Fuzzy Metrical System for Compressed Telesurveillance Databases Indexing

Samia. F. Khelifi*, M. Elarbi Boudihir and Rachid Nourine*****

*Computer Science Dept., King Faisal University, Dammam, KSA
** Computer Science Dept., M. Ibn Saoud University, Riyadh, KSA
*** Intelligent Control and Electrical Power Systems Laboratory
Research Centre, Algeria

Abstract:

This paper proposes a video retrieval system from compressed outdoor video surveillance databases. The aim is to extract moving objects from frames provided by MPEG video stream in order to classify them into predefined categories according to image-based properties, and then robustly index them. The principal idea is to combine between useful properties of metrical classification and the notion of temporal consistency. Fuzzy geometry classification is used in order to provide an efficient method to classify motion regions into three generic categories: pedestrian, vehicle and no identified object. The temporal consistency provides a robust classification to changes of objects appearance and occlusion of object motion. The classified motion regions are used as templates for metrical training algorithms and as keys for tree indexing technique.

Key words: Video database retrieval and indexing, Compressed video, Temporal consistency, Fuzzy geometry classification, Fuzzy inference system.

Introduction:

The large volume of images and videos pose a significant challenge for storage, retrieval and indexing the visual information from multimedia databases. Two approaches have been commonly used: a content indexing approach, where the index terms serve to encode the content of images; and a structural approach where images are represented as a hierarchy of regions, objects, and portions of objects. The content indexing approach is based on features such as colour, texture, shape and sketch extracted from an image, which essentially serve as the index. The structural approach is based on spatial relationships between objects or regions in a scene. In video indexing techniques using temporal features as keys, image sequences are indexed based on the motion properties of objects within the sequence. Temporal features allow the user to specify queries that involve the exact positions and trajectories of the objects in the shot. The survey of what has been achieved on the content-based image retrieval in the past few years and what are the potential research directions can be found in [Bru 99] [Hab 99] [Fer 98] [Gud 95] [Idr 97] [Smo 94] [Sch 00] [Tiz 97].

Many content-based image search systems have been developed for various applications in order to extract intrinsic image features suitable of automatic indexing and retrieval. These features are used to reduce the complexity of image comparisons and to improve the organisation of image database. Unfortunately, automatic retrieval of suitable features is very hard; it is usually only feasible for retrieval systems that incorporate a high degree of domain-specific knowledge about the type of image contents to be retrieved. In unconstrained images, the set of known object classes is not available. Also, use of the image search systems varies greatly. The knowledge of the image content can be used to index specific images in the database for purposes of rapid retrieval [Ike 01] [Nib 93] [Sch 00]. In this context, we developed a system for retrieval and indexing telesurveillance MPEG videos in relation to the dynamic content of image sequences. It includes a robust fuzzy inference system to classify motion regions into pedestrians, vehicles and no-identified objects.

System Overview:

The system consists of five stages (figure 1). In the first stage (section 3), the digital video is segmented into elementary shots. In the second stage (section 4), all the moving objects are detected and segmented into motion regions. In the third stage (section 5), the principal idea is to exploit on one hand, the useful properties classification of fuzzy metrical classification in order to distinguish between types of motion regions, and on the other hand, the notion of temporal consistency in order to provide a robust classification against changes of objects appearance, occlusion, and cessation of object motion. In the fourth stage (section 6), once a motion region has been classified, it can be used as training template for the indexing and retrieval process.

