A Smart Methodology to Supervised The Human Activities

Haneen Hassan Al-Ahmadi

Software Engineering Department College of Computer Science and Engineering University of Jeddah.
Jeddah. Saudi Arabia

Abstract

Economics, social norms, lack of confidence can be a hurdle for people to learn activities that require some degree of perfection to execute, e.g., praying in case of Muslims, dancing, fitness exercises. Religious, cultural, and social activities are strictly practiced in some communities. The recent technological advancements are playing an essential role as teaching and training as known as computer-aided learning tools that help users learning using oral pronunciation, image, and video recognition. Most of the systems provide only a passive learning experience, and no feedback makes them less efficient learning tools. In this paper, the authors propose a system that uses techniques from pattern matching and image processing for content matching integrated with knowledge-based systems to handle the variations in the activities. The primary purpose of this paper to tell the user about those rules that are necessary to follow in that place where the complex sequence of action could be performed. If these rules are not following, then all activities could be performed again, and it could be considered as wrong. For the testing of the system, also for checking the validation off the system, authors take the data information about the prayer movements for Muslims, yoga, and classical dance.

Key words:

image processing, video content similarity, sift features, knowledgebase system, and pattern matching

1. Introduction

Society, culture, and religion can sometimes make learning some activities that are carried out physically as gestures or movements like dance, exercise, prayer mandatory. Some are of immense importance and carried out publicly, like praying in a Mosque or dancing at a social event [1]. The most obvious solution to the problem is to learn from an expert. Several factors reduce the viability of this solution for most people, like the availability of experts, cost, of course, fear of lack of confidence, investment of time, and money [2]. Most of the above-stated problems can be addressed using a Computer-aided learning system. Computers are a relatively common commodity and are increasingly becoming an integral component in daily y lives with features like entertainment, processing documents, and communication media [3].

To a certain degree, we will find lots of systems that may assist in the inspection procedure. Base around the arrangement of this video, a productive online video of similarity dimension might be executed. Mostly to figure out the approximate similarity that they genuinely are changed into the smaller chunks of eyeglasses, which precisely the like a pre-defined arbitrary image cancel. On-line mastering around the massive scale about the grounds of picture similarity is quite helpful; for example, finding a graphic that has resemblance into any picture or author needs to discover an online video very similar for our memory videogame [4]. Extracts voice attributes in Quran's poetry recitation for its Quran subscribers and maps with the info accumulated throughout the practice period. The system stated precisely the mistake in the event the info of voice is not matched. If their clients do such tasks according to a string, those assistive approaches tend not to support. Concerning their moves and also a chain of tasks, this particular system does not ship any info. A number of those spiritual or cultural activities take a stringent arrangement and method of implementing like begging for Muslims 5 days every time, classical dance, operation, and Yoga. The number of issues of tasks can be different from straightforward to hard, which makes an error can lead to shame, financial reduction, or replicating the actions [5].

For Example, Sequence dancing is an exceptionally well-known type of social dancing in the United Kingdom and abroad. Dissimilar to the advanced move styles where a learner can feel awkward on account of his/her constrained information of a substantial number of steps and varieties that can be performed. It depends on the ordinary Ballroom and Latin dancing; however, everybody moves the same institutionalized 16 bar grouping in the meantime and is named particular terms from the expert that made it. If any grouping of movement is missed, then the students need to play out the entire action again, or the creator can state to make hone still he/she gets expertise in that action. Muslims have a well-defined methodology of praying specifying movements and prayers, variations to the standard are not allowed (Unless there is a medical condition). These steps vary in number depending on the time and type of prayer and must be executed without mistakes [6]. Just like dancers need practice, people perform their yoga and another physical exercise for health and fitness. Proper execution in case of physical training like lifting weight and yoga is essential as wrong positions or movements may cause pain, strained or damaged muscles or bones. The use of media like training videos for learning different activities like dancing, or audiotapes for learning languages or sentences is not a new idea and has been used for a long time. The

learning is passive, and there is no external feedback on the progress of learning. So, it can conclude that the above-stated activities require a 3rd person perspective and evaluation to establish the validity of the action. The quality of learning is limited to the perception and understanding of the learner, having no prior experience of the task he/she is learning. It is safe to assume that an expert system can not only provide a learning opportunity with much-needed feedback for learning [7].

