International Journal of Engineering & Technology, 7 (4) (2018) 5196-5205

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.12700
Research paper

Exploiting tensor-space similarity measures in image and video processing

Adib Akl [1] *, Charles Yaacoub [1]

[1] Faculty of Engineering, Holy Spirit University of Kaslik (USEK), P. O. Box 446, Jounieh, Lebanon
**Corresponding author E-mail: adibakl@usek.edu.lb*

Abstract

Measuring the similarity between images is a crucial process for several image and video processing applications. For instance, it is used in image inpainting, image retrieval, pattern recognition, as well as image and video compression. Traditional intensity-based approaches have shown efficiency in different situations and scenarios. However, failures still exist, especially in the case of high local structural variation images. In this paper, different tensor-based metrics that reveal the local structural information are presented and evaluated in several dissimilarity estimation contexts. Results show that using these metrics, relevant dissimilarities of local patterns can be better detected, thus helping standard intensity-based image inpainting algorithms, motion estimation methods and the analysis of medical images.

Keywords: *Gradient Fields; Image Analysis; Local Structures; Motion Estimation; Similarity Measure; Structure Tensor*.

1. Introduction

Image and video analysis techniques have gained attention in the research community, due to the diversity and importance of their applications. For instance, they are used in image synthesis [1-3], image inpainting [4] and video compression [5]. It is also used in medical imaging for brain tumor and bone diseases detection [6-12]. Other applications include identifying regions of interest and retrieving similar regions in remote sensing [13], [14], image compression [15], [16] and composite materials simulation and design [17], [18].
Several methods have been proposed for image and video analysis [1], [3], [19 - 27]. Regardless of the application, most of them rely on the use of similarity estimation measures and/or statistical information. For instance, Paget and Longstaff [20] proposed a method for image modeling based on estimating the conditional probability density function. This method is able to capture the visual characteristics of the image into a statistical model describing the pixels interactions. Portilla and Simoncelli [21] used the wavelet transform for parameterizing the image by statistical coefficients that correspond to several functions at neighboring positions. Peyre [3] used the grouplet transform to analyze the structural layer of the image. Pless and Simon [22] proposed a method for image analysis based on image similarity using locally linear embedding and isomap. In this method, comparisons or distances are only meaningful between nearby points. Bhandarkar and Chen [23] proposed an image similarity analysis approach for video sequences using artificial neural network, where the similarity measure is based on correlation. The methods of Akl and Iskandar [19], [24] relied on estimating the texture pattern size in the images using geometrical information. Hassani et al [12] studied the anisotropy of the image textures using Morlet wavelet and the Renyis entropies [28] as inputs for a neural network. Sevestre-Ghalila et al [10] proposed a morphological study establishing a basic frame of the image local architecture. A one-dimensional local binary pattern descriptor was used by Houam et al [29] for texture characterization. Fractal-based analysis methods have also been proposed [30-32].
Most of these methods have shown success in different situations. However, unsatisfactory results are always present when treating images of high local structural variations.
This paper deals with the estimation of image similarities using structure-based measures, mainly the structure tensor field which represents local orientations, the degree of anisotropy and the energy. The advantage of using these structural measures, in image inpainting, motion estimation (for video compression), and medical image analysis, is evaluated and discussed.
The remainder of this paper is organized as follows. The structure tensor field computation and the tensor-based similarity metrics are presented in section 2. Section 3 details the proposed exploitation of the tensor-based similarity metrics in image inpainting, in motion estimation, and in the analysis of medical images. Practical results are discussed in section 4. Finally, conclusions and prospects are drawn in section 5.

2. The structure tensor

2.1. Structure tensor field computation

The structure tensor, also known as the second-moment matrix, is a mathematical toot that summarizes the gradient directions in the neighborhood of a pixel, and the extent of the coherence of those directions [2], [33], [34]. It is a symmetric matrix, typically used in image processing for edge and local patterns representation.
The structure tensor field T of an image Y, which assigns a structure tensor T(u) for each site u of Y, is computed from estimated gradient fields $GY = [Y_x\ Y_y]$ where $Y_x = Y * H_x$ and $Y_y = Y * H_y$; H_x and H_y are Gaussian derivative kernels and '*' denotes convolution. The structure tensor field [2],

$$T = \begin{pmatrix} T_{xx} & T_{xy} \\ T_{xy} & T_{yy} \end{pmatrix}, \tag{1}$$

Is computed by spatial smoothing of the gradient outer product:

$$T = Q * GY\ (GY)^t = Q * \begin{pmatrix} Y_x^{\ 2} & Y_x Y_y \\ Y_x Y_y & Y_y^{\ 2} \end{pmatrix}, \tag{2}$$

Where Q is a Gaussian averaging kernel and 't' is the transpose operator.

Note that spatial smoothing is used to increase the robustness of the structure tensor against artifacts and noise. It is important to mention that the structure tensor field T has the same size as the image Y except the outer edges (contours) where tensor values are discarded.

At each position u, an eigenvalue decomposition of T(u) results in two eigenvectors $T^1e(u) = [T^1e_x(u)\ ,\ T^1e_y(u)]$ and $T^2e(u) = [T^2e_x(u)\ ,\ T^2e_y(u)]$ that point orthogonal and parallel to the local edge, respectively, and two eigenvalues $\gamma_1(T(u))$ and $\gamma_2(T(u))$ that represent the strength of the intensity change.

It is common to represent the structure tensor T(u) as an ellipsoid which orientation $O_r(T(u))$ is calculated from eigenvector $T^1e(u)$ associated to $\gamma_1(T(u))$ and varies between -π/2 and π/2:

$$O_r(T(u)) = tan^{-1}\left(\frac{T^1e_y(u)}{T^1e_x(u)}\right), \tag{3}$$

And which ellipticity is given by the ratio of the eigenvalues [33].

A relevant indicator based on the structure tensor is the coherence factor which is obtained from the eigenvalues and varies between 0 and 1:

$$C_h(T(u)) = \frac{\gamma_1(T(u)) - \gamma_2(T(u))}{\gamma_1(T(u)) + \gamma_2(T(u))}. \tag{4}$$

Coherence is used to estimate the local anisotropy and it characterizes the local variability of image geometries.

The energy of the structure tensor, defined as the sum of its eigenvalues (i.e. the expectation of the squared amplitude of the gradient), is also an essential factor characterizing the dynamics. In other words, the energy of the structure tensor at a certain position expresses the local contrast in the vicinity of this position [33].

In the sequel, the structure tensor field will be represented by its local orientations image. Fig. 1 presents an example of structure tensor field calculation; from left to right: image showing structured patterns, its structure tensor field represented by its orientation image and the palette used for orientations representation.

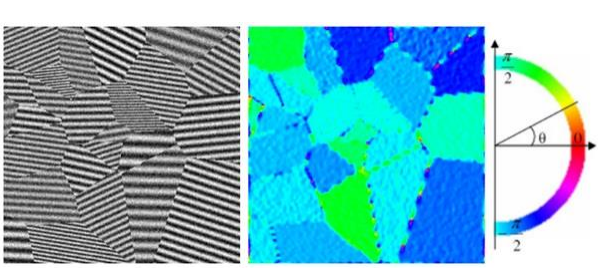

Fig. 1: Example of Structure Tensor Field Calculation. Left to Right: Input Image, Its Structure Tensor Field Orientation Image and the Palette Used for the Representation of Orientations.

