

A 3D approach to Co-occurrence Matrix Features used in Dynamic Textures Indexing

Muguras Mocofan^{1, a} and Radu Vasile^{2, b}

¹Politehnica University of Timisoara, Bd. Vasile Parvan, Nr. 2, 300323, Timisoara, Romania

²Politehnica University of Timisoara, Bd. Vasile Parvan, Nr. 2, 300323, Timisoara, Romania

^amugur@cm.upt.ro, ^bradu@cm.upt.ro

Keywords: dynamic texture, image databases, image processing.

Abstract. The area of applications of dynamic texture is increasingly wide: video surveillance, transaction systems, medical application and video synthesis. The paper presents an indexing model in large databases of dynamics texture using the co-occurrence matrix features. The data from the video sequence that represents the dynamic texture are loaded in a 3D matrix. The application of the co-occurrence matrix is performed for each frame of the data parallelepiped covered in 3 directions. This enables/facilitates the integration of the temporal features of the dynamic texture in the mathematical description of the behaviour. Additionally, we use more translations to compute the indexing vector from the 2D+T space of dynamic textures.

Introduction

The multimedia indexing is a significantly active research theme in the community of image processing. Beyond indexing of textual records, currently, it is able to index video content, video or, generally speaking, multimedia. Indexing corresponds to the organization of data according to an order defined by one or more attributes. These are similar to signatures which specify the contents of the image or video, often linked to the characteristics as visual colour, texture or shape. One of the most important features of the image is the texture. The characterization of the static texture (2D) has been extensively studied, and is part of the descriptors of the MPEG-7 [1]. Either dynamic or temporal, the texture is a spatially repetitive, time-varying visual pattern that forms an image sequence with certain temporal stationary. In dynamic texture, the notion of self-similarity central to conventional image texture is extended to the spatiotemporal domain [2].

The extension of these visual features to the temporal dimension poses certain difficulties. The notion of texture in image sequences raises many problems. The 2D textures plus time dimension are in many cases a simple extension of the 2D structures to the 3D.

One may wonder also if in the area of video, the concept of 2D texture + T is a relevant descriptor, because the motion information represents a real contribution to the recognition. Moreover, the movement can extract new descriptors for indexing, such as measures related to accelerating, turbulence or vortices for the fluid. It is therefore reasonable to think that the extension of texture to the time domain will have its contribution to video indexing.

In the specialised literature, the temporal extension of the notion of texture is found as the dynamic, more rarely as temporal texture. A flag in the wind, a field of waving grass, the waves of the sea, the lake surface, the movement of the drill, smoke, fire, an ant colony, the wings of a windmill turn, fountains, waterfalls, are all examples of dynamic textures presented in the literature [3, 4, 5].

The dynamic textures are found in many natural and artificial scenes, and therefore, any video indexing system must be able to analyse and characterize these dynamic textures. Increasing computer power and the importance of the video data stream has generated the recent outbreak of numerous studies on the dynamic textures. Research topics and potential applications of dynamic textures are indeed numerous:

- Video indexing: presenting the scenes of dynamic textures is very common and their recognition would facilitate the indexing step. The extraction of relevant features of the dynamics of textures sets the requests to semantics ("turbulent flow", "large sea waves", ...) [6].

- Video surveillance: an example application involving dynamic textures for the monitoring of departs of fires in forests: an advanced analysis of dynamic textures like smoke or fire is necessary in order to report a fire and not false alarms.

- The spatial-temporal segmentation: it is an important step in video analysis, as it may, subsequently, be used to make a video resume or detect a disturbance in a dynamic texture. It can be applied to the dynamic subtraction of background [7].

- Monitoring: in some applications, the object to follow can be like a dynamic texture (tracking a vortex in a fluid) [8].

- Video synthesis: this is an important application area of dynamic textures, including video games or animated films: realistic rendering of a fire, animal fur, the movement of waves on the surface of the sea [9].

In this paper, we focus on the study and characterization of dynamic textures with the goal of indexing in large databases of videos. A relevant characterization will also lead to other applications, such as spatial-temporal segmentation or removal of dynamic backgrounds.

Since dynamic textures are very complex, it is crucial to be able to simplify the understanding of the underlying phenomena. The 3D model of dynamic textures that we propose involves the use of the co-occurrence matrix features such as contrast, homogeneity, correlation, directivity, entropy calculated for 4 directions: front to back, left to right, top to bottom and for the 3D spatial translations. These are related to the human visual system.

The Spatial Grey Level Dependence Method (SGLDM) is one of the most important statistical texture description methods, especially in medical image analysis. Co-occurrence matrices are employed for the implementation of this method; however, they are inefficient in terms of computational time and memory space, due to their dependency on the number of gray levels (gray-level range) in the entire image [10]. Their inefficiency puts up barriers to the wider utilization of SGLDM in a real application environment.

Currently, the load of a 10-second video sequence in the RAM memory of a computer is no longer difficult. Although the dynamic texture indexing is slower, the use of the targeted vector (that describes efficiently the texture behaviour) in the research process compensates the indexing. The reduction of its computational time and storage requirements should be an aim of continuous research.

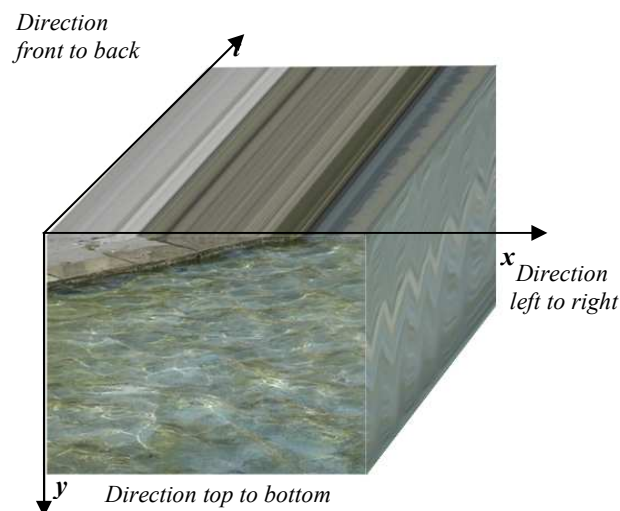


Figure 1. 3D representation for dynamic textures from DynTex database

The co-occurrence matrix features for dynamic textures

The co-occurrence matrix contains the L -nd order spatial averages. For a translation t , the co-occurrence matrices \mathbf{MC}_t of a region R are defined for all couplets of grey levels (a, b) through the relation:

$$\mathbf{MC}_t(a, b) = \text{card}\{(s, s+t) \in R^2 \mid A[s] = a, A[s+t] = b\} \quad (1)$$

As a result, $\mathbf{MC}_t(a, b)$ represents the number of pixel couplets $(s, s+t)$ from the considered region $A[s]$, separated through the translation vector t .

