Research Article

Video Retrieval System for Meniscal Surgery to Improve Health Care Services

Sana Amanat,[1] Muhammad Idrees,[2] Muhammad Usman Ghani Khan,[1] Zahoor Rehman [iD],[3] Hangbae Chang,[4] Irfan Mehmood [iD],[5] and Sung Wook Baik[5]

[1]*Department of Computer Science and Engineering, University of Engineering and Technology Lahore, Lahore, Pakistan*
[2]*Department of Computer Science and Engineering, University of Engineering and Technology Lahore Narowal Campus, Lahore, Pakistan*
[3]*Department of Computer Science, Comsats Institute of Information Technology Attock Campus, Attock, Pakistan*
[4]*Department of Industrial Security, College of Business and Economics, Chung-Ang University, Seoul, Republic of Korea*
[5]*Department of Software, Sejong University, Seoul, Republic of Korea*

Correspondence should be addressed to Irfan Mehmood; irfanmehmood@ieee.org

Meniscal surgery is considered the most general orthopedic process that deals with the treatment of meniscus tears for human health care. It leads to a communal contusion to the cartilage that stabilizes and cushions the knee joints of human beings. Such tears can be classified into different categories based on age group, region, and occupation. Further, a large number of sportsmen and heavy weightlifters even in developed countries are affected by meniscus injuries. These patients are subjected to arthroscopic surgery, and during surgical treatment, the perseverance of meniscus is a very crucial task. Current research provides a significant ratio of meniscal tear patients around the globe, the critical expanse is considered as having strikingly risen with a mean annual of 0.066% due to surgery failure. To decumbent this ratio, an innovative training mechanism is proposed through video retrieval system in this research. This research work is focussed on developing a corpus and video retrieval system for meniscus surgery. Using the proposed system, surgeons can access guidance by watching the videos of surgeries performed by an expert and their seniors. The proposed system is comprised of four approaches to the spatiotemporal methodology to improve health care services. It entails key point, statistical modeling, PCA-scale invariant feature transform (SIFT), and PCA-Gaussian mixture model (GMM) with a combination of sparse-optical flow. The real meniscal surgery dataset is used for testing purposes and evaluation. The results conclude that using PCA-SIFT approach improves the results with an average precision of 0.78.

1. Introduction

A meniscus tear is a public discolouration of the cartilage that alleviates and pads the knee joints of human beings. This is common malfunctioning across the globe, with a mean annual rate of 0.066%. It is noteworthy to mention that a large number of meniscus injury incidences are observed at about 66 per 100,000 around the globe. These incidences can be categorized under various aspects, that is, age group, region, and occupation. Most of the sportsmen and heavy weightlifters affected with age less than 40 years in developed countries recovered using meniscus surgery. They cover the areas of America, Latin, Asia, and Africa. On the other hand in Denmark, more degenerative changes in the annual incidence are observed. These changes are increasing from 164 to 312 over the past 10 years and about 2 million cases observed in meniscus surgery treatment. Details of registered patients, diagnosis, and procedures maintained by Danish National Patient Register can be found in [1] and Summit (2016).

Previous studies have reported a very high rate of repetitive surgery due to the failure of meniscus surgery. Different causes are observed for the failure of meniscus surgery. One of the major causes is the inappropriate removal of meniscus

because unexperienced surgeons do not know how much or which part has to be removed properly. More precisely, it deals with the interaction of tissues and radiation, which may be represented in the form of a static image and video or recordings. These recordings can be used to train surgeons and practitioners to perform successful surgeries. One way to access these recordings is through video retrieval system that aims to retrieve videos from a source of the multimedia repository. Upon surgeon's request and need, video retrieval system can provide the most relevant and accurate videos of meniscal surgery. Likewise, there have been several studies in relation to surgery, that is, for the eyes, lungs, heart, and kidney, but meniscal surgery holds a unique position since it is a commonality in most of the countries and in literature neither corpus nor video retrieval systems precisely for meniscal surgery yet found [2].

To retrieve retinal surgery videos using compressed video streams, the required motion info is extracted at regular intervals. Precisely, two types of motion information are excerpted: one is region trajectory and the other encompasses the residual error. A heterogeneous feature vector is used to combine the motion information, and sequence segmentation is utilized to extract the motion information. The Kalman filter and generalized Gaussian distribution are utilized to formulate region trajectory and residual information. In last step, extension of fast dynamic time warping is also used along with feature vector to compare videos [3].