Note that this palette is used in all the results that follow. It can be clearly seen in the orientation image that the structure tensor field reflects the changes of the paving directions.

2.2. Tensor-space similarity metrics

To measure the similarity between two structure tensors $T_1(u)$ and $T_2(u')$ (respectively, on sites u and u'), three different tensor-space metrics are used: the Square of the Euclidean Distance (SED), the Orientation and Coherence metric (O_rC_h) and the Frobenius Norm (FN).

The SED is computed on the components of the structure tensors as follows:

$$\begin{aligned} SED(T_1(u), T_2(u')) = &\left(T_{1xx}(u) - T_{2xx}(u')\right)^2 \\ &+ \left(T_{1yy}(u) - T_{2yy}(u')\right)^2 \\ &+ 2\left(T_{1xy}(u) - T_{2xy}(u')\right)^2 \end{aligned} \tag{5}$$

The O_rC_h estimates the dissimilarity between structure tensors by means of geometric properties [35], i.e. the orientation and coherence factors of the tensors:

$$\begin{aligned} O_rC_h(T_1(u), T_2(u')) = &\lambda_O(T_1(u), T_2(u'))^{\beta} \\ &\cdot \lambda_C(T_1(u), T_2(u'))^{1-\beta}, \end{aligned} \tag{6}$$

Where β is a weight ranging between 0 and 1, λ_O is the dissimilarity measure of orientation and λ_C is the dissimilarity measure of coherence respectively expressed as:

$$\begin{aligned} \lambda_O = &\left|\sin\left(O_r(T_1(u)) - O_r(T_2(u'))\right)\right| \\ &\cdot \min\left(C_h(T_1(u)), C_h(T_2(u'))\right), \end{aligned} \tag{7}$$

And

$$\lambda_C = \max\left(\frac{C_h(T_1(u))}{C_h(T_2(u'))}, \frac{C_h(T_2(u'))}{C_h(T_1(u))}\right) \tag{8}$$

The FN metric [35] is given by:

$$FN(T_1(z), T_2(z')) = \sqrt{Tr(\alpha'\alpha)}\ , \tag{9}$$

Where $\alpha = \left[T_1(u) - T_2(u')\right]$ and Tr is the trace operator.

3. Exploitation of tensor-space metrics

This section deals with exploiting the tensor-space similarity metrics in image inpainting, in motion estimation, and in the analysis of medical images.

The method consists in combining the intensity information of the image with its structural information calculated using the structure tensor field. More precisely, in order to measure the similarity between two image neighborhoods (or patches) $PY_1(u)$ and $PY_2(u')$, respectively having u and u' as center pixel positions, we linearly combine the similarity measures as follows:

$$\begin{aligned} S(u,u') = &k \cdot SSD\left(PY_1(u), PY_2(u')\right) \\ &+ (1-k) \cdot SSTD\left(PT_1(u), PT_2(u')\right), \end{aligned} \tag{10}$$

Where k is a weighting factor ($0 \le k \le 1$), SSD is the Sum of Squared Distances calculated between the intensity values of $PY_1(u)$ and $PY_2(u)$, and SSTD($PT_1(u)$, $PT_2(u')$) is the Sum of Structure Tensor Dissimilarities calculated between the structure tensor fields $PT_1(u)$ and $PT_2(u')$ of patches $PY_1(u)$ and $PY_2(u')$,

respectively. The measures SSD and SSTD can be respectively expressed as:

$$SSD\left(PY_1(u), PY_2(u')\right) = \sum_{n=0}^{N-1}\left(PY_1(n) - PY_2(n)\right)^2 , \quad (11)$$

And

$$SSTD\left(PT_1(u), PT_2(u')\right) = \sum_{n=0}^{N-1}\Delta\left(PT_1(n), PT_2(n)\right), \quad (12)$$

Where $PY_i(n)$ is the n^{th} intensity value within image patch $PY_i(z)$ ($i \in \{1,2\}$, $z \in \{u,u'\}$), N is the number of pixels in each patch, and Δ is one of the three tensor-space similarity metrics ($\Delta \in \{SED, O_r C_h, FN\}$) presented in section 2.2.

In other words, each neighborhood has two components; an intensity neighborhood in the image patch itself, and a structure tensor neighborhood belonging to the structure tensor field of this image patch. In this paper, square neighborhoods of size N are used. Note that N can be as large as the number of pixels in the image. In this case, the similarity is globally calculated between two images of size N, as will be shown later.

It is important to mention that by setting k to 1, the structural information is deactivated and the similarity measurement is only based on the image intensities. On the contrary, intensity information can be deactivated by setting k to 0. This results in taking into consideration only the tensor-space information in the image similarity estimation process.

3.1. Tensor-based image inpainting

Image inpainting is an active research field used in several applications, including video animations, video compression and image editing [36-39]. It is also used for the removal of undesired objects and for filling missing areas in images [40]. Two image inpainting examples are illustrated in Fig. 2. For concision, we will denote the area that we want to inpaint (i.e. the missing area) as 'AI', the reference patch from which the grey-levels (or colors) are copied to the AI as 'RF', the image containing the AI and the RF as 'input sample' and the output image obtained after inpainting the AI as 'output image'.

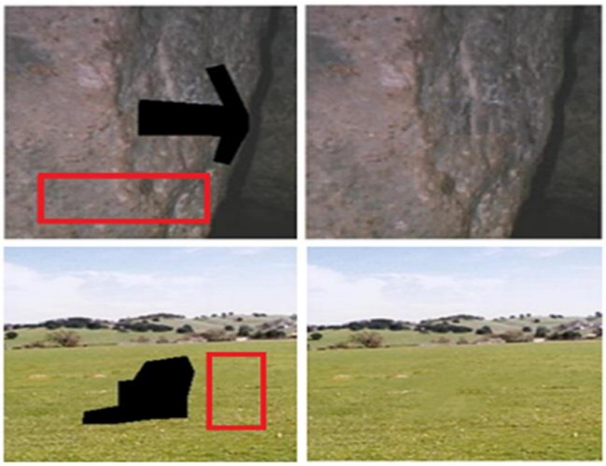

Fig. 2: Examples of Image Inpainting. Left Column: Input Images Showing the AI (Area to Inpaint) Marked in Black and the RF (Reference Patch) Contoured in Red. Right Column: Output Images.

Several image inpainting algorithms have been proposed [40-43]. Most of these methods were successful in different cases. However, failures still exist, especially with high local structural variations. Therefore, the use of the structure layer of the image represented by the structure tensor field can be exploited in order to enhance the inpainting process of the standard intensity-based methods.

The proposed method consists in adapting an intensity-based algorithm to the structure tensor field specificities. Consider for example the method of Wei and Levoy [44] due to its flexibility and to its pixel-based principle which allows the inpainting of a wide variety of images (other inpainting methods could have also been considered, without loss of generality). The adaptation consists in inpainting the structure layer of the input sample in a first stage, then using the inpainted structure layer in order to constrain the image inpainting, in a second stage.