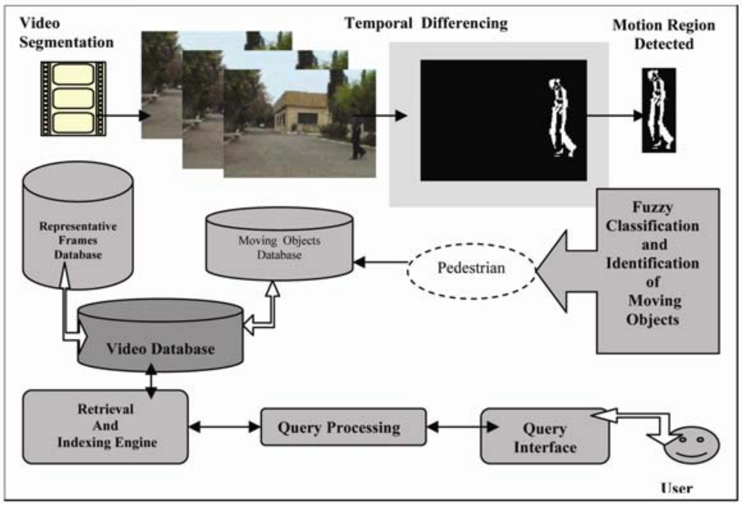

Fig. (1) : The video retrieval system overview

Video Segmentation:

The input data of the system consists of image sequences taken from outdoor video surveillance scenes. Video has both spatial and temporal dimensions and hence a good video index should capture the spatio temporal contents of the scene. In order to achieve this, the first step in video indexing is to decompose a video sequence into shots. Video shots may be associated with key or representative frames that best represent the shot. Several shot detection algorithms on compressed and uncompressed video are presented in [Yeo 95, 96] [Shn 96] [Bru 99] [Idr 97].

We propose to use a unified approach for scene change detection in motion JPEG and MPEG. This algorithm is based on the use of only DC coefficients. First we have to construct DC frame f_m^{DC} for every frame in the sequence. The DC coefficients in JPEG and I-frames in MPEG are obtained directly from each block. The DC coefficients for B- and P-frames are also estimated. The sum of the difference magnitude of the DC frames f_m^{DC} and f_n^{DC} is used as a measure of similarity between two frames.

$$D\left(f_m^{DC}, f_n^{DC}\right) = \sum_{i=1}^{X/8} \sum_{j=1}^{Y/8} \left| P(f_m^{DC}, I, i, j) - P(f_n^{DC}, I, i, j) \right| \tag{1}$$

Where $P(f_m^{DC}, I, i, j)$ is the DC coefficient of block (i, j). A scene change from f_m to f_n is declared if: (i) $D(f_m^{DC}, f_n^{DC})$ is the maximum within a symmetric sliding window and (ii) $D(f_m^{DC}, f_n^{DC})$ is 2-3 times the second largest maximum in the window.

Video shots may be associated with a key frame that best represents the shot and can later be used for the retrieval process. Let a shot represented by its first frame. Subsequent frames are then compared to the first frame, looking for a frame whose difference is above a given threshold T_s. If such a frame is found, it is considered as a key if it is followed by a continuous sequence of frames differing by at least T_s from the previous key frame. Choosing those frames of a video shot as key frames is based on the observation that consecutive frames are often almost identical. In addition, the shot is usually characterized by the first few frames, before the camera begins to zoom or close-up. So in our application it is a sufficient choice.

Motion Region Detection :

Then, all the moving objects must be accurately isolated from the background in order to be classified. Two methods are possible: temporal differencing (TD) and template correlation matching [Bre 97] [Bru 99] [Hua 83] [Kru 98] [Lip 98]. Both approaches have advantages and drawbacks. TD is impossible if there is a significant camera motion. It also fails if the target becomes occluded. On the other hand, the template correlation matching is not robust to changes in object size, orientation or even changing in light conditions. His use is most appropriate when the target size is small. So, the properties of these two methods are complementary. This is the motivation for combining TD and the notion of temporal consistency. The idea is to use TD to detect moving regions and apply temporal consistency algorithm to reduce misclassified motion regions.