Another approach to retrieve medical videos is a key frame which is initially extracted and its global features are excerpted, which has to be saved in XML script format in the form of an object to reduce the storage space. Furthermore, neural network and naïve Bayes are used to enhance the performance. Features of an image are extracted using entropy, Huffman coding, and fast Fourier transforms [4].

In order to retrieve laparoscopy videos from a colossal repository of videos, a feature signature is utilized which is composed of abstract and local details of the image. Initially, position, colour, and texture features of an image are considered for this signature. Weight is assigned to each feature describing the potential to be used as a feature signature representative. For comparison between two signatures, earth mover's distance, signature quadratic form distance, and signature matching distance are used in combination [5]. To retrieve endoscopy videos from repository, feature signature is computed based on global features, that is, colour, position, and texture. Weight is assigned to each feature, and zero-weighted features are considered as irrelevant for signature. Those features have weighted value unequal to zero denoted as signature representatives. Furthermore, cluster centroid of these features is used as feature representative. To measure the similarity between endoscopic images, signature matching distance and adaptive binning feature signature are used by using the visual characteristics [6].

In another approach to retrieve endoscopic videos, local, global, and feature fusion techniques are employed. To extract features, various algorithms CEDD, SURF, and SIMPLE are used. Additionally, early and late fusion approaches are also considered. These approaches are tested on 1276 video clips. LIRE software library is used to index

test frames also capable to extract 20 visual features of an image [7].

To retrieve cataract surgery videos from the repository, a video is divided into a number of overlapping clips. Each clip contains the same number of frames. Various features are extracted from these frames using histogram of oriented gradients (HOG), BoVW, and motion histogram. These features are statistically modeled using the Bayesian network, CRFs, and HMMs. Thirty cataract surgery videos are used for the testing purpose having 720×576 pixels. A mean area under ROC of 0.61 was achieved in recognizing surgical steps from video stream [8]. In order to retrieve and analyse female reproductive system through hysteroscopy, a video summarization framework is proposed. To compute feature set, colour information is extracted from key frames using COC colour model. Other salient features are also extracted such as texture, motion, curvature, and contrast [9].

Another approach to retrieve cataract videos from the source that formulate a feature set is by using global features such as colour, texture, and motion information of the frame. Furthermore, the wavelet transform is performed on these extracted channels of colour. In this study, cataract and epiretinal surgeries are considered [10]. To retrieve similar videos in the domain of retinal surgery, a motion-based approach has been found in literature which utilized MPEG-4 to extract features from the frame. For video encoding, a group of pictures and macroblocks are used where three types of frames are considered. Motion histogram, classification, and motion trajectories are used to formulate motion feature set. Furthermore, signatures are also composed and to measure their ratio of similarity, fast dynamic time warping approach is used. This approach is applied to epiretinal membrane surgery.

In order to control the failure rate of surgery and to overcome the mentioned drawbacks, there is a dire need for a sophisticated approach. It could be possible via the fabrication of a connexion between information technology and medical sciences. Information technology is capable of providing medical diagnosis, clinical analysis, and treatment planning through a practice known as medical imaging [11, 2, 12].

Our proposed system firstly focuses on spatial features, that is, the location of objects in reference to static objects [13]. Secondly, it deals with the temporal information of an image, which provides a glimpse of moving and operational objects [14]. This system is comprised mainly of four methods, that is, interest point, statistical modeling, PCA-SIFT, and PCA-GMM-based video retrieval. The first method incorporates interest points of videos which are excerpted using up-to-the-minute feature explication and indicator algorithms known as SIFT, HOG, and covariant feature detector (CFD) with distance transform (DT) addressed by Bai et al. [15] and Bharathi et al. [16]. The second method encompasses statistical modeling of features using GMM and Kullback-Leibler divergence. Moreover, in the third and fourth method, dimension reduction technique is integrated with SIFT and GMM, respectively. To extract the temporal features, Lucas-Kanade sparse-optical flow is incorporated in each method [17]. This proposed system can help new

and less-experienced surgeons to get appropriate information about surgical methods, procedures, and tools to perform a successful surgery [11, 2, 12]. These meniscal recordings can additionally be used for the explanation of the patients and their follow-up sessions [9].

Table 1 illustrates a summarized comparison of previous approaches to retrieve videos. It is clear from this comparison that there are so many systems for medical video retrieval, but a video retrieval system for meniscal surgery practitioners is not found in spite of its need around the globe. In this system, we also have used a hybrid of state-of-the-art techniques for feature detection, recognition, and comparison. The most important point that can be concluded in this summarization is a step towards new research direction.