3.1.1. Structure layer inpainting

The structure layer inpainting starts by the calculation of the structure tensor field from the input sample luminance component (cf. section 2.1). The luminance component is calculated as in [45]. The inpainting process ensures that the inpainted area is locally similar to the neighboring regions of the input sample by adapting the algorithm of Wei and Levoy [44] to the specificities of the structure tensor as follows: the AI of the tensor field is first initialized by a random noise, i.e. tensors chosen randomly from the tensor field of the RF, then the neighborhood (vector of matrices) of each structure tensor of the AI is captured, the neighborhood of the tensor field of the RF having the best similarity with the current neighborhood is found, and its central tensor is copied to the current position in the AI, as illustrated in Fig. 3. This can be iteratively repeated in order to obtain a coherent inpainted area with neighboring tensors. The similarity between two structure tensor neighborhoods is calculated using the Sum of Structure Tensor Dissimilarities (cf. equation 12). Note that the inpanting process starts with the outer borders of the AI and ends at its center.

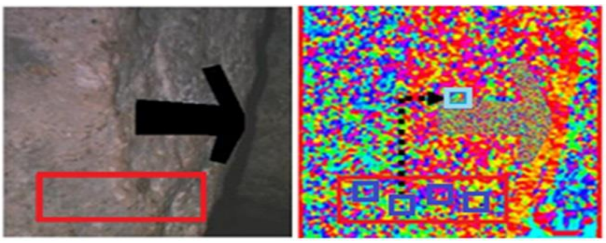

Fig. 3: Illustration of the Structure Tensor Field Inpainting Process. Left: Input Sample. Right: It's Structure Tensor Field Orientation Image with Its AI Initialized by A Random Noise. The Most Similar Neighborhood (Blue Square) of the Current Neighborhood (Cyan Square) Is Searched for in the RF (Contoured In Red), and the Corresponding Tensor Is Copied to the Target Position in the AI.

3.1.2. Structure-constrained image inpainting

In the image inpainting stage, the inpainted structure layer is used to constrain the inpainting of the output image. The method of Wei and Levoy is also used in this inpainting process. However, the neighborhoods take also into consideration the information held by the inpainted structure tensor field. More precisely, the inpainting algorithm takes as inputs, the RF, its tensor field, the inpainted tensor field and the AI initialized by random noise. The AI pixels are then iteratively updated in order to ensure their local similarity with neighboring pixels of the input sample: the neighborhood of every pixel of the AI is captured, the most similar neighborhood is searched for in the RF and copied entirely to the target position in the AI. However, the underlying neighborhoods have two components: an intensity component in the AI ($PY_1(u)$) and in the RF ($PY_2(u')$) of the output image, and a structure tensor component in the inpainted tensor field ($PT_1(u)$) and in the tensor field of the RF ($PT_2(u')$) as shown in Fig. 4. The similarity between two neighborhoods located at sites u and u' is therefore computed using the S(u,u') metric (cf. equation 10).

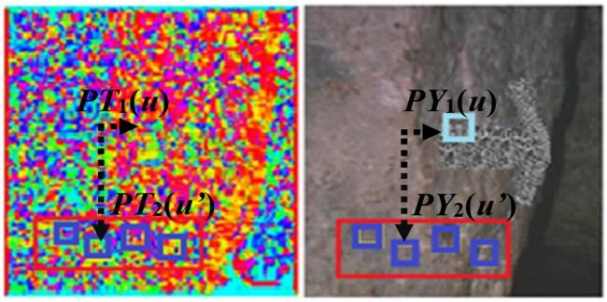

Fig. 4: Illustration of the Structure-Constrained Image Inpainting Process. for Each Current Neighborhood ($PY_1(U)$ in AI and $PT_1(U)$ in Its Inpainted Tensor Field), The Most Similar Neighborhood ($PY_2(U')$ in RF and $PT_2(U')$ in Its Tensor Field) Is Found, and the Corresponding Intensity Component ($PY_2(U')$) Is Entirely Copied to the Target Position ($PY_1(U)$).

The pseudocode of the whole inpainting process is given in Fig. 5. Note that, in order to reduce the computational load, the update of AI pixels is patch-based (i.e. the most similar neighborhood is entirely copied to the target position) and not point by point as is the case of structure layer inpainting. In fact, this can be achieved with merely no blockiness or artifacts, thanks to the local structures information held by the inpainted structure tensor field.

Structure layer inpainting

$T \leftarrow$ STRUCTURETENSORFIELDCALCULATION (Y)

$AI^T \leftarrow$ STRUCTURETENSORNOISEINITIALIZATION (AI^T)

loop through all positions u of AI^T

  $u' \leftarrow$ argmax { BESTSIMILARITY($PT_1(u)$ vs $PT_2(u')$)}

  $PT_1(u) \leftarrow PT_2(u')$

endloop

Structure-constrained image inpainting

$AI \leftarrow$ NOISEINITIALIZATION (AI)

loop through all positions u of AI

  $u' \leftarrow$ argmax{BESTSIMILARITY($PY_1(u)$, $PT_1(u)$ vs $PY_2(u')$, $PT_2(u')$)}

  $PY_1(u) \leftarrow PY_2(u')$

endloop

Fig. 5: Pseudocode of the Structure-Based Image Inpainting Algorithm.

3.2. Tensor-based motion estimation

Video compression consists in digital video format conversion by redundancy removal, i.e. components that are not necessary for the reproduction of the original data [47].

It has been shown that lossless compression standards such as JPEG-LS [46] cannot achieve a compression ratio greater than 4:1. Lossy techniques are therefore used to obtain a better compression. This relies in the removal of subjective redundancy, i.e. the frame data which is able to be removed without disturbing the visual quality [5], [47]. Most of the compression methods apply both spatial and temporal removal of redundancy. The removal of spatial redundancy relies on the closeness of adjacent samples, while the removal of temporal redundancy is based on the fact that successive frames are normally highly correlated [5].

Most compression standards share the same common principles. For instance, MPEG-2 Video [48], MPEG-4 Visual [49], H.263 [50], H.264 Advanced Video Coding [47], and H.265 High Efficiency Video Coding (HEVC) [51] reduce temporal redundancy using block-based motion estimation.

The block-based motion estimation consists in compensating for movement of rectangular blocks of the current frame: each neighborhood (block) of size M×N of the current frame is compared with all the neighborhoods of the same size in a P×L area of one or more reference (past or future) frames, i.e. area centered on the collocated position of the current neighborhood in the reference frame, and the relative position (i.e. motion vector) of the best matching neighborhood is then determined [47]. Block similarity is usually measured using the Sum of Squared Distances (cf. equation 11).

This widely used approach takes only into consideration the intensity information of the underlying frames. Therefore, jointly using the structure layer and the intensity information in order to find the best match is proposed. Note that in our approach, we only focus on motion estimation (temporal correlation).

The proposed approach starts by applying motion estimation in the tensor domain. More precisely, it estimates the structural motion of the frame. It then constrains motion estimation of the intensity layer using the obtained structure layer. In both stages, the block-based approach is adapted to the specificities of the structure tensor.

The structure layer motion estimation computes the reference tensor field (i.e. the reference frame structure tensor field) and the current tensor field (i.e. the current frame structure tensor field). Then it extracts the neighborhood of size M×N of each current frame tensor, compares its tensor values with all the neighborhoods of the same size in a P×L search area of the reference tensor field. The neighborhood having the best match is determined, and its central value is copied to the current tensor field target position. The similarity between two structure tensor neighborhoods is calculated using the Sum of Structure Tensor Dissimilarities (cf. equation 12).

