A comparative study between Spatio-temporal orthogonal moments for volumes description

Manel Boutaleb, Imen Lassoued and Ezzeddine Zagrouba
Team of research SIIVA, RIADI laboratory
High Institute of computer science (ISI), University of Tunis El Manar
manel.bout@yahoo.fr, lassoued.imen@yahoo.fr, ezzeddine.zagrouba@fsm.mu.tn

Abstract— Object motion description in videos is one of the most active research topics in pattern recognition and computer vision. In this paper we study and compare Krawtchouk, Tchebytchev and Zernike spatio-temporal moments for volumes description. Indeed, reconstruction process of sequences volumes is elaborated to select the best moment descriptor for spatio-temporal volumes. Structural and temporal information of a video sequence can be captured by this moments. The first step of this method is to segment the video into volume space-time images. Then, all objects silhouettes will be extracted from these images. So this set will define the space-time form. The next step is to apply the orthogonal space-time moments on the resulting shape or just on the silhouettes defines patches around the interest detected points. This approach allows to define a descriptor for each video in the database. These descriptors will then rebuild the volumes of silhouettes with different orders to select the optimal for description process.

Keywords: video; moments; reconstruction; description.

I. INTRODUCTION

Object description, in videos, is a particularly complex topic in the area of computer vision which involves the automatic classification of objects. Indeed, the selection of a good descriptor is crucial in many areas such as video surveillance, human-computer interaction and indexing of video.

Indeed, this paper presents a comparative study between three orthogonal spatio-temporal moments: Krawtchouk, Tchebytchev and Zernike. This study is used to reconstruct volumes silhouettes based on the use of these moments to judge which is are more superior for object description. We denote Section 2 to review previous works of moments. Then we will describe our work in Section 3. Section 4 will deal with evaluating Weizmann action dataset and reporting experimental results. And finally, we end up by presenting some possibilities for future work in Section 5.

II. THEORY OF MOMENTS

Moments were widely used for many applications such as image security (insertion and detection of watermark [1], [2]), pattern recognition and object identification (recognition of facial features [3], aircraft identification [4], etc.), edge detection [5], texture analysis [6], image indexing [7], image recognition [8], etc.. They have the ability to represent the global characteristics of the image shape. The low order moments represent the global characteristics of form, while the higher order moments characterize the details of the image. There are two types of moments: not-orthogonal moments and orthogonal moments. Teague showed that reconstruction with orthogonal moments is much less demanding and quicker than not orthogonal moments because of the characteristic function of orthogonality which is in the form of a bijective function. In this work, we focus on the theory of three orthogonal moments: Zernike, Tchebytchev and Krawtchouk.

A. Zernike moments

Zernike moments were first proposed by Teague [9]. They are based on continuous orthogonal polynomials defined in the polar domain to represent the properties of an image without redundancy or overlap of information between the moments. The formulation of these moments seem to be one of the most popular, surpassing the alternatives (in terms of noise resilience, information redundancy and reconstruction possibilities). The magnitudes of Zernike moments are invariant to rotation. They are also robust to noise and minor variations in shape. Since the basis is orthogonal, they have minimum information redundancy. In many applications, Zernike moments were used to classify images and shapes of objects [10] and 2D image analysis [11].

In [11], results shows that Zernike moments are more robust in noisy and more efficient in terms of image representation. They are able to reach a value close to zero redundancy in a set of functions when the moments correspond to the independent characteristics of the image.

B. Tchebytchev moments

Tchebytchev moments were first proposed by Mukundan [12]. They are orthogonal in the field of the image coordinate space. They have a feature that completely eliminates the need for any discrete approximation in their numerical implementation. They also reduces the overall cost calculation. These moments are used in several studies [12], [13], [14] for the description and analysis of 2D images. Other work is dedicated to the security images (Insertion and detection of watermark [15]). the results of this work shows that these moments are more robust and less sensitive to noise compared to other type of moments.

C. Krawtchouk moments

A new series of moments orthogonal polynomials on the basis of discrete classical Krawtchouk is introduced by Yap and al [16]. The Krawtchouk polynomials play an important role in theories of coding and graphs [16], [17], [18]; they
forms an important family of orthogonal polynomials. This
moments have properties well suited to the characteristics of
model in the analysis of two-dimensional images. In [16],
Krawtchouk moments can be used to extract local features of
an image, unlike other times orthogonal, which tend to capture
the overall characteristics. Studies shows that Krawtchouk
moments are superior in terms of reconstruction error
compared to Zernike, pseudo-Zernike, Legendre and
Tchebytchev moments. Another watermarking technique based
on Krawtchouk moments is proposed in [19] to estimate the
gometric distortion parameters including rotation angle, the
new scale and translation parameter facing a rotation,
translation and resizing. Experimental results shows that the
technique of image watermarking is robust against the
proposed treatment of the image, such as translation, rotation,
additive noise and image compression.

III. PROPOSED APPROACH

Our work is based mainly on the use of orthogonal space-
time moments to describe and characterize silhouettes of
different pattern in videos as accurately and faithfully as
possible. We focus our attention on the stage of silhouettes
volumes description and reconstruction, while studying the
impact of choice of the moments orders in the description
silhouette movements. We makes a comparative study of
spatio-temporal Zernike, Krawtchouk and Tchebytchev
moments to reconstruct 3D shapes based on the order of
moments that present the main criterion for choosing the right
result. In our work, the action of a moving object is described
by silhouettes volumes which generates over time. This
volume is considered as a form of 3D. Fig.1 shows the
different stages of labor.