2. Materials and Methods

This section sheds light on the proposed methodology for meniscal surgery video retrieval. The surgical video is considered to be composed of frames which are chronologically bound together to define a tale, that is, live surgical process. Each frame grasps an instant of surgery, producing a coherent association with contiguous frames. Algorithms in this framework are applied to the frame instead of a video because the frame is a building block of a video. Therefore, these videos are decomposed into frames and shot boundary is detected using XOR. As a result, the collections of master frames depicting master shot are extracted. These master frames are further utilized to model spatiotemporal features of a particular video. Detail of spatiotemporal feature extraction and modeling technique is as follows.

2.1. Spatiotemporal Modeling Using Interest Points. The interest point represents the identification of interest or key points within an image used for further processing. A description of the interest point is as follows:

(i) It has a demarcated location and mathematical description of an image.

(ii) Local image organization is high in the context of local information content.

(iii) In the image domain, it has steady trepidation of both local and global.

2.2. SIFT and CFD-Based Approach. SIFT is used to detect interest points using various scales. With the help of Gaussian filters, the change in succeeding Gaussian-blurred images is figured out to excerpt interest points [17]. A difference of Gaussian (DoG) image $\text{DoG}(a, b, \Delta)$ of an original image $\text{Im}(a, b)$ is given by

$$\text{DoG}(a, b, \Delta) = L(a, b, L_i\Delta) - L(a, b, L_j\Delta), \quad (1)$$

where $L(a, b, L\Delta)$ describes the convolution operation of the original image $\text{Im}(a, b)$ with Gaussian blur $G(a, b, L\Delta)$, at scale $L\Delta$. [19].

CFD uses an affine adaptation to approximate the figure of an affine region in order to compute covariant features of an image. With the help of affine adaptation, interest points can be calculated as follows [10]:

$$\text{Cfd}(U) = \begin{bmatrix} \text{De}_{aa}(U) & \text{De}_{ab}(U) \\ \text{De}_{ab}(U) & \text{De}_{bb}(U) \end{bmatrix}, \quad (2)$$

where $\text{De}_{aa}(U)$ is defined as the 2nd fractional derivative in the direction a and $\text{De}_{ab}(U)$ represents the mix fractional 2nd derivative in directions of a and b. These computed features of input and target images are matched using Euclidean distance to get the value of similarity [20]. These calculated distances are sorted to make the ranked list based on similarity of input video with target videos.

2.3. HOG and Distance Transformation (DT). The HOG feature detector uses edge directions to compute local object appearance and shape within an object.

Each image is split into small cells, and histogram of gradient directions is computed for all cells.

In order to improve accuracy, contrast normalization of local histograms is computed by measuring the intensity across image block. Dalal-Triggs method can be used for normalization such as [3].

$$f = \frac{\mathbf{N}}{\sqrt{\|N\|_2^2 d^2}}. \quad (3)$$

All histograms in a given block are represented by a vector $\mathbf{N}$, which is not normalized, a small constant value is represented by d and $\|N\|k$ be its k-norm for $k = 1, 2$ [21].

DT [3] descriptor is also called distance map which labels the pixels by measuring distance with the support of Canny's edge detector. A framework of key point base is present in Figure 1.

2.4. Statistics-Based Framework for Video Retrieval. Statistics based technique is an approach to scrutinize condense and to interpret data. For this purpose, we have used GMM with Kullback-Leibler divergence to model the spatiotemporal content of videos. Following is the description of a method.

2.5. Spatiotemporal Modeling through GMM. The surgical footage is an object which contains highly complex multidimensional contents based on time series folded data. Assume that a video V is an arrangement of frames and has length l, $\{v_1, \ldots, v_n\}$, with every frame v_i, providing a sequential pattern of an arrangement.

A video frame is described by each element v_i, and its spatial features are modeled in d-dimensional space. At this phase, i_s presents the spatial features of frame x_i in d dimension feature space. Spatial features of a frame constitute facts of locality with the magnitude of features, for example, colours of every pixel (red, green, and blue), and collected objects may be easily traced. In order to model RGB colour distribution, pixel position can be stated as shown in Figure 2.