Firstly each I frame of a shot is smoothed with the second derivative in time of the temporal Gaussian function. If fn is the intensity of the n^{th} I frame of the shot, then the absolute difference function Δ_n is:

$$\Delta_n = | f_n - f_{n-1} | \qquad (2)$$

The result of the difference is binarized in order to separate changed pixels from others. To do this, a threshold function is used and a motion image M_n can be extracted.

$$M_n(u, v) = \begin{cases} f_n(u,v) & \text{if } \Delta_n(u,v) \geq T \\ 0 & \text{if } \Delta_n(u,v) < T \end{cases} \qquad (3)$$

Where T is an appropriate threshold chosen after a several tests according to the exterior environment with different acquisition conditions [Khe 03].

To separate the regions of interest from the rest of image, binary statistical morphological operators (erosion and dilatation) are used. This allows decreasing the number of connected components. Then, the moving sections must be grouped into motion regions $R_n(i)$. This is done using a connected component criterion (figure 2). It allows to group different motion sections susceptible to be a part of the same region, or allows grouping the residual motion parts into one motion region. This propriety is useful to identify pedestrian who are not rigid and also useful in occultation of the moving object and other target.

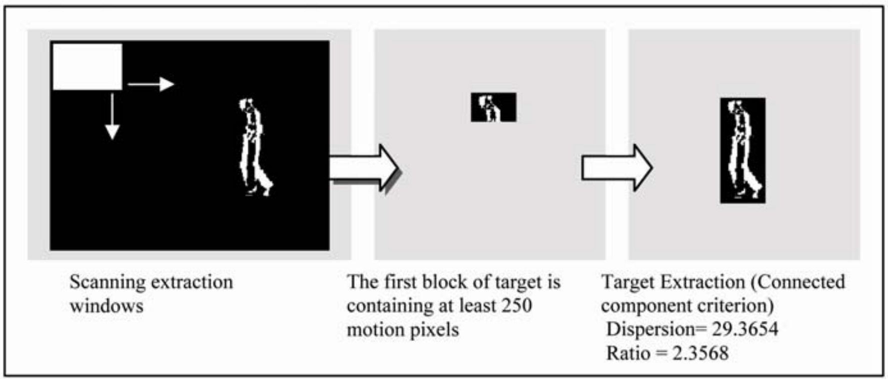

Fig. (2) : Grouping moving objects into motion regions using a connected component criterion.

Fuzzy Motion Region Classification System :

The task of the system is to distinguish the cars from pedestrians from other moving and stationary objects like animals, trees, roads and buildings in the image sequences and identify them as vehicles, human or non-identified object. The principal idea is to exploit useful properties of fuzzy metrical classification in order to provide a robust method to classify motion regions. Indeed, the regions are not always crisply defined, it is sometimes more appropriate to regard them as fuzzy subsets of the image [Bar 92] [Bez 92] [Dzu 01] [Rud 94] [Tiz 97]. The motivation of the use of the geometry features is that is computationally inexpensive and invariant to lighting conditions. On the other hand, it is obvious that the human, with its small and more complex shape, will have larger dispersion than a vehicle (figure 3).

If we define an appropriate membership function μ for the object [Khe 04], the area a and the perimeter p of the object can be calculated as follows:

Area of fuzzy sets:

$$a(\mu) = \sum\mu \qquad (4)$$

Perimeter of a fuzzy set:

$$p(\mu) = \sum_{m=1}^{M} \sum_{n=1}^{N-1} \left| \mu_{mn} - \mu_{m,n+1} \right| + \sum_{n=1}^{N} \sum_{m=1}^{M-1} \left| \mu_{mn} - \mu_{m+1,n} \right| \qquad (5)$$

Where M and N are the dimensions of the image.

Based on the perimeter and the area, the dispersion and the ratio of a fuzzy set can be determined as follows:

$$Dispersion = \frac{(Perimetre)^2}{Area} \qquad (6)$$

$$Ratio = \frac{Length}{width} \qquad (7)$$

The classified motion regions are used as templates for metrical training algorithms. The strategy adopted in this paper is to find the parameters of a fuzzy system by means of learning method obtained from neural networks. A common way to apply a learning algorithm to a fuzzy system is to represent it in special neural-network architecture and train the system using a learning algorithm, such as back propagation.