The co-occurrence matrices contain a lot of information, but it is difficult to deal with all items of information [11]. Generally, there are some characteristics of the image that are used:

$$\text{Homogeneity} = \frac{1}{N_c} \sum_a \sum_b (\mathbf{MC}_t(a, b))^2 \quad (2)$$

$$\text{Contrast} = \frac{1}{N_c (L-1)^2} \sum_{k=0}^{L-1} k^2 \sum_{|a-b|=k} \mathbf{MC}_t(a, b) \quad (3)$$

$$\text{Entropy} = 1 - \frac{1}{N_c \cdot \ln(N_c)} \sum_a \sum_b \mathbf{MC}_t(a, b) \cdot \ln(\mathbf{MC}_t(a, b)) \cdot \mathbf{1}_{\mathbf{MC}_t(a, b)}$$

$$\mathbf{1}_{\mathbf{MC}_t(a, b)} = \begin{cases} 1, & \text{if } \mathbf{MC}_t(a, b) \neq 0 \\ 0, & \text{if not} \end{cases} \quad (4)$$

$$\text{Correlation} = \frac{1}{N_c \cdot \sigma_x \cdot \sigma_y} \left| \sum_a \sum_b (a - m_x) \cdot (b - m_y) \cdot \mathbf{MC}_t(a, b) \right| \quad (5)$$

The first step is to evaluate the co-occurrence matrices features from the dynamic texture indexing perspective. We compare several approaches and different descriptors. The main objective is to evaluate these approaches and identify the most relevant descriptors.

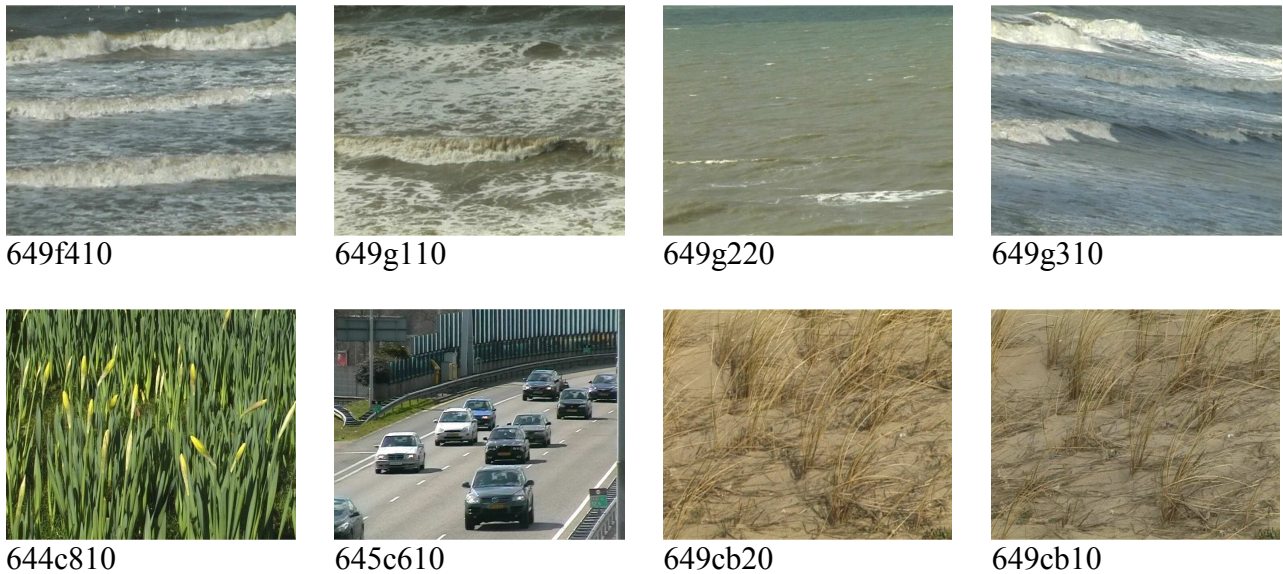


Figure 2. Examples of dynamic textures from DynTex database

The three databases used for our experimentations are made from DynTex videos. These databases differ in their difficulty, number of classes and number of elements:

- Alpha Database: 60 image sequences of dynamic textures grouped into three relatively simple classes: sea, herbs and trees.
- Beta Database: 162 image sequences of dynamic textures grouped into 10 classes: sea, herbs, trees, flags, still water, fountains, smoke, escalator, traffic, rotate.
- Gamma Database: 275 image sequences of dynamic textures grouped into 11 classes: flowers, sea, trees without leaves, dense foliage, escalator, calm water, flags, herbs, traffic, fountains and fire. In this database, the classes are equipped with many samples to cover many issues (change of scales, orientation). This is a very complex data set [12].

A 3D approach to the co-occurrence matrix features

Each of our experiments in indexing is conducted in the following manner:

- Analysis of image sequences (spatial-temporal approach);
- Descriptor calculation and construction of a feature vector for their indexing;
- Classification of signatures.

For instance, below are some examples of dynamic textures analyzed on the basis of the co-occurrence matrix features; in this paper we randomly chose the dynamic texture 649cb10, which is a field of dry grass windy. For this dynamic texture, the contrast variations are very small for all translations, if the direction from the front to back is used. The greater contrast value corresponds to a larger translation, but is not a rule for the entire DynTex database.