A spatiotemporal probabilistic model can be constructed through statistical observation of samples by using time and space features. The sampling of the video is performed at a

TABLE 1: Comparison of proposed methodology with existing ones.

| Reference papers	Feature set	Feature descriptor	Similarity/cluster/classification measurement	Dataset	Research domain	Video type
Droueche et al. [18]	A motion vector, residual information	Kalman filter, generalized Gaussian distribution	Extension of fast dynamic time warping (EFDTW)	69 video-recorded retinal surgery steps, 1400 video-recorded cataract surgery, and 1707 movie clips with classified human actions	Eye surgery	MPEG-4 AVC/H.264
Ramya et al. [4]	Global features	Entropy, histogram, frequency	Neural network	15478 XML scripts based on key frames	Medical videos	mp4
Schoeffmann et al. [5]	Position, colour, texture	N/A	Earth mover's distance, signature quadratic form distance, and signature matching distance	25 stored video files per procedure	Laparoscopic surgery	N/A
Beecks et al. [6]	Spatial information (colour, coarseness, contrast value)	Local feature descriptor	k-mean cluster	25 segment videos per procedure, resulting in total 1276 short video clips ranging from 60 sec to 250 sec	Endoscopic videos	N/A
Carlos et al. [7]	Local and global features, feature fusion	SIMPLE, CEDD, SURF	k-mean	1276	Endoscopy videos	N/A
Charriere et al. [8]	N/A	BoVW	k-nearest neighbour (KNN), motion histogram, and hidden Markov model	30	Cataract surgery	DV format
Muhammad et al. [9]	Colour opponent colour space (CoC), curvature map, texture	N/A	N/A	10	Hysteroscopy	N/A
Quellec et al. [10]	Texture, colour, optical flow	N/A	N/A	23 retinal, 100 cataract surgery, 69 hollywood movie	Eye surgery	MPEG 2
Droueche et al. [18]	Motion histogram	Fast dynamic time warping	k-mean, Kalman filter, and extended of fast dynamic time warping	23 epiretinal surgery video	Epiretinal membrane surgery	MPEG-4 AVC/H.264
Proposed	The key point, colour	SIFT, HOG, DT, CFD, PCA-SIFT, spatiotemporal feature vector	GMM, Euclidean distance, and Kullback-Leibler divergence	20 meniscus surgery	Meniscus surgery	Flv, Avi, mp4
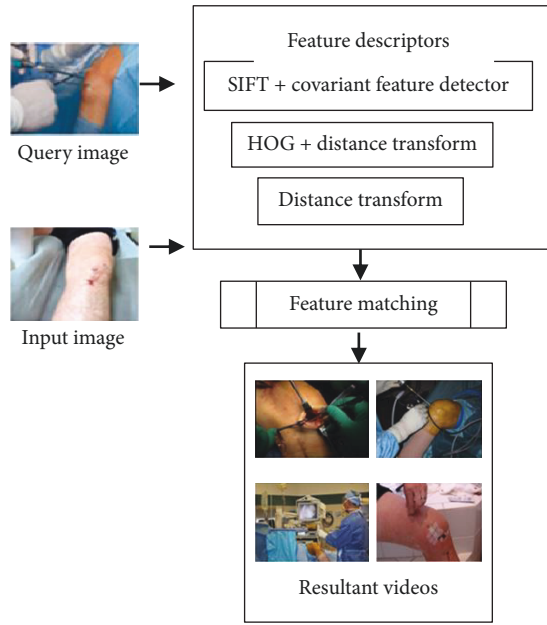
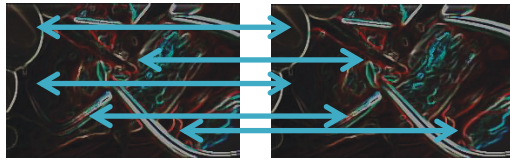FIGURE 1: Framework of key-point-based retrieval.

FIGURE 2: Temporal connection is created to form continuity between the consecutive frames.

FIGURE 3: Spatiotemporal probabilistic model for a video frame sequence $\{x1, \ldots, xn\}$.

FIGURE 4: GMM of an input query.

presumed rate, preserving temporal enslavement by their adjoining images. Spatial features of every image depend on the spatial context of every image and may depend on its former image, which is known as spatiotemporal association.

Furthermore, GMM [22] is applied to an approximate likelihood density of collected items. GMM is an unsupervised learning approach in which probabilistic model is built with the assumption that data points are produced by an assortment of a finite amount of Gaussian distributions through unidentified parameters.

Moreover, a d-dimensional spatial element for the image x_i may be demonstrated through GMM along with K quantities and a presumed number of a frame structure is represented by K, for instance, the numeral core objects instituted in particular section. By using MLE, an association can be computed through the probability of spatiotemporal continuousness amongst two successive images as presented in Figure 3.