The fuzzy system is based on two entrances: the dispersion and the ratio of the motion regions, and one exit. For every entrance, we have two fuzzy sets: one for the category of humans and other for the category of vehicles. We have three fuzzy sets for the entrance: one for the human, one for the vehicle and one for no identified objects.

The system leads good performances (98%) over databases of 270 examples where 116 are pedestrians, 124 are vehicles and the rest represent states that are no identified (table 1).

The accuracy of the classification is largely independent of target size, appearance shape or speed. However, the main difficulty with metrical classification is that: when multiple humans close together, they can be misclassified as a vehicle according to the simple metric, if the target is very small, it tends to be rejected as no identified object, and a partly occluded vehicle may look like a human, or some background clutter may appear as a vehicle. To overcome this problem, an additional hypothesis is used. The main idea is to record all potential motion regions PR_n from the first frame of the shot. Each one of these potential regions must be observed along some frames of the shot to determine if they persist or not, and so decide to continue classifying them. To do this, for each new frame, each previous motion region PR_{n-1} is matched to the spatially closest current motion region R_n according to a mutual proximity rule. After this process, each previous potential motion region PR_{n-1} whish have not been matched to current region are removed from the list of accepted motion regions. And any current motion region R_n whish has not been matched is considered new potential region. The metric operators, dispersion and ratio of each frame, are used to update the classification hypothesis [Khe 04]. The most advantage of this method is that if an occluded object is misclassified it will be correctly classified with the passage of time. Another advantage is that the instable motions appearing at the background, such as leaves blowing in the wind, will be misclassified as no-identified regions.

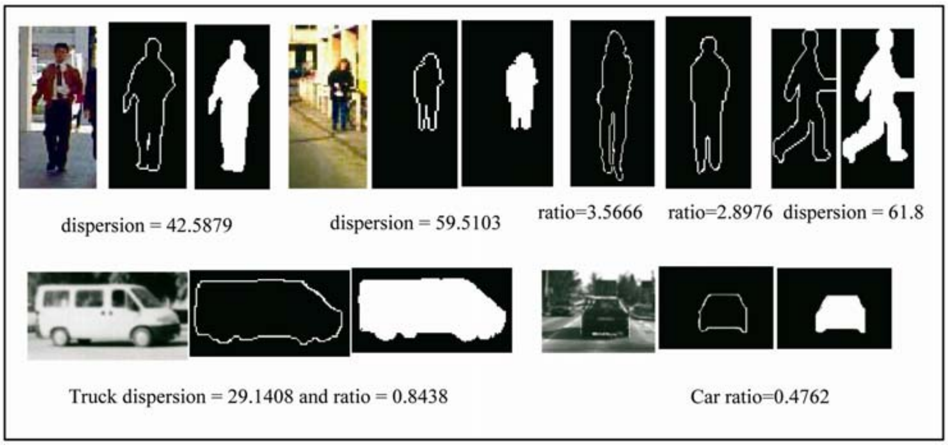

Fig. (3) : Human and vehicles dispersion/ratio values calculated for some image of the learning database.

Table (1)
Results of the learning algorithm

Class	Vehicle	Pedestrian
Dispersion	[17 45]	[23.2 125]
Dispersion Concentration	[20 30]	[30 60]
Ratio	[0.1 1.1]	[0.59 4.47]
Ratio Concentration	[0.2 0.7]	[1.9 3.2]

Indexing and Retrieval

Indexing digital video, based on its content, can be carried out at several levels of abstraction, beginning with indices like the video program name to much lower level aspects of video like the specified motion objects and their locations of the video [Idr 97] [Ray 96].