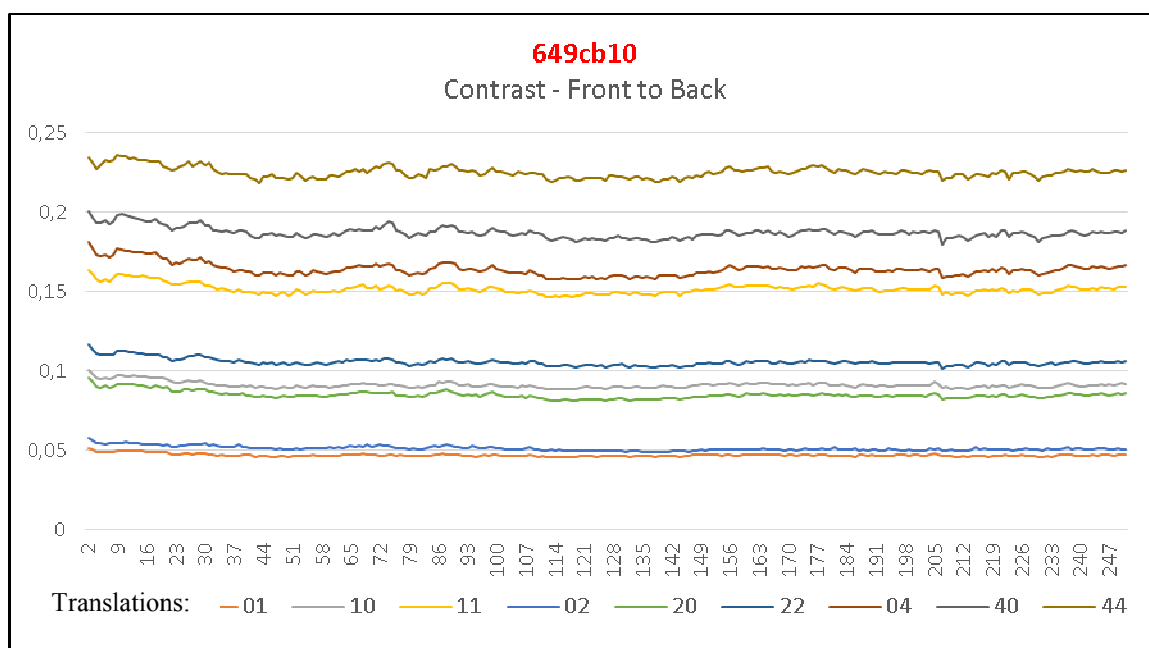


Figure 3. The contrast evolution of the co-occurrence matrix feature for the 649cb10 video from DynTex database in direction front to back 250 frames (720x576) and 9 translations

The contrast features of the co-occurrence matrix calculated in direction from left to right highlight in the image where we have texture of grass and the areas where is no grass, which cannot be revealed by the direction front to back (649cb10 dynamic texture from DynTex database).

For a scroll in direction top to bottom of the contrast co-occurrence matrix we find a high level of uniformity that differs greatly from contrast co-occurrence matrix calculations that consider translating not in plan, but between successive frames. We found a much higher sensibility for the features of the co-occurrence matrix characteristics for direction between frames, 3D type.

We calculated the following characteristics of the co-occurrence matrix: correlation, homogeneity and directivity.

Directivity in the left to right direction gives best results to present a texture with vertical elements. Agitation of grass in the wind is emphasized by the directivity calculated for translating 3D4 (translation between 4-frame is considered).

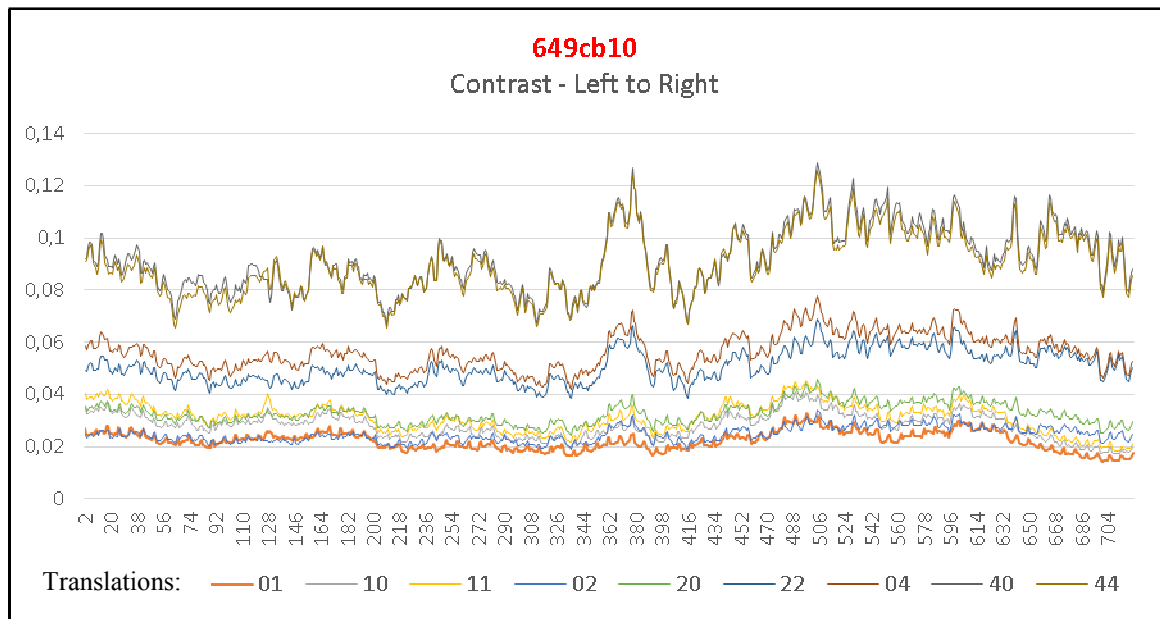


Figure 4. The contrast evolution of the co-occurrence matrix feature for the 649cb10 video from DynTex database in direction left to right - 720 frames at (250x576) and 9 translations