We have acquired Gaussians against input and output videos after applying maximum likelihood as shown in Figures 4 and 5. By analyzing data distribution of these Gaussians, we can determine that dissimilarity is higher between input and output. To measure the dissimilarity ratio, k-Lealer divergence, also known as relative entropy, was used.
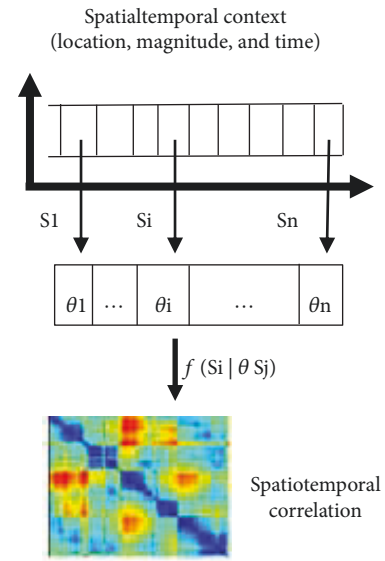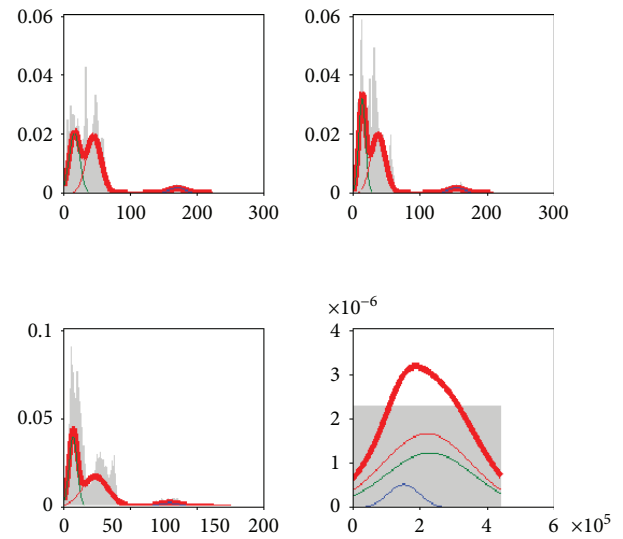
2.6. Compression-Based Video Retrieval.
In order to reduce insignificance and redundancy, datum compression technique can be used. The intention of compression is to cut down the data dimensions for effective storage, retrieval, and manipulation. For this purpose, principal components analysis (PCA) is used to the significant components from data that immaculately defines the entire image. Moreover, a hybrid approach is applied by the combination of PCA along with SOFT and GMM to improve the precision of retrieval method. Details of the method are as follows.

2.6.1. Summary of PCA.
Principal component analysis is the dimensional reduction method which calculates linear projections to maintain maximum variance in original data.
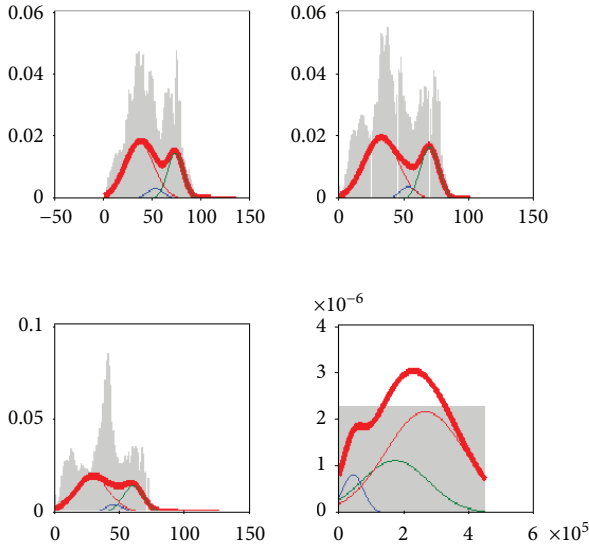
FIGURE 5: GMM of the output query.

Eigenvectors to Eigenvalues λ of a matrix of the covariance E_x of datum $D \in R^{dxM}$ where d defines the dimensions of the original data and M shows the total count of points.

$$E_x a = \lambda a. \tag{4}$$

By visualizing the point x_i via projecting

$$y_i = A x_{is}. \tag{5}$$

A defines a matrix having eigenvectors associated with 2 or 3 of the biggest eigenvalues, and Y_i is a calculated lower dimension sketch of X_i. PCA is the most commonly used method to visualize new data. Images produced by PCA are very easy to deduce and also works moderately well even when the difference in data stands primarily intense in solitary limited directions [23].