As a way to eliminate the above-stated dilemmas dilemma, the writer suggests that a video articles fitting technique (VCM). VCM works by using pattern fitting and picture processing methods to coordinate with the activities of men in both two flows. The strategy employs the database to get video lessons from pros improved for optimum fitting. Authors utilize picture processing approaches for noise elimination and regular fitting to find gaps inside the finished tasks of their consumer's video in comparison to this skilled. An understanding base process can be utilized to save exactly the specifics of tasks involving forms, repetitions, along with timings of activities to be achieved. Mcdougal defines a similarity dimension between video clip graphics while in the circumstance of recovery. We would aim to supply a platform that will aid its clients to master and also implement a succession of activities to finish a task (dancing moves, yoga, and prayer) even though reducing period limitations, the price tag of education, and also additional telltale variables discussed above.

2. Literature Review

Computer-aided learning has revolutionized the learning experience and the reach of students. E-learning and distance learning is now the norm and practical methods of learning. There have been many learning tools developed with different target areas; the author has discussed some of the systems in this section. Voice systems to learn other languages with the help of electronic disks are also presented, but they are passive and do not interact with the learner. Several tools for learning the phonetic delivery of words and languages have been developed with feedback for the correctness of the uttered words [8]. The system uses speech signal processing and feature extraction and content matching. Voice content matching system for Quran readers extracts voice features from the Quran's verse recitation and maps them with the data collected during the training phase. Every mismatch mistake is pointed out. A similar system was developed to cope with phonetic diversity and learning languages. The system used a multi-classifier system to match the voice content, with the ability to find a phonetic mismatch and missing words. The above-discussed feedback learning system reinforces our argument for the third person (or Computer-assisted) evaluation can shorten the learning curve and help achieve the required skill level [9].

In this paper, writers discuss several calculations that their applications, benefits, restrictions, and also for that which application that is employed by most people to our graphics fitting system utilizing different approaches and distinct graphics profiling calculations. The concept dependent around the summarization of each series of online video investigation introduces and also productive video resemblance dimension. The video is changed over into a little section of edges that are generally like predefined irregular pictures offsets using Visit strategy for Video similitude location with video signature Clustering and figure inexact likeness. The aim of the system is the design of video sequences can be matched, and to handle excellent contents in the database. They use a digital signature of the video, which is a small number of frames from the original [10]. The system is practical for an extensive database. That is a necessary trade-off because frame by frame matching is not a practical matching methodology for more extensive databases. However, this approach cannot filter out smaller details that are essential for our problem. If the author ignores any number of frames in the test video may result in an incorrect result. Active recovery of shape classes from pictures and video utilizing neighborhood self-similarities can be utilized for applied introduction of shape to perceive recognizable examples in pictures and recordings that are often missed by standard framed essential techniques. The procedure performs well for a picture with various scaling; however, the method is not tried with the more significant corpus [11].

Largescale on-line finding out of film recognition through placement frame is tremendously precious in large scale, as by way of instance, locating an image which has resemblance into the provided film or founder must detect an online video such as our stock-piling online video. The online calculation for versatile picture likeness (OASIS) framework is the online calculation for adaptable picture comparability that learns closeness estimation by the utilization of the bilinear framework. In light of a dataset on a vast scale, the OASIS proposes that query similarity can be precisely learned [12]. It is precious for individuals who need to learn and find new information. Likeness estimating calculations depend on the prepared information that contained real-valued distances. All the above-stated system cannot handle well the matching process when the videos are out of synchronization. In our problem, it can be widespread as the user may start too early or too late to act, deeming the use of above-stated algorithms less viable for use. People also use online dictionaries and websites to learn foreign languages. However, all these applications teach their users in one-way like mode. Users can view activities in a video of Yoga or classical dance, or they can hear the exact pronunciation of any word/sentence uttered by a native speaker. However, if the learner wants to check

his/her accuracy/level, he/she has no way to do it other than he/she consults a human expert, which involves some limitations like availability, expertise level, and economy [13].