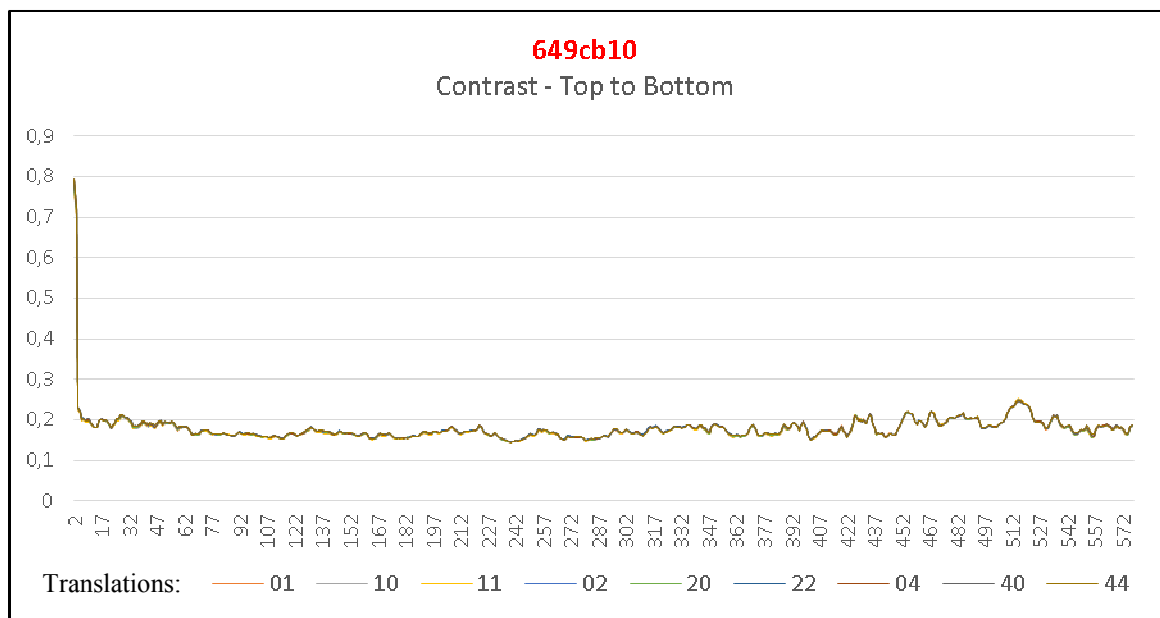


Figure 5. The contrast evolution of the co-occurrence matrix feature for the 649cb10 video from DynTex database in direction top to bottom - 576 frames at (250x720) and 9 translations

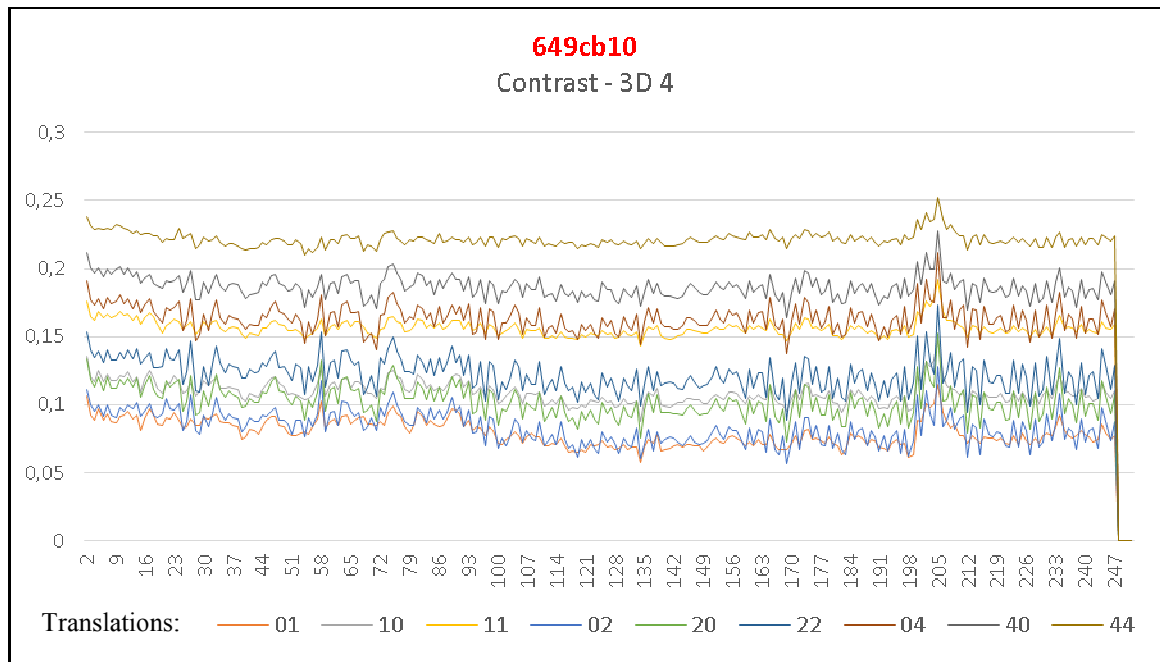


Figure 6. The contrast evolution of the co-occurrence matrix feature for the 649cb10 video from DynTex database in direction front to back with 4 frames translation (3D4)

We studied the time evolution of dynamic textures. Figure 7 shows a high degree of variation of the average characteristics of co-occurrence matrix, which means that the video sequences contain very complex dynamic textures. We present the mean of the contrast features of the co-occurrence matrix for 250 frames where only the translation is used (1,1).