The obtained motion-estimated structure tensor field is then used as a constraint for the motion estimation process of the frame's pixel layer. The inputs of the algorithm are: the reference frame with its structure tensor field, the current frame, and the motion-estimated tensor field. The motion estimation principle is the same as the standard block-based algorithm: for each neighborhood of size M×N in the current frame, search in a P×L area of the reference frame for the neighborhood leading to the best similarity. However, the neighborhood structure takes the motion-estimated structure tensor field into consideration. Therefore, the neighborhood has two components: a pixel-domain neighborhood in the reference and current frames, and a structure tensor neighborhood in the reference and motion-estimated tensor fields. The metric used to measure the resemblance between two neighborhoods, combining the Sum of Squared Distances (SSD) for the pixel-domain neighborhood components and the Sum of Structure Tensor Dissimilarities (SSTD) for the structure tensor neighborhood components, is that of (10). The two-stage structure-based motion estimation approach is illustrated in Fig. 6.

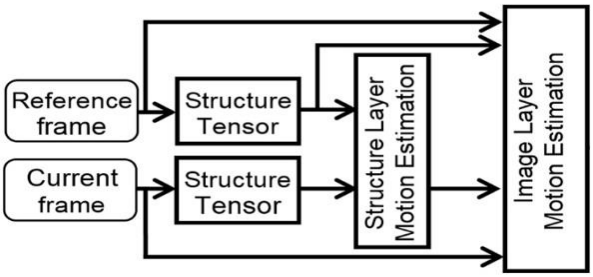

Fig. 6: Illustration of the Structure-Based Motion Estimation Approach.

Note that jointly applying motion search in both the structure and the image layers is also a possible solution. This is advantageous with respect to the two-stage method in terms of complexity, which increases naturally in this latter (two-stage) case because of the computational load associated with the reconstruction of the frame's structure layer.

3.3. Tensor-based analysis of medical images

Image analysis is a wide area of study used in different image processing applications. For instance, it is used in medical imag-

ing for brain tumor detection [6-9]. Different image analysis methods have been proposed in the literature [10], [22 - 31]. Most of these methods rely on computing the similarity between images or image patches. However, the underlying structure is not taken into consideration.
The proposed similarity computation approach simply combines intensity and structure tensor similarity values using the metric given in (10), which allows to better detect local structural dissimilarities with different types of images, as will be shown in section 4.

4. Results

This section presents the results obtained using different examples in inpainting, motion estimation, and medical image analysis.

4.1. Image inpainting results

4.1.1. Influence of the structure/intensity weighting

The structure/intensity weighting factor (k) used in (10) directly influences the quality of the output images.
Fig. 7 presents inpainted images using different values of this weighting factor with the SED metric; from left to right: the input image showing the AI (in black) and the RF patch (highlighted in red), and the obtained output images using k = 0.15, 0.5 and 0.85. It can be seen that a low weighting factor of 0.15 results in discontinuities at the borders of the AI. In this case, the intensity information is negligible with respect to the structural information. However, the structures and the dynamics of the obtained AI are acceptable. On the contrary, a high weighting factor of 0.85 leads to missing important structures of the reference image. Finally, it can be clearly observed that the inpainted AI obtained using a weighting factor of 0.5 better preserves the dynamics and the structural information.

Fig. 7: Results of the Structure-Based Inpainting Approach with Different Values for the Structure/Intensity Weighting Factor. Left to Right: the Input Image, the Output Images Obtained Using K = 0.15, 0.5 and 0.85.

4.1.2. Qualitative and quantitative evaluations

In this section, the proposed method and the algorithm of Wei and Levoy [44] are compared.
Fig. 8 presents results obtained on four different input images. Each result shows from upper-left to bottom-right, the input image showing the AI (in black) and the RF (highlighted in red), the orientation image of its structure tensor field, the inpainted structure tensor field, the output images obtained with a weighting factor k = 1 (i.e. Wei and Levoy's algorithm) and with the proposed algorithm using k = 0.5, and a closer view of a particular region of interest in output images obtained with k = 1 and k = 0.5. The SED metric is used in results A and B while the O_rC_h and FN metrics are used in results C and D respectively. In result A obtained on a regular structured image, the proposed approach succeeds in leading to a well inpainted structure tensor field which seems artefacts-free. Therefore, the obtained output image is of good quality and without edge effects or distortions, while the AI of the output image obtained while deactivating the structural information presents artefacts and undesired patterns. The same applies to result B where the proposed approach is better than Wei and Levoy algorithm in terms of preserving the local structures. The AI of the image obtained with Wei and Levoy presents over-smoothing while the one obtained by the proposed approach shows more structural details that are similar to those of the RF. In result C, the AI obtained using k = 1 seems more regular than the exemplar. More precisely, unwanted periodic patterns that do not exist in the exemplar are clearly present. In result D, the dynamics and structures of the output images obtained by both algorithms are almost similar. However, despite some discontinuities (in both results) that do not significantly affect the image, the result with k = 0.5 is sharper and better preserves stairs structure compared to the case with k = 1.
Note that we limited our comparison to the algorithm of Wei and Levoy [44] without loss of generality, since we aim at emphasizing on intensity-based similarity measures that can be enhanced by taking into consideration the structural information.
Besides the subjective qualitative analysis, a quantitative evaluation is also performed. It consists in comparing intensity and orientation histograms of the exemplar and the inpainted area by calculating the Kullback-Leibler divergence [52] between them as follows:

$$KL(H_{AI} \| H_{RF}) = \sum_i H_{AI}(i) \log \frac{H_{AI}(i)}{H_{RF}(i)} + \sum_i H_{RF}(i) \log \frac{H_{RF}(i)}{H_{AI}(i)}, \quad (13)$$

Where H_{AI} and H_{RF} are respectively the histograms of intensity (or orientation) of the inpainted area and the reference patch.
Table 1 presents the difference values of Kullback and Leibler calculated on the histograms of the output images of Fig. 8. In general, the subjective evaluation is confirmed. In results A and B, the differences of both histograms (intensity and orientation) are higher with k = 1 than with k = 0.5, which verifies that the proposed method outperforms Wei and Levoy that results in artefacts and dynamics distortions of the AI. The repetitive patterns present in the AI obtained using k = 1 result in a high difference of the orientation histogram. Finally, the last row of the table verifies that images obtained by the proposed approach and by the method of Wei and Levoy are of nearly same quality.

Table 1: Kullback and Leibler Difference Values Obtained on the Output Images of Fig. 8

Result	Intensity Histograms		Orientation Histograms	
	k = 1	k = 0.5	k = 1	k = 0.5
A	0.398	0.28	0.379	0.278
B	0.412	0.201	0.399	0.297
C	0.308	0.31	0.402	0.289
D	0.398	0.356	0.3	0.306

4.2. Motion estimation results

In this section, a comparison between the proposed tensor-based motion estimation method and the original block-based algorithm is presented and evaluated using different standard video sequences [53].
Fig. 9 shows, from upper-left to bottom-right, the 10th frame of a video sequence used as a reference frame, the 13th frame of the