The interactive retrieval system proposed in this paper includes a query interface sub-module and a query by content retrieval sub-module as shown in figure 1. To facilitate storage and retrieval in visual information systems, flexible data structures should be used. Structures such as R-tree family, R^*-tree, quad-tree, and grid file are commonly used. Each structure has its advantages and disadvantages; some have limited domains and some can be used concurrently with others. To achieve a fast retrieval speed and make the retrieval system truly robust, a quad-tree indexing technique is applied [Khe 04, 05]. The goal of the system is to be able to retrieve a set of sequences, which have motion objects similar to that specified by the query. Reference frames are stored in the database to represent each shot. Image indexing is then applied on the referenced frames. Each frame is indexed according to the descriptors of its moving objects that are defined in the section 4. The metrical features are used as a primary indexing and querying elements

Results :

The system has been applied to large amounts of different video environments where human and vehicular activities are present. Fig. 4 shows some examples of target classification. For single targets, the system provides a robust classification. Note that trees blowing in the wind are completely rejected. Furthermore, the accuracy of the classification is largely independent of target size, appearance shape, speed, lighting conditions or viewpoint. It is also computationally inexpensive. However, when multiple humans close together for a long time, they can be misclassified as a vehicle according to the simple metric. Another limitation of the system is that if the target is very small, less than 4x4 pixels, it tends to be rejected as no identified object. The main problem with vehicles recognition is that when, vehicle is partially occluded for long times, it could be rejected. Also, pedestrians tend to move in close groups that can be misclassified as vehicles according to the simple metric. Fig. 4 and 5 show some results of the system.

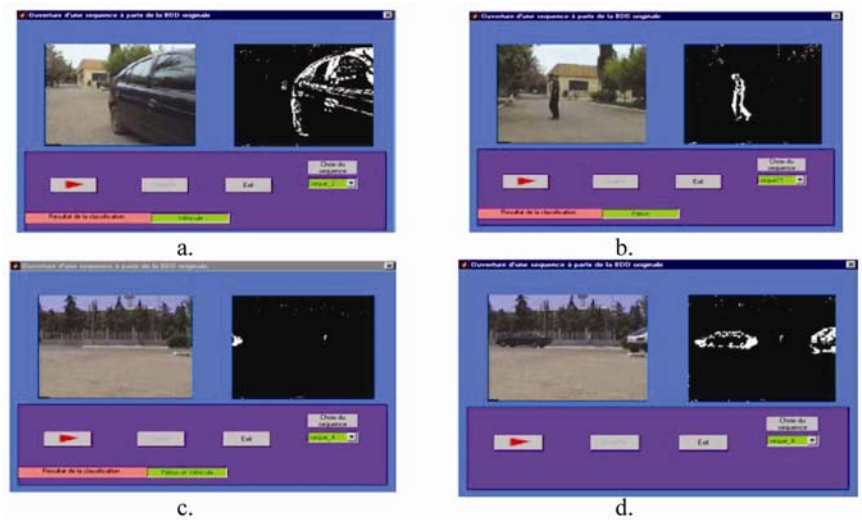

Fig. (4) : Sequences from the ICEPS Laboratory Database automatically segmented and classified as vehicle regions (a), pedestrian regions (b), pedestrian and vehicles regions (c and d)

Fig. (5) : Searching the sequences that contain mobile pedestrian (a), mobile vehicle and pedestrian (b). Vehicle (c and d).

Conclusion :

The work presented here is concerned with motion region detection, classification and indexing moving regions from MPEG surveillance video sequences. The first stage is to decompose the video sequences into shots saving unnecessary decompression. Then, a set of representative frames is selected. The representative frames of a shot are used to the image pre-processing stage in order to generate a collection of moving regions of interest. A robust fuzzy system is proposed to classify moving regions into predefined categories; humans and vehicles, according to image-based properties. Classification is based on simple rules whish are largely independent of appearance or 3D models. Consequently, the metrical classification whish is explored in this paper, is based purely on object's shape, and not on its image content. An additional hypothesis on temporal consistency is used to make the classification system robust to changes of objects appearance and occlusion of motion regions. However, some problems remain to solve: it is necessary to study the problem that when multiple humans close together and when a target is very small.