2.6.2. PCA-SIFT-Based Video Retrieval. In this method, PCA is applied on SIFT which is highly vigorous to image distortion. In PCA-SIFT firstly computes an Eigenspace to define an inline image of the native patch. In the second step, the local image gradient is calculated on a given patch. In the third step, it projects gradient image vector by the Eigenspace to originate the dense feature trajectory. That feature trajectory is expressively negligible than SIFT feature trajectory and may be used with the identical matching method. Euclidean distances amongst binary feature trajectories can be used to define either these two trajectories resemble identical interest point in dissimilar frames. The following are the steps to retrieve videos using PCA-SIFT approach.

(1) Input shot is transformed to image.

(2) The master image is extracted using shot boundary detection.

(3) For each mater frame, PCA-SIFT is calculated.

(4) Steps from 1–3 are repeated for target videos from 1–N.

(5) After step 4, two feature vectors were gotten, that is, for input and target videos $[1 - N]$.

(6) The feature vector of input and target videos is compared using Euclidean distance to compute the similarity.

2.6.3. PCA-GMM-Based Video Retrieval. The analyses of Figures 4 and 5 show that spatial context is comprised of location and magnitude features against each pixel such as RGB colour channel. These colour channels are extracted from input and target videos and passed through dimension reduction process using PCA. These compressed channels are modeled using GMM for spatiotemporal analysis. Furthermore, Kullback-Leibler divergence is applied in similar way as presented in Section 3.2.

2.6.4. Temporal Points Extraction from Surgical Videos by Sparse-Optical Flow. Surgical video is a dynamic entity in which the motion of an object plays a vital role in understanding an ongoing surgical process. To capture the movement or motion of surgical procedure, we have used optic flow method, which is an arrangement of moving object in the photographic sight prompted through a relative movement amongst a scene plus an eye, that is, camera. The Lucas-Kanade method was used to capture the motion of objects within a video of live surgery [17].

3. Results and Discussion

In this section, we present experimental settings including the data set and evaluation method.

3.1. Dataset. Datasets are comprised of 20 real meniscus surgical videos collected from various online publicly available repositories MedlinePlus [24]; AAOS [25]; vjortho (2017); and Northgate (2017). These shots are sufficient and beneficial in the current scenario because it covers a wide range of multiple scenarios and uniquely identifies a number of instances. It also provides a higher degree of relevance with reference to its surrounding and properly differentiates from the nonadjacent objects. Furthermore, it is clear from Table 1 that for testing purposes, the state-of-the-art systems have considered a small testing dataset. However, for result evaluation of a proposed system, we have requested three surgeons to choose the ground truth from our dataset. A ground truth is selected in terms of video, a relevant approach, and a mechanism can also be found in [9]. Sample frames from meniscal surgery are shown in Figures 6–8.

3.2. Tool. To precede the evaluation phase, an input video is provided to our proposed system. Its spatiotemporal features are computed against 1–N targeted surgical videos, and the most relevant videos are displayed in the right side of the interface as shown in Figure 9.

3.3. Evaluation Measurement. For the purposes of a system evaluation, there are many parameters such as precision,

Figure 6: Frame extracted from reconstruction surgery of meniscal (1).

Figure 7: Frame extracted from reconstruction surgery of meniscal (2).

Figure 8: Frame extracted from reconstruction surgery of meniscal (3).

recall, and ROC area. But in the information retrieval system, precision metric is used because it provides the proportion of retrieved videos that are actually relevant. In order to measure the precision of our system, we have split the video into small segments and these segments are further compared with the ground truth. Precision can be computed using the following metric [26, 27].

$$\text{Precision} = \frac{\text{TP}}{\text{TP}} + \text{FP}. \qquad (6)$$
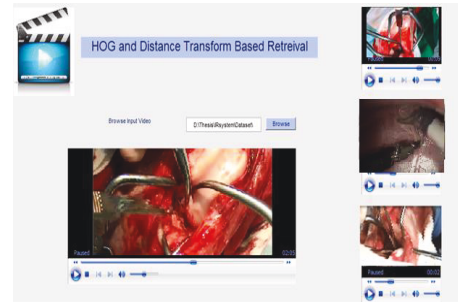

Figure 9: Sample screenshot of relevant videos using HOG and CFD.

Herein, a video is said to be the true positive (TP) if it is selected by both our system and the surgeon, or said to be false positive (FP) if it is selected by our system but not by the surgeon. Using this scheme, we measure the average precision of our system using our four proposed methods as shown in Table 2. It has already been mentioned that video retrieval system specifically for meniscus surgery was not yet found in the literature. Therefore, we have considered a number of general parameters for video retrieval system for result comparison. A brief comparison of medical video retrieval system and our proposed system is shown in Table 2.