Psycho-visual display top-quality foundation on multi-scale likeness arrangement demonstrates the chance of metric predicated display caliber on detecting comparability in the arrangement border. That will be done by utilizing the degree of the uprooting of a specific location, making use of the wavelet switch module amongst our benchmark, and also store out advice. Factual proximity seems related to content-based online video copy identification calculation is dependent upon the discovery of substance-based replicate. That is but one of those climbing landscapes and audio software, which is a requirement of an unflinching project for our database place and also the computer system vision modern society. In all the above-discussed systems, the use of the application is of one-way like fashion. Users of these applications are not sure if they are learning accurately or not. For this purpose, in this paper, the author proposes a VCM system. The video matching contains certain activities that are performed in multiple sequences, and the performance of every sequence is necessary for sight person to complete the activity. In the video content similarity perform certain activities like (classical dance, yoga, and 5- time prayers), there must be a system that can guide those people who have less time to attend dance class, may hesitate consulting dance master due to the economics, or another factor like attach to their social, and religious norms. The task is achieved by our video content matching system for the supervision of human activities. The ability of the user to use the system offline and in isolation is another merit of the system, making it a more practical and user-friendly solution [14].

3. Architecture and Working of VCM system

Our proposed system supervises human activities (guide untrained people in performing certain activities). VCM provides a user with a variety of activities to learn. Users are given a preview of the activity and are requested to mimic the actions in the video. Our system then uses the camera to capture the motion and gestures of the user and compares it to the database of action by using feature matching of the SIFT technique. The results are communicated to a user after Analysis using the Knowledgebase system. The user interface displays both videos (from the user and expert side by side for further details).

Video and live video support are two parts of stored data that our system consists of. All the detail in the video that is stored in the system is shown in Fig. 1 that is used to find out the similarity and mistakes according to live video, these activities guide the user to provide a sequence of

activities and to interact with the system. With the help of the image acquisition toolbox, these videos can be created on MATLAB. Our system can be divided into three modules 1) User interface 2) Content matching Or SIFT Agent, 3) Knowledge Base System, in this section, the author discusses the implementation details and working of the system. The typical structure of our work depends on three levels application level, document support level, and information level.

A. User Interface

The user interface (UI) module is responsible for the interaction of the user with the system and also serves the Video Acquisition tasks. Our system uses stored videos of different activities for the uses that may be interested in learning Fig.1. The user can select the activity they want to learn. Users can also preview the activity before trying to perform. When the user is ready to perform, the next task is video acquisition. For Video acquisition author use RGB (Red Green Blue color format) with the resolution of 1024x 768 using the image acquisition toolbox of MATLAB. Higher resolutions are possible, but the selected configuration works well for our implementation. Video acquisition time is dependent on the type and length of activity. The author used a webcam for hardware support.
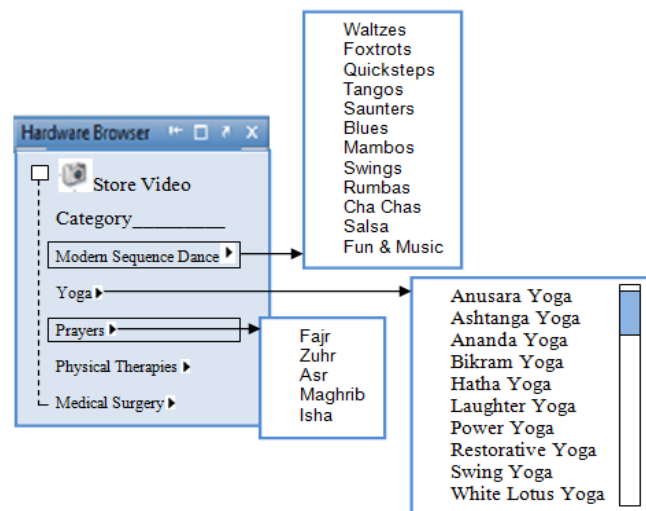