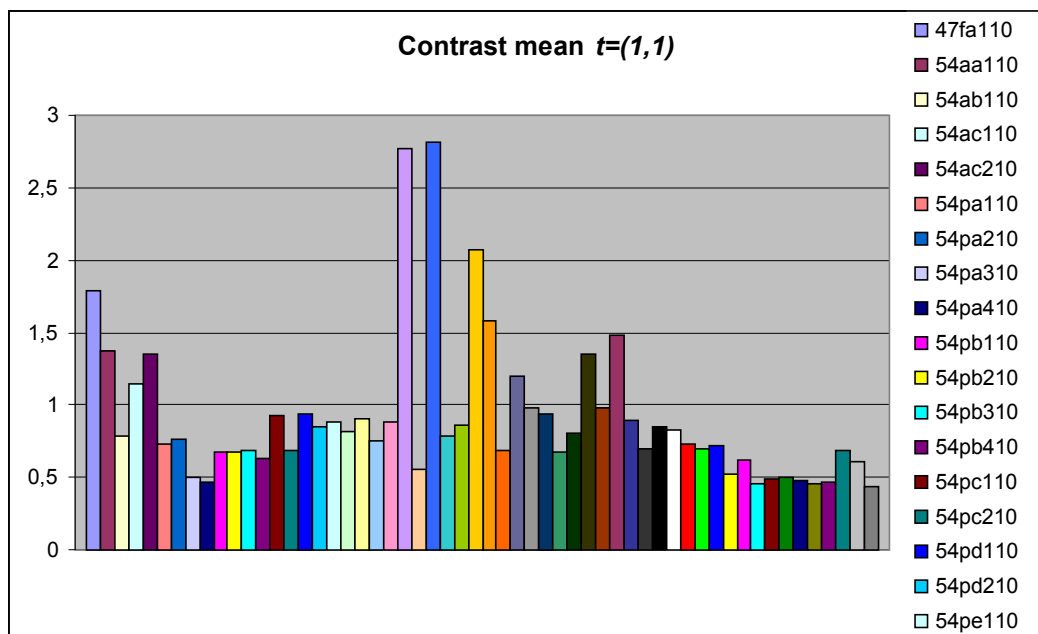


Figure 7. The mean value of the contrast feature for a few dynamics textures from DynTex database

There is a close connection between the degree of motion in video sequence and the dispersion of characteristics. The mean and dispersion of characteristics are presented only for translation of (1,1).

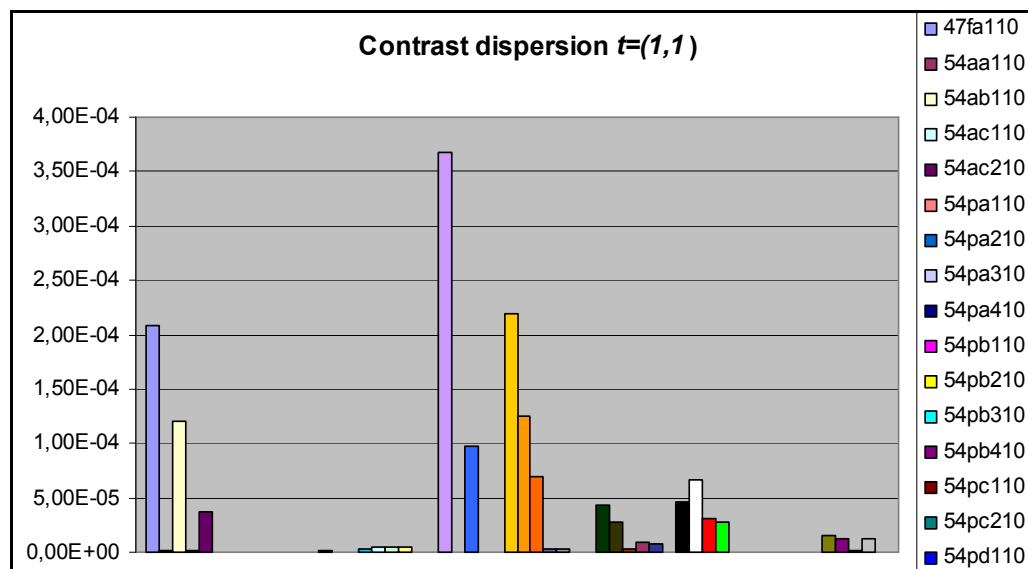


Figure 8. The dispersion value of the contrast feature for a few dynamics textures from DynTex database

We present the evolution in time of 649cb10 dynamic texture from DynTex features of the co-occurrence matrix. Due to space reasons, in the figures 9 and 10, we trace the correlation, homogeneity, directivity only for 250 values, using the same translations like in figures 3, 4, 5 and 6.