We can deduce from Table 2 that PCA-SIFT from our proposed method performs well. Observable key factors of this dominant performance are that the incline areas nearby SIFT points are extremely organized to be presented through PCA. Eigenvalues against the patch decline are considerably faster than Edigenvalues for arbitrarily designated incline patches. For the reason that patches adjoining points entirely share definite characteristics, reducing since SIFT point recognition stage; they are totally adjusted, interchanged to align central inclines with vertical, and are scaled correctly. That streamlines the trade PCA necessity do and enabling the PCA-SIFT to exactly characterize patch by the small number of dimensions along with Lucas-Kanade optical flow method.

4. Conclusion

Our proposed video retrieval system for meniscal surgery modeling is based on four methods. These methods include key point-based retrieval, statistical retrieval, PCA-SIFT-based retrieval, and PCA-GMM-based retrieval. They further exploit spatiotemporal features for the improvement of health services to human beings.

In a key point-based retrieval method, key points of video are extracted using SIFT, HOG, CFD, and DT. Secondly, using the statistical retrieval method, spatial and temporal attributes are statistically modeled using GMM and Kullback-Leibler divergence. Moreover, by using PCA-SIFT-based retrieval and PCA-GMM-based retrieval methods, the dimension reduction technique is used with SOFT and GMM, respectively. To track spatiodevelopment over temporal interims, Lucas-Kanade sparse-optical flow method was used, which is efficient and robust to noise.

TABLE 2: Comparative analysis of proposed system with already existing systems.

	Dataset	Clip size	Average retrieval time	Average number of match frames/ videos	Recall	Average precision	ROC	Description
1	Medical videos & classified human action	Avg 20 sec	7 min 3 sec	0	0%	0.43	0	*Time* required to compute feature vector for 9 videos 0.43 in case of ERM and CD dataset
2	Medical videos	1.97 min and above	16.4 (retrieval time for all attributes)	111	0%	0.983	0	*Precision* is based on 8 videos using the neural network in terms of relevant key frame retrieval and a total number of key frames for that video
3	Laparoscopic videos		0	0	80%	0	0	N/A
4	Endoscopic videos	60 sec to 250 sec	0	0	88%	0	0	N/A
5	Endoscopy	0	0	0	0%	0.78	0	N/A
6	Cataract	0	0	0	0%		0.7	With motion analysis and 0.9 with the presence of surgical process
7	Hysteroscopy	2 to 3 min	0	0	0%	0.92	0	
8	Eye surgery	0	0	0	0%	0.72	0	In the case of CSD dataset
9	Epiretinal surgery	0	0	0	0%	0.62 0.57	0	0.69 in case of motion histogram
Proposed	Meniscal surgery	60 sec	60 sec	10	78%	0.68 0.64 0.78	0	0.78 for PCA-SIFT

For experimental work, we have used 20 videos of meniscus surgery for human health services, and our evaluation shows that results gained using PCA-SIFT are more significant. The results established using PCA-SIFT approach enhanced the obvious results with an average precision of 0.78.

Our proposed system will help out the medication of diverse patients, and surgeons can especially get guidance by watching the videos of surgeries performed by experts. Various datasets of real meniscal surgery in our system will be explored in the future for testing purposes and evaluation.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] J. B. Thorlund, K. B. Hare, and L. S. Lohmander, "Large increase in arthroscopic meniscus surgery in the middle-aged and older population in Denmark from 2000 to 2011," *Acta Orthopaedica*, vol. 85, no. 3, pp. 287–292, 2014.

[2] K. Charriere, G. Quellec, M. Lamard, G. Coatrieux, B. Cochener, and G. Cazuguel, "Automated surgical step recognition in normalized cataract surgery videos," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4647–4650, Chicago, IL, USA, 2014.

[3] P. Over, J. Fiscus, G. Sanders et al., "TRECVID 2014 – an overview of the goals, tasks, data, evaluation mechanisms and metrics," in *Proceedings of TRECVid*, p. 52, Orlando, FL, USA, 2014, NIST.

[4] S. T. Ramya, P. Rangarajan, and V. Gomathi, "XML based approach for object-oriented medical video retrieval using neural networks," *Journal of Medical Imaging and Health Informatics*, vol. 6, no. 3, pp. 794–801, 2016.

[5] K. Schoeffmann, C. Beecks, M. Lux, M. S. Uysal, and T. Seidl, "Content-based retrieval in videos from laparoscopic surgery," in *Medical Imaging 2016: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 9786, pp. 1–10, San Diego, CA, USA, 2016.

