags, and/or descriptions associated with the image. The term 'content' in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself. CBIR is desirable because most web based image search engines rely purely on metadata and this produces a lot of garbage in the results. Also having humans manually enter keywords for images in a large database can be inefficient, expensive and may not capture every keyword that describes the image. Thus a system that can filter images based on their content would provide better indexing and return more accurate results. There is a growing interest in CBIR because of the limitations inherent in metadata-based systems, as well as the large range of possible uses for efficient image retrieval. Textual information about images can be easily searched using existing technology, but requires humans to personally describe every image in the database. This is impractical for very large databases, or for images that are generated automatically, e.g. from surveillance cameras. It is also possible to miss images that use different synonyms in their descriptions. Systems based on categorizing images in semantic classes like "cat" as a subclass of "animal" avoid this problem but still face the same scaling issues.

IV. CONCLUSIONS AND FUTURE WORK

The proposed system presents an extensive search on face annotation techniques for web facial images and videos. This framework focused on tackling the problems of existing system annotation. Improving the quality of labeling and proposed an algorithm. To improve the performance it also proposed Clustering Based Approximation which gives result without introducing much performance degradation. This proposed system investigated a promising search-based face annotation framework, in which it focused on improving the performance of system. The quality of labeled of images are improved using an algorithm. To further improve the performance, it also proposed a clustering based approximation solution, for reducing the performance evolution.

Future work will focus on further automating the system and performing additional analysis of annotation results. The easiest images typically involve clearly visible faces without any occlusions. Difficult images generally involved large crowds, lower resolutions persons in the background, or ambiguous cases such as pictures on the wall or persons on television screens. Color based image segmentation; it is possible to reduce the computational cost avoiding feature calculation for every pixel in the image. Although the color is not frequently used for image segmentation, it gives a high discriminative power of regions present in the image. This kind of image segmentation may be used for mapping the changes in land use land cover taken over temporal period in general but not in particular.

The proposed algorithm effectively reduces overall time complexity. The proposed algorithm number of non-relevant effectively minimizes the undesirable results and gives a good relevance percentage by giving minimum images.

V. ACKNOWLEDGEMENT

The authors would like to thank professors of NSSCE, Palakkad for suggestions and support on this paper.

REFERENCES

[1] T.L. Berg, A.C. Berg, J. Edwards, and D. Forsyth, “Who’s in the Picture,” Proc. Neural Information Processing Systems Conf.(NIPS),2005.

[2] D. Ozkan and P. Duygulu, “A Graph Based Approach for Naming Faces in News Photos,” Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR), pp. 1477-1482, 2006.

[3] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, “Automatic Face Naming with Caption-Based Supervision,” Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2008.

[4] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, “Face Recognition from Caption-Based Supervision,” Int’l J. Computer Vision, vol. 96, pp. 64-82, 2011.

[5] T. Mensink and J.J. Verbeek, “Improving People Search Using Query Expansions,” Proc. 10th European Conf. Computer Vision (ECCV), vol. 2, pp. 86-99, 2008.

[6] T.L. Berg, A.C. Berg, J. Edwards, M. Maire, R. White, Y.W. Teh, E.G. Learned-Miller, and D.A. Forsyth, “Names and Faces in the News,” Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR), pp. 848-854, 2004.

[7] M. Zhao, J. Yagnik, H. Adam, and D. Bau, “Large Scale Learning and Recognition of Faces in Web Videos,” Proc. IEEE Eighth Int’l Conf. Automatic Face and Gesture Recognition (FG), pp. 1-7, 2008.

[8] Z. Wu, Q. Ke, J. Sun, and H.-Y. Shum, “Scalable Face Image Retrieval with Identity-Based Quantization and Multi-Reference Re-Ranking,” Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 3469-3476, 2010.

[9] D. Wang, S.C.H. Hoi, Y. He, and J. Zhu, “Retrieval-Based Face Annotation by Weak Label Regularized Local Coordinate Coding,” Proc. 19th ACM Int’l Conf. Multimedia (Multimedia), pp. 353-362, 2011.