Fig. 1 By using a Hardware Browser Store Video procedure

System inner working is explained with the help of Fig.2, which provides a broad vision about the role of similarity system.
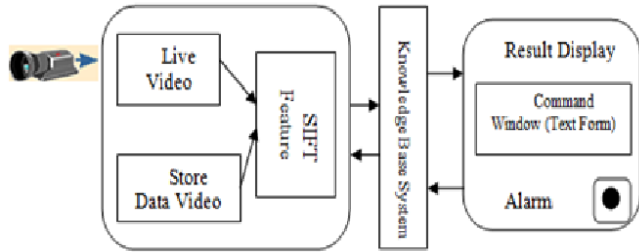
Fig. 2 The architecture of Video Content Similarity System[15]

For the insurance accuracy in the system, users are encouraged to maintain a distance so that the user is in the same frame size as the video of the expert. At that place where the sequence of the movements is strictly observed to get suitable results author takes the prayers of Muslims, yoga, and classical dance for the validation of our system [15]. It can be taken as an example that, as a ballroom dancer, the sequence dancing is a challenging activity. The steps are repeated again and again until the end of the routine, and also all the steps are set in a 6 to 32 bar routine. Table1 shows the Rumbas dance steps sequenc1e that was used as a test case for learning.

Muslims can use it for learning prayer movements that are to be synchronized with prayers. The sequence of movement and proper execution and placement of hands, elbows, and knees are significant as wrong execution deems prayer incomplete. At the end of RAKA'AH (one round), prayer follows ten steps and repeats it again and again, as shown in Table.2. Yoga is good for health, and fitness users can use this, but they have limited knowledge about the steps of yoga that can define to the user that how yoga can be performed. There are three steps in yoga that should follow standing, balancing, twisted, and seated, as shown in Table.2. There could be several different steps that should be performed in yoga, and that all depends on the type of yoga [16].

B. SIFT Featured Agent

For detection and extraction of a local feature from the desired image, the scale-invariant feature transform used. These features are working as a local engineering descriptor. To find the similarity and interest features author should have to transform the content of the image into a local feature. To do so, the author separates the videos into frames. SIFT feature extraction and matching are then applied to the images. The matching is based on not only the presence but also the position of the desired object. The potential problem of losing synchronization of images as users act slowly or faster than the expert's action is solved using brute force matching tactics. The computational overhead is acceptable as the size of the videos is small. For finding the similarity, the author could use the SFIT algorithm in MATLAB. That is necessary that images

should be in sequence when the author is trying to find out the similarity. There should be two conditions should be the following:

- Minimum background variation to remove noise in matching.
- Save frame/images in the sequence of actions.

How the author can get the frame/images data from the stored video and live video is shown in Fig.3, 4. SIFT algorithm applied to it for similarity check after getting the frame/ images.

Table1 1: Rumba's Dance Step Sequence

Sr. #	Routine	Step Sequence	Phase of movements
1	Bar1	Left Feet Forward	Slow
		Right Feet to Side	Quick
		Left Feet Closes to Right Feet	Quick
2	Bar2	Right Feet Back	Slow
		Left Feet to Side	Quick
		Right Feet Closes to Left Feet	Quick
3	Bar3&4	Repeat bar 1 & 2	SQQSQQ
4	Bar5	Left Feet to Side	Slow
		Right Feet placed behind Left Feet Toe - Heel	Quick
		Replace weight to Left Feet	Quick
5	Bar6	Right Feet to Side	Slow
		Left Feet placed behind Right Feet Toe - Heel	Quick
		Replace weight to Right Feet	Quick
6	Bar7&8	Repeat bars 5 & 6	SQQSQQ