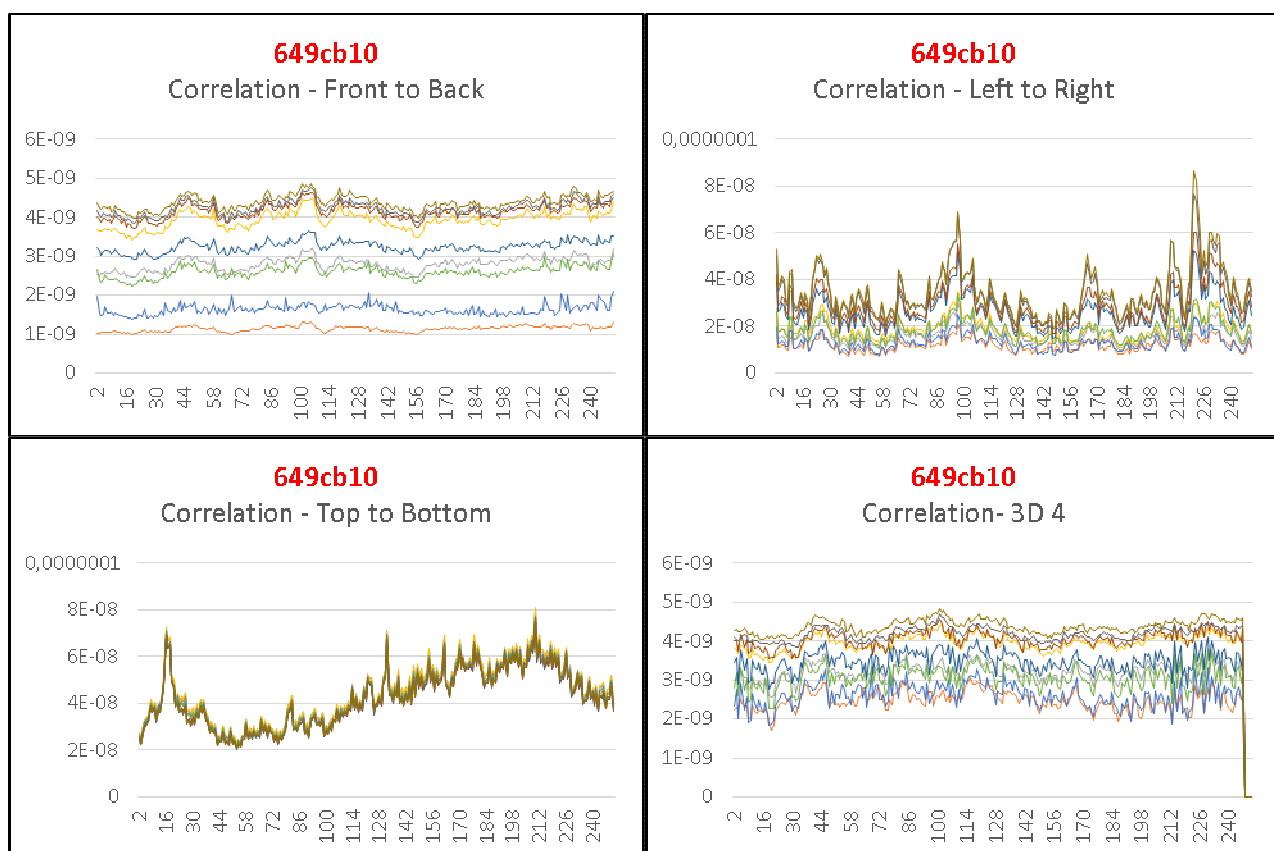


Figure 9. The correlation evolutions of the co-occurrence matrix feature for the 649cb10 video from DynTex database in 4 directions

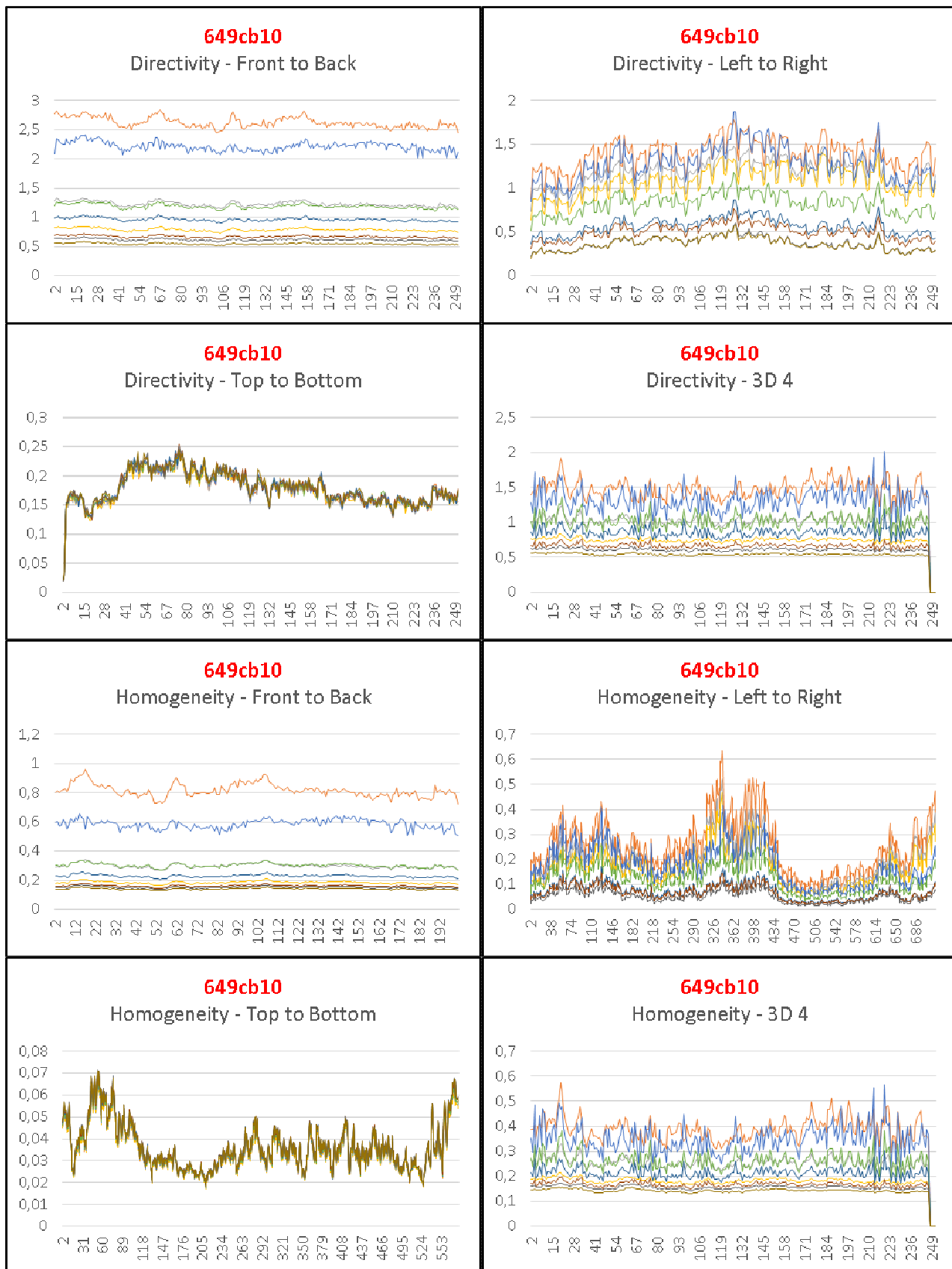


Figure 10. The directivity and homogeneity evolutions of the co-occurrence matrix feature for the 649cb10 video from DynTex database in 4 directions