References :

1. [Bar 92] Bart K., 'Neural Networks and fuzzy Systems.' Prentice Hall. Englewood Cliffs. NJ. 1992.
2. [Bez 92] Bezdek J.C., 'On the relationship between neural networks, pattern recognition and intelligence', *Intenational Jornal of Approximate Reasonning,* Vol 6, p. 85-107, 1992.
3. [Bre 97] Bregler C., 'Learning and recognizing human dynamics in video sequences', *In Proceeding of IEEE CVPR 97*, p. 568-574. 1997.
4. [Bru 99] Brunelli R., Mich O., Modena C. M., 'A survey on the automatic indexing of video data', *Jal. of visual communication and image representation,* Vol 10, p. 78-112, 1999.
5. [Dzu 01] Dzung L. Pham., 'Spatial models for fuzzy clustering', *Jal. Computer vision and image understanding*, Vol 84, p. 285-297, 2001.
6. [Fer 98] Ferman A. M., Murat Tekalp A., 'Efficient filtering and clustering methods for temporal video segmentation and visual summarization', *Jal. of visual communication and image representation,* 99(4), p. 336-351, 1998.
7. [Gud 95] Gudivada V. N., Raghvan V. V., 'Content-based image systems', *IEEE Comput,* 28(9), p.18-22, Sept. 1995.
8. [Hab 99] Habed A., 'Content-based access image and video libraries', Math-info department, Sherbrooke University, 1999.
9. [Idr 97] Idris F., Pandranathan S., 'Review of image and video indexing techniques', *Jal. of visual communication and image representation*, 8(2): 146-166, June 1997.
10. [Ike 01] Iketani A., Nagai A., Kuno Y., Shirai Y., 'Real time surveillance system detecting persons in complex scenes', Jal. of *Real time imaging*, Vol 7, p. 433-446, 2001.
11. [Khe 03] Khelifi S., Boudihir M. E., Nourine R.,'Fuzzy Classification System for outdoor Video Databases Retrieval'. *In Proceeding of AICSSA'03.* IEEE Int. Conf. on Computer Systems and Applications. Gammart. Tunisia. 14-18 July 2003.
12. [Khe 04] Khelifi S., Boudihir M. E., Nourine R., 'Content-Based Video Database Retrieval Using Fuzzy Classification System'. *In proceeding of MediaNet'04. 2nd International Conference on Intelligent Access of Multimedia Documents on Internet.* Tozeur. Tunisia. 25-28 Nov. 2004.

13. [khe 04] Khelifi S., Boudihir M. E., Nourine R., 'Video Database Indexing: an Approach using Fuzzy Classification of Moving Objects in Outdoor Videos'. *In proceeding of MCSEAI'04. 8th Maghrebian Conference on Software Engineering and Artificial Intelligence*. Sousse. Tinisia. p. 555-566. 9-12 May 2004.

14. [khe 04] Khelifi S., Boudihir M. E., Nourine R., 'Fuzzy Classification System for Telesurveillance Databases Retrieval and Indexing'. *In proceeding of International IEEE/APS Conference on Mechatronics and Robotics.* Aachen. Germany. p. 20-25. 13-15 Sep. 2004.

15. [khe 05] S. Khelifi, M. Elarbi Boudihir, R. Nourine, 'Compressed Telesurveillance Video Database Retrieval Using Fuzzy classification System'. *In Proceeding of Springer Verlag Berlin Heidelberg Germany, Lecture Notes in Computer Science (LNCS)*, ISBN: 3-540-29069-9, Vol. 3656/2005, p. 575-584, 2005.

16. [Kru 98] Krüger S., 'Motion analysis and estimation using multi-resolution affine models', Thesis submitted at the university of Bristol, July 1998.

17. [Lip 98] Lipton A. J., Fujiyoshi H., Patil R. S., 'Moving target classification and tracking from real-time video', *Submitted to IEEE WACV 98*, 1998.