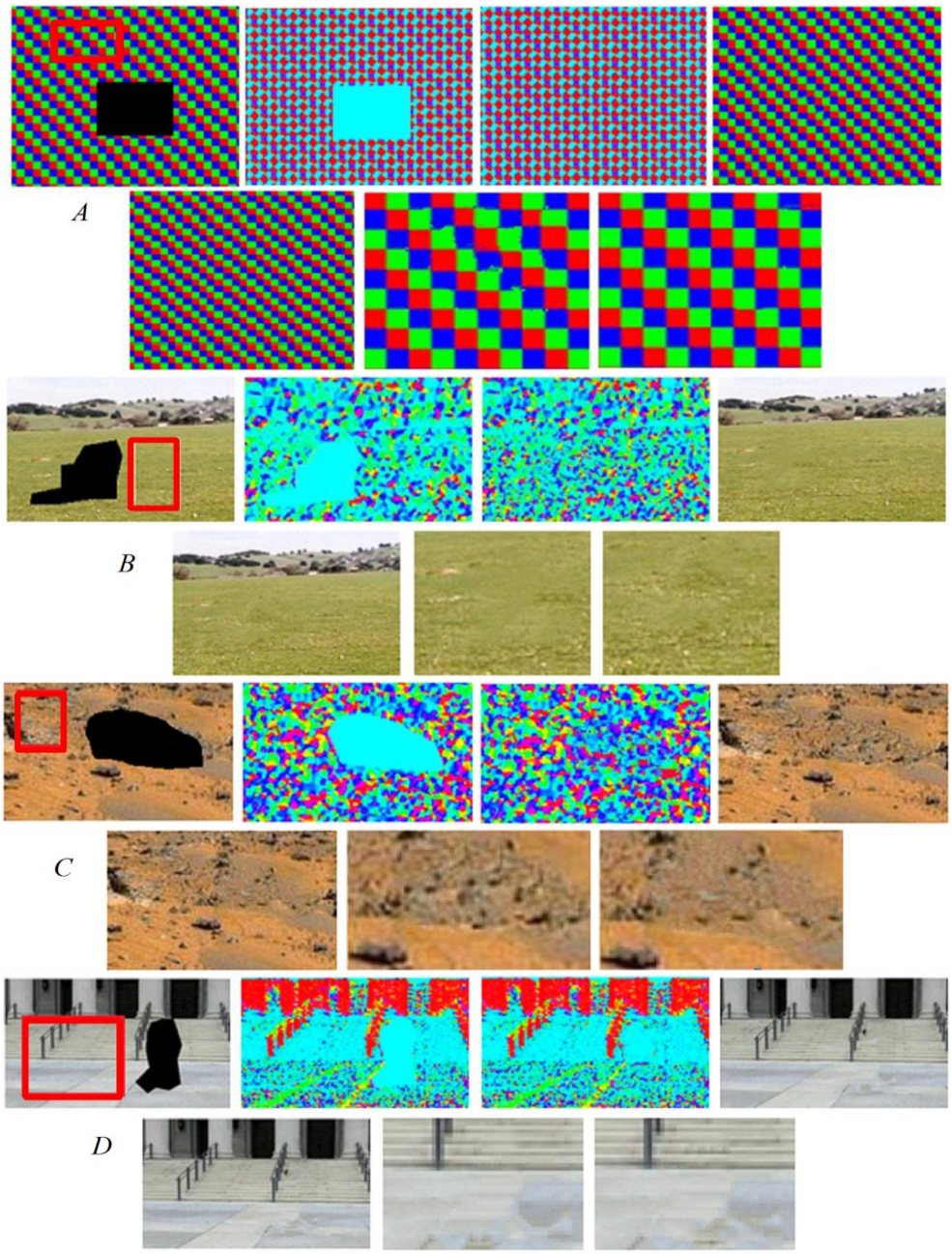

Fig. 8: Inpainting Results. Each Result (Upper-Left to Bottom-Right): Input Image, Its Tensor Field Represented by Its Orientation Image, Inpainted Tensor Field, Output Images Obtained with K = 1 and K = 0.5, Nearer View of the Output Images Obtained with K = 1 and K = 0.5.

Same video sequence used as a current frame, the orientation image of the structure tensor field calculated on the reference frame, the orientation image of the motion-estimated structure tensor field obtained with the tensor-based approach, the reconstructed frames using the standard block-based algorithm and the proposed method using the SED metric. It can be clearly seen that the proposed approach succeeds in well reconstructing the structure tensor field. The resulting orientation images are almost similar to those of the reference image. The reconstructed frame obtained using the proposed algorithm resembles the original frame while the one obtained using the standard method shows artifacts and unwanted patterns (highlighted in red).

Results obtained on samples taken from three different video sequences are present in Fig. 10. Each result shows, from first to fourth row: the reference frame, the current frame, the obtained frames using the standard algorithm and using the proposed approach. For all the results, the current frame is the 12^{th} video frame and the reference frame is the 10^{th}. The SED metric is used in result E while the O_rC_h metric is used in result F and the FN metric is used in result G. Note that the structure tensor and frame motion estimation stages are simultaneously applied in result G.

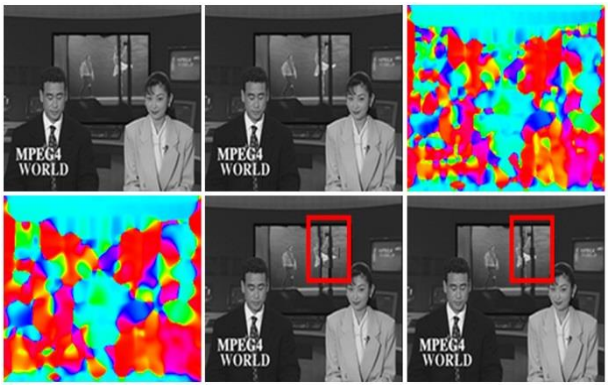

Fig. 9: Results of Motion Estimation. Upper-Left to Bottom-Right: Reference Frame, Current Frame, Orientation Image of the Reference Frame Tensor Field, Orientation Image of the Motion-Estimated Tensor Field, Motion-Estimated Frames Using the Standard and the Tensor-Based Approaches.

In result E, the standard approach results in a frame presenting artifacts (distorted letters highlighted in red) while the proposed

method leads to a reconstructed frame well reproducing the letters. A close view in result F shows that the proposed approach better preserves edge information. The same applies for result G where the frame obtained using the intensity-based approach presents defects (highlighted in red), while the one reconstructed using the proposed structure-based algorithm is of better quality. Note that the improved quality of the reconstructed frames leads to smaller residual errors, which allows for a better video compression. Furthermore, improving the quality of a video frame does not only affect that frame, but rather a sequence of frames that are decoded afterwards due to interframe dependence. This approach is thus expected to improve the overall rate-distortion performance of a given video codec.

Besides the subjective evaluation, an objective analysis is also presented. The Peak Signal-to-Noise Ratio (PSNR) is computed between the original frame and the reconstructed one. The PSNR values obtained on the results of Fig. 11 are shown in Table 2. As expected, PSNR values obtained with the proposed approach are slightly higher than those obtained with the ordinary block-based algorithm. However, this increase in PSNR values is almost negligible; this is due to the fact that the PSNR is a similarity measure averaged over the whole global frame, whereas the artifacts of the motion-estimated frame, where the improvement can be observed, are present in local regions.

Table 2: PSNR of the Results of Fig. 10

Result	PSNR (dB) Ordinary block-based algorithm	Structure Tensor-based algorithm
E	25.81	25.87
F	34.9	35.23
G	31.77	32.1

4.3. Analysis of medical images

In this section, the proposed image similarity estimation approach is evaluated on medical images.