Table 2: Sequence of Activities

For Prayer		
Start	One RAKA'AH (one Round)	Finish

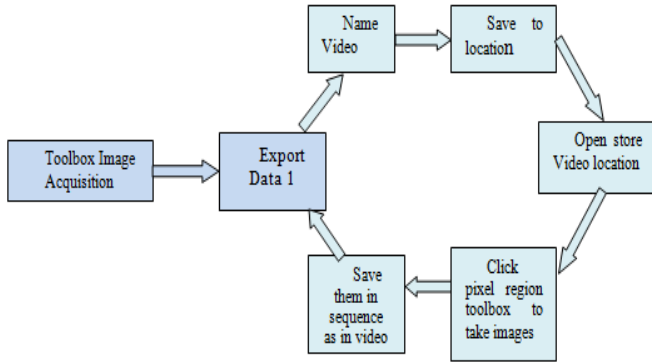| For Yoga		
Sr. #	Step Sequence	
1	Seated & Twisted	
2	Standing	
3	Balancing	Total Body Yoga Flow

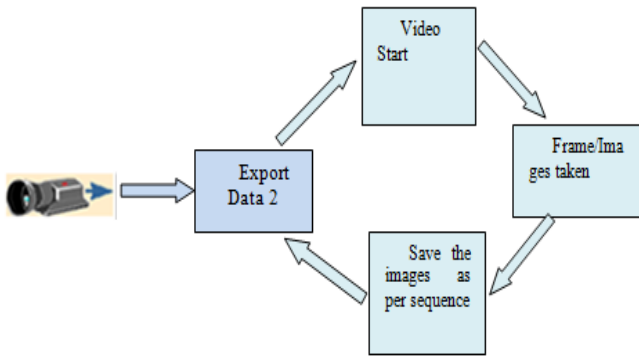Fig. 3 Data were taken from Store Video

Fig.4. Data were taken from Live Video

An abstract description of the SIFT algorithm is discussed below; the first features are selected, scaled, and stored. That is the training phase of the system.

- Scale-space local Extrema Detection: The features selection and location are based on the Difference of Gaussians as known as Dog pyramids. That is done by repetitive convolution of Gaussian kernel α= 1.6. The images are downsampled with increased blur by a factor of 2. The collection of images is called octaves. Gaussian Blur operator is used within the octave as we know that each image size is half of the previous image, so this blur operator is commonly used for the blurred images. The Local extrema are calculated.
- Keypoint Localization: The second-order Taylor function is used to calculate the quadratic function with origin from the same point. Local Extrema that of edges and lower contrast values are discarded because they can cause interference due to noise.
- Orientation Assignment: Local image gradients are used to determine the SIFT feature location. The Gaussian window could be used for weighted the octaves features [17].
- Keypoint Descriptor: Even the gradient advice has

been made to buy into the debut of the keypoint and from then on by a Gaussian using a spoonful of 1.5 key-point scales. This info is subsequently utilized to organize of histograms within a window on the essential purpose. Heavily weighted descriptors make use of an arrangement of 16 histograms, corrected in a 4x4 lattice, every having 8 debut canisters, just one for each among many essential Interface keywords and also we for each of these midpoints of those claws all these results within a factor vector are comprising 128 parts. The made descriptors are subsequently utilized for the magical fit of this thing of attention. Even the above-stated pursuits are performed from the SIFT element representative. SIFT element representative sends the fitting questions into this knowledgebase platform in order that it contrasts the set of this desirable classification, as shown in Fig.5.