[6] C. Beecks, K. Schoeffmann, M. Lux, M. S. Uysal, and T. Seidl, "Endoscopic video retrieval: a signature-based approach for linking endoscopic images with video segments," in *2015 IEEE International Symposium on Multimedia (ISM)*, pp. 33–38, Miami, FL, USA, 2015.

[7] J. R. Carlos, M. Lux, X. Giro-i-Nieto, P. Munoz, and N. Anagnostopoulos, "Visual information retrieval in endoscopic video archives," in *2015 13th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pp. 1–6, Prague, Czech Republic, 2015, IEEE.

[8] K. Charriere, G. Quelled, M. Lamard et al., "Real-time multi-level sequencing of cataract surgery videos," in *2016 14th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pp. 1–6, Bucharest, Romania, 2016, IEEE.

[9] K. Muhammad, M. Sajjad, M. Y. Lee, and S. W. Baik, "Efficient visual attention driven framework for key frames extraction from hysteroscopy videos," *Biomedical Signal Processing and Control*, vol. 33, pp. 161–168, 2017.

[10] G. Quellec, M. Lamard, Z. Droueche, B. Cochener, C. Roux, and G. Cazuguel, "A polynomial model of surgical gestures for real-time retrieval of surgery videos," in *Medical Content-Based Retrieval for Clinical Decision Support*, vol. 7723 of Lecture Notes in Computer Science, , pp. 10–20, Springer, 2013.

[11] C. A. Kelsey, P. H. Heintz, D. J. Sandoval, G. D. Chambers, N. L. Adolphi, and K. S. Paffett, "Radiation interactions with tissue," in *Radiation Biology of Medical Imaging*, pp. 61–79, John Wiley & Sons, Inc., 2014.

[12] D. Markonis, M. Holzer, F. Baroz et al., "User-oriented evaluation of a medical image retrieval system for radiologists," *International Journal of Medical Informatics*, vol. 84, no. 10, pp. 774–783, 2015.

[13] S. W. Zhang, Y. J. Shang, and L. Wang, "Plant disease recognition based on plant leaf image," *The Journal of Animal & Plant Sciences*, vol. 25, no. 3, Supplement 1, pp. 42–45, 2015.

[14] J. Choi, Z. Wang, S. C. Lee, and W. J. Jeon, "A spatio-temporal pyramid matching for video retrieval," *Computer Vision and Image Understanding*, vol. 117, no. 6, pp. 660–669, 2013.

[15] H. Bai, Y. Dong, S. Cen et al., *Orange Labs Beijing(FTRDBJ) at TRECVid 2013: Instant Search*, International Journal of Multimedia Information Retrieval, 2013.

[16] S. Bharathi, D. Shenoy, R. Venugopal, and M. Patnaik, "Ensemble PHOG and SIFT features extraction techniques to classify high resolution satellite images," *Data Mining and Knowledge Engineering*, vol. 6, no. 5, pp. 199–206, 2014.

[17] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/Kanade meets Horn/Schunck: combining local and global optic flow methods," *International Journal of Computer Vision*, vol. 61, no. 3, pp. 1–21, 2005.

[18] Z. Droueche, M. Lamard, G. Cazuguel, G. Quellec, C. Roux, and B. Cochener, "Motion-based video retrieval with application to computer-assisted retinal surgery," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4962–4965, San Diego, CA, USA, 2012.

[19] Z. Droueche, G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, and C. Roux, "Computer-aided retinal surgery using data from the video compressed stream," *International Journal of Image and Video Processing: Theory and Application*, vol. 1, no. 1, 2014.

[20] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[21] B. V. Patel and B. B. Meshram, "Content based video retrieval systems," *International Journal of UbiComp*, vol. 3, no. 2, pp. 13–30, 2012.

[22] B. Werner, S. Harald, and M. Roland, *Semantic Indexing and Instance Search; Joanne Research at TRECVid*, TRECVID Workshop, 2013.

[23] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: a survey," *Foundations and Trends® in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2008.

[24] MedlinePlusU.S. National Library of Medicine, Bethesda, MD, USA, April 2017 https://medlineplus.gov/surgeryvideos.html.

[25] American Academy of Orthopaedic SurgeonsApril 2017, http://www.aaos.org/videogallery/.

[26] J. Davis and M. Goodrich, "The relationship between precision-recall and ROC curves," in *Proceedings of the 23rd international conference on Machine learning - ICML '06*, pp. 233–240, Pittsburgh, PA, USA, 2006.

[27] M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," 2014, March 2016, https://medtube.net/orthopedics.