Results H and I of Fig. 11 show, from upper-left to bottom-right, an initial brain MRI image, a modified version of this image obtained after smoothing the region highlighted in red with a Gaussian averaging filter, the image of differences between the initial and the modified MRI image obtained with k = 1 and k = 0.5 (cf. equation 10) using the SED, O_rC_h and FN metrics. Smoothing is applied to model some types of brain tumors (as in result J where a real tumor is shown) because we could not obtain MRI images of the same patient before and after the tumor infected the brain; this is applied only for demonstration purpose, which is not the case with result J where a different approach was applied.

It can be seen in result H that the image of differences obtained with the intensity-based comparison (i.e. the structure tensor information is deactivated, k = 1) succeeds in detecting the dissimilarities between the initial and the modified image. The images of differences obtained with the SED and FN metrics are worst in terms of dissimilarities detection. However, the O_rC_h metric outperforms the other tensor-space metrics and the intensity-based comparison in detecting the region of differences. The same analysis is applicable for result I where the smoothed region (highlighted in red) is of smaller size than that of result H.

For result J, approximate left-right symmetry of the brain principle [6, 9] is used. More precisely, the initial MRI image is vertically divided in its middle to two equal halves. Then the obtained left side is correlated with the obtained right side which is flipped horizontally, in order to align the two side images. Result J presents, from left to right, the initial brain MRI image showing a tumor on its bottom-left side, the obtained right and left sides of the initial image, the image of differences between the two side images obtained with k = 1 and k = 0.5 (cf. equation 10) using the SED, O_rC_h and FN metrics. It can be seen that the intensity-based dissimilarity calculation barely detects the differences between the two side images, and therefore fails in clearly showing the pattern of the tumor. On the contrary, due to the additional structural information provided by the structure tensor field, the three images of differences obtained with the proposed approach better show the dissimilarities between the side images and therefore the tumor pattern.

5. Conclusion

This paper discussed a structure-based image similarity estimation approach where the structure layer of the images is represented by the structure tensor field. The presented method relies on combining the intensity dissimilarity values with the structural dissimilarity values computed using three different tensor-space metrics. Its use is investigated in three dissimilarity estimation contexts: image inpainting, motion estimation and analysis of medical images.

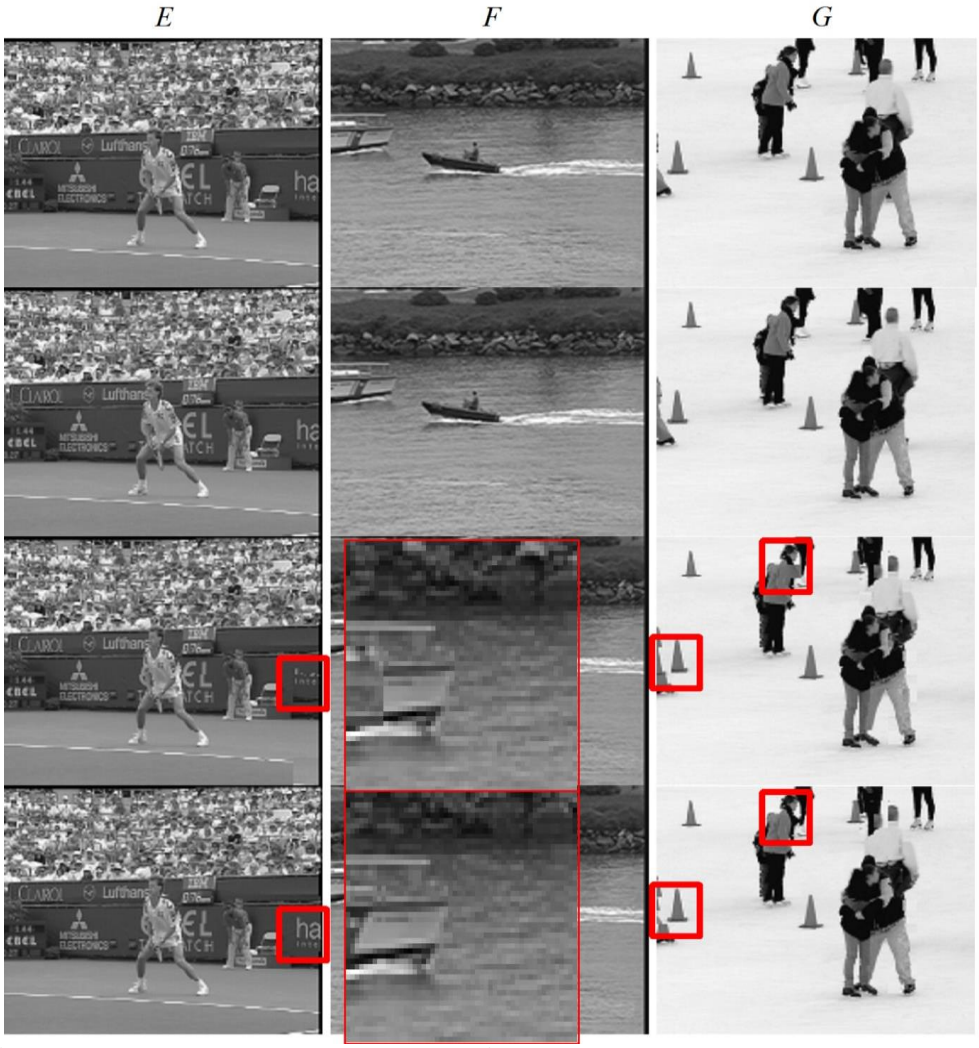

Fig. 10: Results of Motion Estimation. Each Result (1st to 4th Row): Reference Frame, Current Frame, Motion-Estimated Frame Using the Standard Approach, and Motion-Estimated Frame Using the Tensor-Based Algorithm.

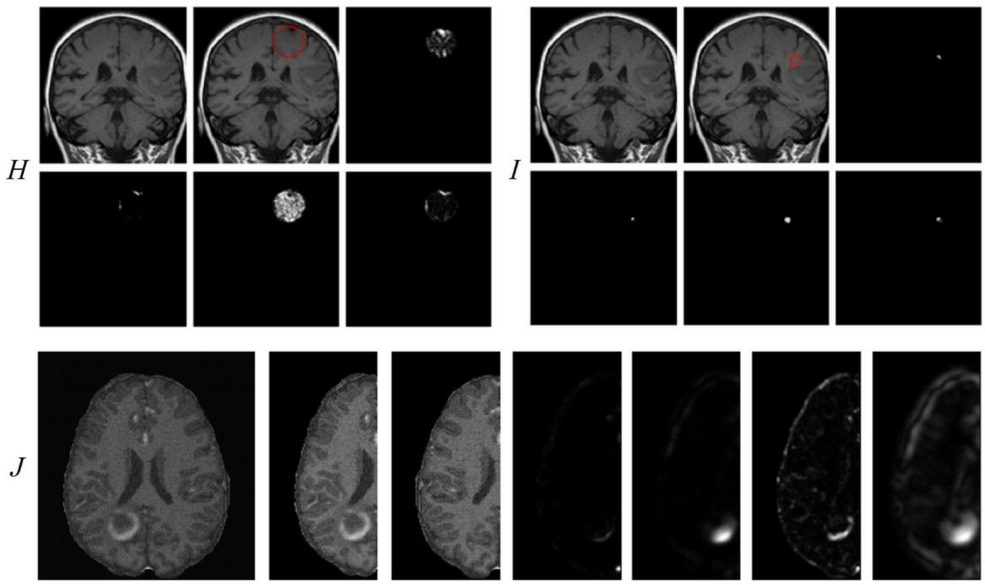

Fig. 11: Results of Similarity Estimation. H and I (Upper-Left to Bottom-Right): Brain MRI Image, A Modified Version of this Image Obtained After Smoothing the Region Highlighted in Red, the Image of Differences between the Initial and the Modified Image Obtained with K = 1 and K = 0.5 (Cf. Equation (10)) Using the SED, O_rC_h and FN Metrics. J (Left to Right): Brain MRI Image, the Right and Left Sides of the Initial Image, the Image of Differences between the Two Side Images Obtained with K = 1 and K = 0.5 (Cf. Equation (10)) Using the SED, O_rC_h and FN Metrics.