C. Command screen and alarm/Beep system

All the mistakes categories detail can be shown on fault management screen that can share the common mistakes and display zone that can also alert the users from the whole mistakes. Fault Display Screen: For displaying the full detail of a single mistake and also the current status based on the knowledge base, this screen should display the status of whether the user made a mistake or not. Alarm/Beep indicator: When the system has detected a fault warning condition this indicator only becomes lit and visible
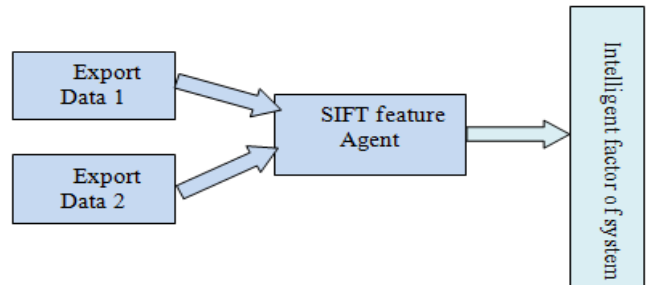
Fig.5. SIFT agent architecture

```
STARTRULE "Classical Dance Category"
IF          User (train_1) =A
Request _Case (A) = "Classical Dance"
            Body Angle Movement<=60°/ as per required sequence
            Movement of Position<= foot/Arms/
            Time Space<=20sec
            Check Similarity
THEN
 Store the category as classical dance with 60° angle foot movement in
 Every 20 second.
```

D. Knowledge-Based System

According to the sequence of activity, time-space, movement position, and angle movement, the SIFT feature access the knowledge base that stores all the activity. The factor of intelligence is based on the Rule base (RB) and Knowledgebase (KB). For the storage of the sequence of the action step sequence and also for the retrieval of possessions, the knowledge base can be used. To store of activity into the knowledge base is defined with some rules that are defined the list of the desired category could be comparing by sending a query from the SIFT agent to the knowledge base.

4. Case study

In this case, the study to find out the similarity and mistakes and the sequence of action is essential. The author tells about the VCM's guiding human activities. The user should have to select the category like fitness exercises, Yoga, physical therapy, prayer, and classical dance. We can show with an example that there are several steps in classical dance that are very difficult for the user to memorize, and also, there is more variation in many kinds of steps in the classical dance. Let us consider Lucy is interested to learn classical dance without the help of dance master due to low economics or any other factor like she has not enough time to attend dance class, and she also wants the system to tell our mistakes for correction. She selects the category of class dance. Data of the classical dance stored in the memory as stored in the video data. Lucy is the user who wants to make her activity is just like an expert user. She should have to select the camera toolbox to make a live video. When Lucy starts classical dance practice, the camera captures the video and system match the performance of the person to its stored data by using the SIFT feature and send to the KB system that is populated based on the sequence of activity, the angle of movements, movement phase, and time-space. After checking similarity, the system tells about activity perform well or not as Figure6. If no difference exists, then the system shows result in a command window in the form of text. An example of Rule is as follows:

```
START RULE "Classical Dance Training"
        IF        User (train_x) = trainer /* 'x' is Lucy */
                  Request _category (x) = "Dance"/*Lucy
choose classical dance*/
                  Check Body Angle Movement
                  Check Movement of Position
                  Check Phase of Movement
                  Check Similarity
        THEN
                  Display<=Check_SimilarityWith "Classical
Dance"
                  Train<=Display (Similarity match: /*Activity
perform well*/
```


Fig. 6 Results of Some movements of Classical Dance

If some difference exists, the system declared as wrong in the form of Alarm/ Beep signal and text form as Fig.7 so that she makes sure that mistake is done and perform activity again or the whole exercise is to be repeated. An example of Rule is as follows

```
START RULE "Classical Dance Training"
        IF        User (train_x) = trainer /* 'x' is Lucy */
                  Request _category (x) = "Classical
Dance"/*Lucy choose classical            dance*/
                  Check Body Angle Movement
                  Check Movement of Position
                  Check Phase of Movement
                  Check Similarity
        THEN
Display<=Check_Similarity with "Classical Dance"
Train<=Display (Similarity not match: /* Arm angle Movement
difference_Declare wrong with Alarm and Declared whole Exercise
repeated */)
```


Fig. 7 Results of Some movements of Classical Dance

In the second usage instance author view that Muslims offer prayers per single day five-time and of the Muslims beg without errors. Inside this usage situation, the author

employed for the recovery of online video graphics with is living activity played by human beings additionally revolves around the recovery of video clip graphics in just a massive video series predicated on saved info. Amir can be a teenaged youngster that starts off to state their prayers with no blunders. The entire method of 5 Truth is saved from memory. After Amir maintains his "FAJIR," prayer starts the digicam, create the online video match the functionality of this man to the stored info should a gap is different that he tends to make specific mistakes is performed afterward Amir needs to state his prayer as shown in Fig.8.