[10] D. Wang, S.C.H. Hoi, and Y. He, “A Unified Learning Framework for Auto Face Annotation by Mining Web Facial Images,” Proc.21st ACM Int’l Conf. Information and Knowledge Management (CIKM), pp. 1392-1401, 2012.

[11] Dayong Wang, Steven C.H. Hoi, Ying He, and Jianke Zhu,” Mining Weakly Labeled Web Facial Images for Search-Based Face Annotation” IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 1, January 2014

[12] W. Dong, Z. Wang, W. Josephson, M. Charikar, and K. Li, “Modeling LSH for Performance Tuning,” Proc. 17th ACM Conf. Information and Knowledge Management (CIKM), pp. 669-678, 2008.

[13] J. Zhu, S.C.H. Hoi, and L.V. Gool, “Unsupervised Face Alignment By Robust Non rigid Mapping,” Proc. 12th Int’l Conf. Computer Vision (ICCV), 2009.

[14] C. Siagian and L. Itti, “Rapid Biologically-Inspired Scene Classification Using Features Shared with Visual Attention,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol.29, no. 2,pp. 300-312, Feb. 2007.

[15] S.E.D. Avila, A.B.P. Lopes, L.J. Antonio, A.d.A. Araújo, VSUMM: a mechanism designed to produce static video summaries and a novel evaluation method, Pattern Recognition Letters 32 (1) (2011) 56–68.

[16] A. Hanjalic, R.L. Langendijk, J. Biemond, A new key-frame allocation method for representing stored video-streams, in: 1st International Workshop on Image Databases and Multimedia, Search, 1996, pp. 67–74.

[17] A. Hanjalic, R.L. Langendijk, J. Biemond, A new method for key frame based video content representation, Image Databases and Multimedia Search, World Scientific, Singapore, 1998.

[18] S.H. Hoon, K. Yoon, I. Kweon, A new technique for shot detection and key frames selection in histogram space, in: 12th Workshop on Image Processing and

Image Understanding, 2000, 217–220.
[19] L.M. Tieyan, X. Zhang, J. Feng, K.T. Lo, Shot reconstruction degree: a novel criterion for key frame selection, Pattern Recognition Letters 25 (12) (2004) 1451–1457.
[20] Y. Zhuang, Y. Rui, T.S. Huang, S. Mehrotra, Adaptive key frame extraction using unsupervised clustering, in: International Conference on Image Processing, 1998, 866-870.
[21] G. Ciocca, R. Schettini, Innovative algorithm for key frame extraction in video summarization, Journal of Real-Time Image Processing 1 (1) (2006) 69–88.
[22] R.M. Jiang, A.H. Sadka, D. Crooks (Eds.), Advances in video summarization andskimmingM. Grgic, K. Delac, M. Ghanbari (Eds.), Recent Advances in Multimedia Signal Processing and Communications, 231, Springer Berlin, Heidelberg, 2009, pp. 27–50.
[23] N.D. Doulamis, A.D. Doulamis, Y.S. Avrithis, K.S. Ntalianis, S.D. Kollias, Efficient summarization of stereoscopic video sequences, IEEE Transactions on Circuits and Systems for Video Technology 10 (4) (2011) 501–517.
[24] X.D. Zhang, T.Y. Liu, K.T. Lo, J. Feng, Dynamic selection and effective compression of key frames for video abstraction, Pattern Recognition Letters 24 (9–10) (2003) 1523–1532.
[25] M. Furini, F. Geraci, M. Montangero, M. Pellegrini, STIMO: STIll and moving video storyboard for the web scenario, Multimedia Tools and Applications 46 (1) (2010) 47–69.
[26] S.E.D. Avila, A.B.P. Lopes, L.J. Antonio, A.d.A. Araújo, VSUMM: a mechanism designed to produce static video summaries and a novel evaluation method, Pattern Recognition Letters 32 (1) (2011) 56–68.