Applied on different images, structure-based similarity measures show that they are advantageous in better detecting the local differences between the underlying images, especially in the case of local structural variations where the intensity-based metric fails in detecting dissimilarities.

As for future perspectives, we aim at using parallel processing algorithms that can make use of GPU capabilities in order to re-

duce the computational load of the point by point difference calculation.

Acknowledgement

This project has been jointly funded with the support of the National Council for Scientific Research in Lebanon CNRS-L and the Holy Spirit University of Kaslik.

References

[1] A. Akl, C. Yaacoub, M. Donias, J.P. Da Costa, C. Germain, "*Structure tensor-based synthesis of directional textures for virtual material design*", *Proceedings of the 21st IEEE International Conference on Image Processing (ICIP)*, (2014).

[2] A. Akl, C. Yaacoub, M. Donias, J.P. Da Costa, C. Germain, "Texture synthesis using the structure tensor", *IEEE Transactions on Image Processing*, Vol. 24, No. 11, (2015), pp. 4082-4095. https://doi.org/10.1109/TIP.2015.2458701.

[3] G. Peyré, "Texture Synthesis with grouplets", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 4, (2009), pp:733-746. https://doi.org/10.1109/TPAMI.2009.54.

[4] A. Akl, E. Saad, C. Yaacoub, "Structure-Based Image Inpainting", *Proceedings of the 6th International Conference on Image Processing Theory, Tools and Applications*, (2016). https://doi.org/10.1109/IPTA.2016.7820976.

[5] A. Akl, R. Gemayel, N. Alkhoury, C. Yaacoub, *"*Structure-Based Motion Estimation for Video Compression*", Proceedings of the International Multidisciplinary Conference on Engineering Technology*, (2016). https://doi.org/10.1109/IMCET.2016.7777420.

[6] S. Ghanavati, T. Liu, P.S. Babyn, W. Doda, G. Lampropoulos, "Automatic brain tumor detection in magnetic resonance images", *Proceedings of the 9th IEEE International Symposium on Biomedical Imaging*, (2012).

[7] T. Sugimoto, S. Katsuragawa, T. Hirai, R. Murakami, Y. Yamashita, "Computerized detection of metastatic brain tumors on contrast-enhanced 3D MR images by using a selective enhancement fil*ter", Proceedings of the 2010 World Congress on Medical Physics and Biomedical Engineering*, (2010).

[8] R. Ambrosini, P. Wang, "Computer-aided detection of metastatic brain tumors using automated three-dimensional template matching", *Journal of MRI*, Vol. 31, No. 1, (2010), pp: 85-93. https://doi.org/10.1002/jmri.22009.

[9] N. Ray, B.N. Saha, M. Brown, "Locating brain tumors from MR imagery using symmetry", *Proceedings of the 41st Asilomar Conference on Signals, Systems and Computer*s, (2007). https://doi.org/10.1109/ACSSC.2007.4487200.

[10] S. Sevestre-Ghalila, A. Benazza-Benyahia, A. Ricordeau, N. Mellouli, C. Chappard, C. Benhamou, "Texture image analysis for osteoporosis detection with morphological tools", *Barillot C, Haynor D R, Hellier P (eds) Medical Image Computing and Computer-Assisted Intervention. Lecture Notes in Computer Science, Springer*, (2004), pp: 87-94. https://doi.org/10.1007/978-3-540-30135-6_11.

[11] D.W. Dempster, "The contribution of trabecular architecture to cancellous bone quality", *J Bone Miner Res*, Vol. 15, No. 1, (2000), pp: 20-23. https://doi.org/10.1359/jbmr.2000.15.1.20.

[12] A.S. Hassani, M. Hassouni, R. Jennane, M. Rziza, E. Lespessailles, "Texture analysis for trabecular bone X-ray images using anisotropic Morlet wavelet and Rényi entropy", *Proceedings of the International Conference on Image and Signal Processing*, (2012). https://doi.org/10.1007/978-3-642-31254-0_33.

[13] C. Zhu, "Remote sensing image texture analysis and classification with wavelet transform", *International Archives of Photogrammetry and Remote Sensing*, Vol. 19, No. 16, (1996).

[14] K. Wikantika, A. Harto, R. Tateishi, "The use of spectral and textural features from Landsat TM image for land cover classification in mountainous area", *Proceedings of the 2001 IECL Japan workshop*, (2001).

[15] A. C. Beers, M. Agrawala, N. Chaddha, "Rendering from compressed textures", *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, ACM*, (1996), pp: 373-378.

[16] W. Sun, Y. Lu, F. Wu, S. Li, J. Tardif, "High-Dynamic-Range Texture Compression for Rendering Systems of Different Capacities", *IEEE Trans. on Visualization and Computer Graphics*, Vol. 16, No. 1, (2010), pp: 57-69. https://doi.org/10.1109/TVCG.2009.60.

[17] J.M. Leyssale, J.-P. Da Costa, C. Germain, P. Weisbecker, G. Vignoles, "An image-guided atomistic reconstruction of pyrolytic carbons", *Applied Physics Letters*, Vol. 95, No. 23, (2009). https://doi.org/10.1063/1.3272949.

[18] C. Chapoullie, "Analyse/synthese tridimensionnelle de textures fibreuses ", *PhD dissertation, University of Bordeaux*, (2014).

[19] A. Akl, J. Iskandar, *"*Structure tensor regularization for texture analysis*", Proceedings of the 5th International Conference on Image Processing Theory, Tools and Applications*, (2015). https://doi.org/10.1109/IPTA.2015.7367217.

[20] R. Paget, I. Longstaff, "Texture synthesis via a noncausal nonparametric multiscale Markov random field", *IEEE Trans. on Image Processing*, Vol. 7, No. 6, (1998), pp: 925-931. https://doi.org/10.1109/83.679446.

[21] J. Portilla, E.P. Simoncelli, "A Parametric Texture Model based on Joint Statistics of Complex Wavelet Coefficients", *International Journal of Computer Vision*, Vol. 40, No. 1, (2000), pp: 49-71. https://doi.org/10.1023/A:1026553619983.

[22] R. Pless, I. Simon I, "Using thousands of images of an object", *Proceedings of the 6th Joint Conference on Information Science, (CVPRIP)*, (2002).

[23] S. Bhandarkar, F. Chen, "Similarity Analysis of Video Sequences Using an Artificial Neural Network", *Applied Intelligence*, Vol. 22, No. 3, (2005), pp: 251-275. https://doi.org/10.1007/s10791-005-6622-3.

[24] A. Akl, J. Iskandar, *"*Second-moment matrix adaptation for local orientation estimation*", Proceedings of the 23rd International Conference on Systems, Signals and Image Processing*, (2016). https://doi.org/10.1109/IWSSIP.2016.7502721.