Fig.8. Results of the Muslim Community

5. Conclusion and Future Work

Our system has proved a useful tool for learning sequenced activities using the SIFT feature. Our system is tolerant of scaling and other minor variations in the scene. VCM provides a chance to learn in solitude for a variety of activities. Although many contents matching systems are available, they do not perform well with video content, where the scene is varied and out of synchronization. Our system matches test video sequences with all the videos in the dataset to evaluate the user's actions. This is computationally expensive; this can be ignored as the length of the sequences the author used was short.

Later on, we are able to fit it into the saved info video rather than shooting frames/images in your video, and we are able to fit online video contents without even shooting video clip frames/images.

References

[1] Popoola, O.P., K.J.I.T.o.S. Wang, Man,, and P.C. Cybernetics, Video-based abnormal human behavior recognition—A review. 2012. 42(6): p. 865-878.

[2] Mannini, A. and A.M.J.S. Sabatini, Machine learning methods for classifying human physical activity from on-body accelerometers. 2010. 10(2): p. 1154-1175.

[3] Dimitrova, N. and R. Jasinschi, System for parental control in video programs based on multimedia content information. 2015, Google Patents.

[4] Bronstein, A., M. Bronstein, and S.S. Rakib, Methods and systems for representation and matching of video content. 2013, Google Patents.

[5] Dudkiewicz, G.G., D.K. Hitt, and J.P. Barker, System and method for determining the desirability of video programming events using keyword matching. 2005, Google Patents.

[6] Petkovic, M., Content-based video retrieval. 2000.

[7] Geetha, P. and V. Narayanan, A survey of content-based video retrieval. 2008.

[8] Renner, G. and A.J.C.-A.D. Ekárt, Genetic algorithms in computer aided design. 2003. 35(8): p. 709-726.

[9] Hillyard, R. and I.J.C.-A.D. Braid, Analysis of dimensions and tolerances in computer-aided mechanical design. 1978. 10(3): p. 161-166.

[10] Chua, L. and P. Lin, Computer-Aided Analysis of Electronic Circuits: Algorithms and Computational Techniques (Prentice-Hall series in electrical computer engineering). 1975.

[11] Cohen, E., et al., Discrete B-splines and subdivision techniques in computer-aided geometric design and computer graphics. 1980. 14(2): p. 87-111.

[12] Yusuf, S., et al., Effects of fondaparinux on mortality and reinfarction in patients with acute ST-segment elevation myocardial infarction: the OASIS-6 randomized trial. 2006. 295(13): p. 1519-1530.

[13] Malmberg, K., et al., Impact of diabetes on long-term prognosis in patients with unstable angina and non–Q-wave myocardial infarction: results of the OASIS (Organization to Assess Strategies for Ischemic Syndromes) Registry. 2000. 102(9): p. 1014-1019.

[14] Marcus, D.S., et al., Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. 2007. 19(9): p. 1498-1507.

[15] Flickner, M., et al., Query by image and video content: The QBIC system. 1995. 28(9): p. 23-32.

[16] Ismail, J., N.L.M. Noor, and W.A.R.W. Mohd. Smart prayer mat: a textile-based pressure sensor to assist elderly with cognitive impairment in praying Activity. in Proc. of the 5th Int. Conf. on Computing & Informatics. 2015.

[17] Lowe, D.G.J.I.j.o.c.v., Distinctive image features from scale-invariant keypoints. 2004. 60(2): p. 91-110.