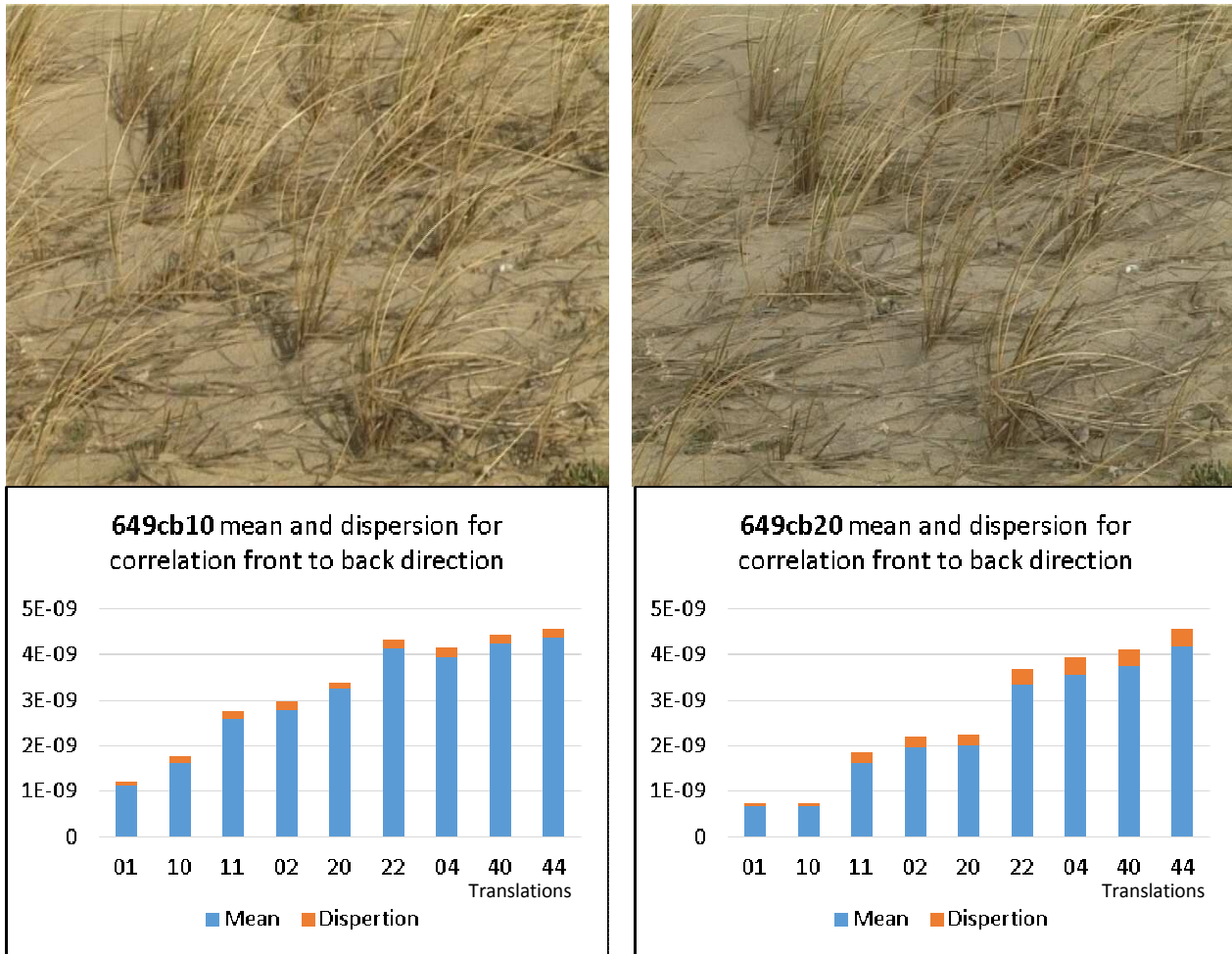


Figure 11. Mean and dispersion for correlation of the co-occurrence matrix feature in direction front to back for 649cb10 and 649cb20 videos from DynTex database

This paper shows that using these features is useful for describing dynamic textures, and particularly for the definition of similarity between dynamic textures. For example, the mean square error between the 649cb10 and 649cb20 correlation mean is $2.31 \cdot 10^{-9}$, and for dispersion is $3.67 \cdot 10^{-10}$.

Construction of indexing features vectors

We have analysed the co-occurrence matrix features for translations in several directions and distances by the following pairs: (0,1), (1,0), (1,1), (2,1), (1,2), (2,2), (0,2), (2,0), (0,4), (4,0), (2,4), (4,2), (4,4) [13].

After more analyses using a 3D approach in 3 directions, the conclusion was: too many pairs, big complexity without an increase of the quality characterization [14]. In literature we find different approaches for indexing dynamic textures. One of them uses the multi-resolution wavelet transform in the space 2D+T. The method provides an acceptable recognition rate on all databases [3]. When looking at the dimensions of vector characteristics, depending on the multi-resolution method employed, we observe that for some methods, the number of descriptors is greater than the number of samples to classify. This poses several difficulties; the samples that have to be classified contain too much redundant information and can deteriorate the classification or the multi-scale approaches that are not in the same conditions. Indeed, the classification of 10 classes in a three dimension space is not as difficult as it is in a space that has 5508 dimensions, and the approach classification used is not necessarily adapted. To correct this problem, it is important to reduce the size of signatures. With the aim of improving the indexing performance, we could also search in more details the invariance of our descriptors. Indeed, at present, our vectors are by construction invariant features in translation and rotation, but not in scale.

For the indexing of dynamic textures in multimedia databases, we argue in favour of using co-occurrence matrix features in the following manner:

- Use the translations pairs: (0,1), (1,0), (1,1), (2,2), (0,2), (2,0), (0,4), (4,0), (4,4);
- Use followings directions: front-to-back, left-to-right, top-to-bottom and the proposed 3D approach (tested for 1, 2, 4, 8 and 16 frames translations – recommended 4 frames distance).
- Indexing vector is obtained by concatenating the mean and variance values calculated for each translation and direction for all frames.