[25] J. P. Antoine, P. Carrette, R, Murenzi, B. Piette, "Image analysis with two-dimensional continuous wavelet transform", *Signal Processing*, Vol. 31, No. 3, (1993), pp: 241-272. https://doi.org/10.1016/0165-1684(93)90085-O.

[26] W. Tang, Y. Wang Y, W. He, "An image segmentation algorithm based on improved multiscale random field model in wavelet domain", *Journal of Ambient Intelligence and Humanized Computing*, Vol. 7, No. 2, (2016), pp: 221-228. https://doi.org/10.1007/s12652-015-0318-3.

[27] Z. Wu, J. Yuan, J. Zhang, H. Huang, "A hierarchical face recognition algorithm based on humanoid nonlinear least-squares computation", *Journal of Ambient Intelligence and Humanized Computing*, Vol. 7, No. 2, (2016), pp: 229-238. https://doi.org/10.1007/s12652-015-0321-8.

[28] P.A. Sahoo, "Thresholding method based on two-dimensional renyis entropy", *Pattern Recognition*, Vol. 37, No. 6, (2004), pp: 1149-1161. https://doi.org/10.1016/j.patcog.2003.10.008.

[29] L. Houam, A. Hafiane, R. Jennane, A. Boukrouche, E. Lespessailles, "Trabec*ular bone anisotropy characterization using 1d local binary patterns", Blanc-Talon J, Bone D, Philips W, Popescu D, Scheunders P (eds) Advanced Concepts for Intelligent Vision Systems. ACIVS 2010. Lecture Notes in Computer Science, Springer*, (2010), pp: 105-113. https://doi.org/10.1007/978-3-642-17688-3_11.

[30] C. L. Benhamou, S. Poupon, E. Lespessailles, S. Loiseau, R. Jennane, V. Siroux, W. Ohley, L. Pothuaud, "Fractal analysis of radiographic trabecular bone texture and bone mineral density: two complementary parameters related to osteoporotic fractures", *J Bone Miner Res*, Vol. 16, No. 4. (2001), pp:. 697-704. https://doi.org/10.1359/jbmr.2001.16.4.697

[31] R. Jennane, W.J. Ohley, S. Majumdar, G. Lemineur, "Fractal analysis of bone x-ray tomographic microscopy projections", *IEEE Trans. Med. Imaging*, Vol. 20, No. 5, (2001), pp: 443-449. https://doi.org/10.1109/42.925297.

[32] L. Pothuaud, E. Lespessailles, R. Harba, R. Jennane, V. Royant, E. Eynard, C.L. Benhamou, "Fractal Analysis of Trabecular Bone Texture on Radiographs: Discriminant Value in Postmenopausal Osteoporosis", *Osteoporosis International*, Vol. 8, No. 6, (1998), pp: 618-625. https://doi.org/10.1007/s001980050108.

[33] L. Dryden, A. Koloydenko, D. Zhou, "Non-Euclidean Statistics for Covariance Matrices, with Applications to Diffusion Tensor Imaging", *Annals of Applied Statistics*, Vol. 3, No. 3, (2009), pp: 1102-1123. https://doi.org/10.1214/09-AOAS249.

[34] J. Angulo, "Structure Tensor Image Filtering using Riemannian L1 and L∞ Center-of-mass", *Image Analysis and Stereology*, Vol. 33, No. 2, (2014), pp: 95-105. https://doi.org/10.5566/ias.v33.p95-105.

[35] V. Toujas, M. Donias, Y. Berthoumieu, "Structure Tensor Field Regularization Based on Geometric Features", *Proceedings of the 18th European Signal Processing Conference (EUSIPCO)*, (2010).

[36] P. Fillard, V. Arsigny, N. Ayache, X. Pennec, "A Riemannian framework for the processing of tensor-valued images", *Fogh Ol-*

sen O, Florack L, Kuijper A (eds) Deep Structure, Singularities, and Computer Vision. Lecture Notes in Computer Science, Springer, (2005), pp: 112–123. https://doi.org/10.1007/11577812_10.
[37] V. Kwatra, A. Schödl, I.A. Essa, G. Turk, A.F. Bobick, "Graphcut textures: Image and video synthesis using graph cuts", *ACM Transactions on Graphics*, Vol. 22, No. 3, (2003), pp: 277-286. https://doi.org/10.1145/882262.882264.
[38] A. Bargteil, F. Sin, J.E. Michaels, T.G. Goktekin, J.F. O'Brien, "A Texture Synthesis Method for Liquid Animations", *Proceedings of the Eurographics Symposium on Computer Animation*, ACM, (2006). https://doi.org/10.1145/1179849.1179929.
[39] G. Winkenbach, D.H. Salesin, "Computer-generated pen-and-ink illustration", *Proceedings of the 21st annual conference on Computer graphics and interactive techniques, ACM*, (1994). https://doi.org/10.1145/192161.192184.
[40] M. Bertalmio, G. Sapiro, V. Caselles, C. Ballester, "Image inpainting", *Proceedings of the 27th annual conference on Computer graphics and interactive techniques, ACM*, (2000).
[41] A. Efros, T. Leung, "Texture synthesis by non-parametric sampling", *Proceedings of the 7th International Conference on Computer Vision*, (1999). https://doi.org/10.1109/ICCV.1999.790383.
[42] A. Criminisi, P. Pérez, K. Toyama, "Region filling and object removal by exemplar-based image inpainting", *IEEE Trans. on image processing*, Vol. 13, No. 9, (2004), pp: 1200-1212. https://doi.org/10.1109/TIP.2004.833105.
[43] J. Aujol, S. Ladjal, S. Masnou, "Exemplar-based inpainting from a variational point of view", *SIAM Journal on Mathematical Analysis*, Vol. 42, No. 3, (2009), pp: 1246-85. https://doi.org/10.1137/080743883.
[44] L.Y. Wei, M. Levoy, "Fast texture synthesis using tree-structured vector quantization", *Proceedings of the 27th International Conference on Computer Graphics and Interactive Techniques, ACM*, (2000). https://doi.org/10.1145/344779.345009.
[45] ITU-R (2011) Recommendation BT.601-7. Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios, (2011).
[46] M. Weinberger, G. Seroussi, G. Sapiro, "The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS", *IEEE Trans. on Image Processing*, Vol. 9, No. 8, (2000), pp: 1309–1324. https://doi.org/10.1109/83.855427.
[47] I.E. Richardson, The H.264 Advanced Video Compression Standard, John Wiley & Sons, (2011).
[48] ITU-T (2000) Recommendation H.262, Information technology – Generic coding of moving pictures and associated audio information: Video, (2000).
[49] ISO (2004) Standard ISO/IEC 14496-2:2004 - Information technology -- Coding of audio-visual objects -- Part 2: Visual, (2004).
[50] ITU-T (2005) Recommendation H.263, Video coding for low bit rate communication, (2005).
[51] ITU-T (2015) Recommendation H.265, International Standard ISO/IEC 23008-2, High Efficiency Video Coding, (2015).
[52] S. Kullback, R.A. Leibler, "On Information and Sufficiency. Annals of Mathematical Statistic", Vol. 22, No. 1, (1951), pp: 79-86. https://doi.org/10.1214/aoms/1177729694.
[53] URL (2018) YUV test sequences. URL http://videocoders.com/yuv.html. Accessed on April 3, 2018.