18. [Nib 93] Niblack W., 'The QBIC project: querying images by content using colour, texture, and shape. Storage and Retrieval for Image and Video Databases', *In proceeding of the SPIE. no 1908.* San Jose California SPIE, Bellingham, p.173-187, Feb. 1993.

19. [Ray 96] Raymond Ng., Sedighian A., 'Evaluating multi-dimensional indexing structures for images transformed by principals component analysis', *In Proceeding of SPIE Storage and retrieval for image and video databases,* 1996.

20. [Rud 94] Rudolf K., Gebhardt. J. and Klowonn. F., 'Fundations of fuzzy systems'. John Wiley and Sons Ltd. Chichester. 1994.

21. [Smo 94] Smoliar S. W., Zhang H. J., 'Content-based video indexing and retrieval', *In proceeding of IEEE Multimedia,* 1(2), p.62-72, Summer 1994.

22. [Sch 00] Schonfeld D., Lescu D., 'VORTEX: video retrieval and tracking from compressed multimedia databases- multiple object tracking from MPEG-2 bit stream', *Jal. of visual communication and image representation,* Vol.11, p. 154-182, 2000.

23. [Shn 96] Shneier M. and Abdel M. M., 'Exploiting the JPEG compression scheme for image retrieval', *In Proceeding in IEEE Trans. Patt. Anal. Mach. Intell.* 18(8), p. 849-853. August 1996.

24. [Tiz 97] Tizhoosh H., 'Fuzzy image processing', Springer, 1997.

25. [Yeo 95] Yeo B.-L. and Liu B. 'A unifiedapproach to temporal segmentation of motion JPEG and MPEG compressed videos', *In proceeding of the International Conference on Multimedia Computing and Systems*, p. 81-88, May 1995.

26. [Yeo 96] Yeo B.-L. and Liu B. 'Efficient processing of compressed images and video', Ph.D. thesis, Dept. Of Electrical Engineering, Princeton University, Jan. 1996.

نظام قياسي ضبابي لفهرسة قواعد بيانات أنظمة المراقبة عن بعد

سامية خليفي*, محمد العربي بودهير**, رشيد نورين***

*قسم علوم الحاسب، جامعة الملك فيصل، الدمام، المملكة العربية السعودية

**قسم علوم الحاسب والمعلومات، جامعة الإمام محمد بن سعود ، الرياض، المملكة العربية السعودية

***مركز أبحاث التحكم الذكي و نظم الطاقة، الجزائر

الملخص:

نقترح في هذا البحث بناء نظام استرجاع مرئي من قواعد بيانات تحتوي على فيديو مضغوطة MPEG خاصة بنظام مراقبة خارجية. الهدف هو تحديد الأشياء المتحركة. بعد ذلك يتم تصنيف هذه المناطق المتحركة إلى فئات محددة وفق الخصائص الأساسية للصور والمؤشرات القوية الموجودة فيها.

الفكرة الرئيسية التي تم تطبيقها هي الجمع بين الخاصيات القياسية ومفهوم التماثل الزمني. تكمن خصوصية طريقتنا في استخدام مفاهيم الهندسة الضبابية التي تضمن أحسن أداء في عملية تصنيف الأشياء المتحركة إلى ثلاث فئات: فئة المترجلون, فئة السيارات و فئة الأشياء غير المعروفة. و قد تم اقتراح التماثل الزمني كطريقة فعالة لمعالجة التغييرات التي تحدث على مظهر الأشياء المتحركة من جهة ومن جهة أخرى لإيجاد الحلول الفعالة للمشاكل التي قد تحدث نتيجة تداخل و تطابق هذه الأشياء المتحركة. تستعمل الأشياء المتحركة التي وقع تصنيفها كأدوات قياس على مستوى خوارزميات التدريب القياسي وأيضا كمفاتيح في بناء تقنية الفهرسة الشجرية.