The index vector length is equal with the result of multiplication of the 9 translations used, in 4 directions, for 5 features of the co-occurrence matrix and with the two average values of the features (mean and dispersion). The length of the proposed indexing vector is 360, it is not a small value but it describes very well the dynamic textures.

For less demanding, one can calculate the average and dispersion of the following characteristics: homogeneity, contrast, entropy, directivity and correlation only for the translation (1,1). Mean and dispersion values are calculated by evaluating the individual characteristics of the co-occurrence matrix corresponding to each frame of video sequence containing dynamic textures. Using a small number of features 2-3 doesn't give good results; we noticed a 28% decrease for 2 features and 21% for three features compared to the situation when all 5 were used [13].

The number of video frames from the sequence where the features are computed has a major influence on the results. Here we found that, using a large number of frames, for textures with uniform motion, defines a better precision of the texture behaviour. Instead, if the analysed sequence contains video camera moves or new objects enter in scene, the mediation for more frames doesn't provide an adequate description and good results. In such situations, one may also analyse the short-term characteristics variation.

Analysis in space and time is achieved by calculating the co-occurrence matrix features using a temporal axis translation, which is made between pixels from successive frames, this time translation is expressed by the temporal distance between frames. We experienced T axis movement and mixed movement 2D + T. This approach offers a better interpretation for the movement presence in dynamic textures.

If video sequences containing dynamic textures like grass windy or ocean waves affected by wind with different speeds, the use of 3D4 direction in the proposed indexing method give a good detection of similarity between textures (the case presented in the paper). This 3D4 direction made a clear discrimination of them. The advantages offered by the introduction of this new 3D4 direction are very difficult to be predicted when we refer to a very large dynamic textures database with varied content. But for cases like the one presented in the paper, the use of this direction allows a good discrimination for dynamic textures with a great similarity between them.

Conclusions

One of the main concerns of our study has been the construction of tools for spatial-temporal analysis adapted to the dynamic texture. As for the static texture in the image domain, it is essential to take into account the properties of dynamic textures in order to characterize them effectively.

Analysis of video sequences using co-occurrence matrix features for various translations (in different directions and at different distances) gives us very useful information about the behaviour of dynamic textures.

Additionally, it provides a good description of the texture changes in time, keeping the usual correlation with verbal expressions of what granularity, contrast and entropy are.

Further work will be orientated on the multiresolution decomposition for the dynamic textures simultaneously with the attempt to detect a small texture part from a video sequence.

Acknowledgment

This paper was supported by the project "Develop and support multidisciplinary postdoctoral programs in primordial technical areas of national strategy of the research - development - innovation" 4D-POSTDOC, contract nr. POSDRU/89/1.5/S/52603, project co-funded from European Social Fund through Sectorial Operational Program Human Resources 2007-2013.

References

- [1] P. Wu, Y.M. Ro, C.S. Won & Y. Choi, Texture Descriptors in MPEG-7, International Conference on Computer Analysis of Images and Patterns (CAIP 01), (2001), pp. 21–28
- [2] R. Peteri and D. Chetverikov. Dynamic texture recognition using normal flow and texture regularity. In Proc. Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 2005), Estoril, Portugal, (2005).
- [3] S. Dubois, R. Peteri, A Michel Menard, A Comparison. of Wavelet Based Spatio-temporal Decomposition Methods for Dynamic Texture Recognition, Proceedings of the 4th Iberian Conference on Pattern Recognition and Image Analysis, voa de Varzim, Portugal, (2009), pp. 314–321
- [4] G. Zhou, N. Dong, Y. Wang, Non-Linear Dynamic Texture Analysis and Synthesis Using Constrained Gaussian Process Latent Variable Model Circuits, PACCS '09 Pacific-Asia Conference on Communications and Systems, Chengdu, China, 16-17 May (2009), pp. 27–30.
- [5] M. Szummer, Temporal Texture Modeling. Technical Report 346, MIT (1995)
- [6] S. Dubois, R. Peteri, A Michel Menard, A 3D discrete curvelet based method for segmenting dynamic textures, Proceedings of the 16th IEEE international conference on Image processing ICIP'09, Cairo, Egypt, (2009), pp. 1365–1368.
- [7] R. Peteri, D. Chetverikov, Qualitative Characterization of Dynamic Textures for Video Retrieval, International Conference on Computer Vision and Graphics ICCVG 04, Warsaw, Poland, September (2004), pp. 33–38.
- [8] T. Bouwmans, F. El Baf, B. Vachon, Statistical Background Modeling for Foreground Detection: A Survey, Handbook of Pattern Recognition and Computer Vision, World Scientific Publishing, Volume 4, Part 2, Chapter 3, January (2010), pp. 181-199.
- [9] R. Peteri, Tracking Dynamic Textures using a Particle Filter Driven by Intrinsic Motion Information, Machine Vision and Applications, January (2010), pages 1–9.
- [10] A.E. Svolos, A. Todd-Pokropek, Time and space results of dynamic texture feature extraction in MR and CT image analysis, Information Technology in Biomedicine, IEEE Transactions on , vol.2, no.2, June (1998), pp.48-54.
- [11] J. P. Cocquerez, S. Philipp, Analyse d'images: filtrage et segmentation, Mason, Paris (1995)
- [12] R. Peteri, S. Fazekas, M. Huiskes, A Comprehensive Database of Dynamic Textures, Pattern Recognition Letters, vol. 31, (2010), pp. 1627–163.
- [13] M. Mocofan, R. Vasiu, Dynamic textures indexing using the co-occurrence matrix features, Applied Computational Intelligence and Informatics (SACI), 2012 7th IEEE International Symposium on , Timisoara, Romania, (2012), pp.327-330.
- [14] M. Mocofan, F. Alexa, Analysis of dynamic textures using a 3D approach for the co-occurrence matrix features, Electronics and Telecommunications (ISETC), 2012 10th International Symposium on , Timisoara, Roania, (2012), pp.235-238.