The first step consists in performing a video temporal
segmentation to obtain a volume of images. Then, the
background of this set of images is extracted to obtain the
silhouettes volume of object. Two methods are presented in
the first step of video description: The first is to calculate directly
the Zernike, Krawtchouk and Tchebytchev spatio-temporal
moments for each series of silhouettes obtained with different
orders. The second method focuses on extracting interesting
points from each volume of silhouettes using the SIFT
detector. Then we calculate the moment vectors for each
patches around the interest detected points.

The next step is to perform a reconstruction of silhouettes
volume using the founded moment descriptors. This helps us to
find the optimal order to describe properly the entire
silhouettes volume of the object in motion. Different orders of
moments are tested to determine which of these moments is the
best for video action description.

A. Space-time volume extraction

The first step of our work is silhouettes extraction of the
object which is performing a video action. We detect the
moving object through a background subtraction. We then use
the dictionary background subtraction described in Kim et al
[51]: this is a method that can handle scenes containing
moving backgrounds or illumination variations. It managed to
have a robust detection for different types of videos.

This is achieved by making a separation between the
intensity and chromaticity in the model background. Is a result,
this model is a multi modal and a multi-layer which models the
basic movement such as tree branches. This robust method of
background subtraction allows us to use a video with several
scenarios to demonstrate our reconstruction method and action
description.

B. Silhouettes reconstruction by orthogonal moments

After extracting the spatio-temporal volume, we apply the
formulas of Krawtchouk, Zernike and Tchebytchev moments
in space-time area on all pixels of silhouettes for volumes
reconstruction.

1) Spatio-temporal Krawtchouk moments

We used the expressions of 3D Krawtchouk moments of
the work presented in [20]. We focused on the condition of
orthonormality, so we define the orthonormal Krawtchouk
polynomial by:

\[ \tilde{K}_n(x; p; N_t) = \frac{w(x; p; N_t)}{h(n; p; N_t)} K_n(x; p; N_t) \]  

(1)

This extension Krawtchouk moments is described in (2D +
t). (2D + t) Krawtchouk moments of the function f with order
(n + m + l) are obtained from the weighted Krawtchouk
polynomials as the following:

\[ \tilde{C}_{n,m,l} = \sum_{x=0}^{N_x} \sum_{y=0}^{N_y} \sum_{t=0}^{N_t} \tilde{K}_{n,m,l}(x,y,t) f(x,y,t) \]  

(2)

Where :

\[ \tilde{K}_{n,m,l} = \tilde{K}_n(x,p,N_x) \tilde{K}_m(y,p,N_y) \tilde{K}_l(t,p,N_t) \]  

(3)
The projection of any 2d+t function \( f(x,y,t) \) can be completely reconstructed using the Krawtchouk moments as follows:

\[
f (x, y, t) = \sum_{x=0}^{N_x} \sum_{y=0}^{N_y} \sum_{t=0}^{N_t} \tilde{K}_{n,m,t} \tilde{C}_{nml}
\]

(4)

We used these moments as a characteristic spatio-temporal videos expressed as a function \( f(x,y, t) \). We recall that \( f (x, y, t) \) is a discrete binary volume. The parameters \( N_x \), \( N_y \) and \( N_t \) are in correlation with the dimensions of the binary volume in video.

2) Spatio-temporal Zernike moments

To make the volumes description and reconstruction, we used the extension of Zernike moments in the space time area presented in [21]. They used expressions of 2D Zernike moments from Teague.

Spatio-temporal Zernike moments noted \( A_{nml} \), are expressed by:

\[
A_{nml} = \frac{m+1}{\pi} \sum_{n=1}^{m} \sum_{m=1}^{n} U(t, \mu, \gamma) S(m,n) P_{m,n}(x,y)
\]

(5)

Where \( n \) is the order of moment and \( m \) is an integer such that \( n - |m| \leq 0 \) and \( P_{m,n}(x,y) \) is the value of the pixel \( (x, y) \) in the image at time \( t \). The expression of orthogonal polynomials is:

\[
S(m,n) = \left[ V_{nm}(r, \theta) \right]^2
\]

(6)

where * denotes the complex conjugate. \( V_{nm}(x,y) \) are the Zernike polynomials and are expressed in polar coordinates with:

\[
V_{nm}(r, \theta) = R_{nm}(r) \exp(jn\theta)
\]

(7)

where \( r, \theta \) are defined on the unit circle and \( R_{nm}(r) \) is the radial orthogonal polynomial defined by:

\[
R_{nm} = \sum_{s=0}^{n-|m|} (-1)^s s! \left( \frac{s!}{s! \left( \frac{s+|m|}{2} - s \right) ! \left( \frac{s+|m|-|n|}{2} - s \right) !} \right)
\]

(8)

With:

\[
F(m,n,s,r) = \frac{(m-s)!}{s! \left( \frac{s+|m|}{2} - s \right) ! \left( \frac{s+|m|-|n|}{2} - s \right) !}
\]

(9)

So the volume of silhouettes can be reconstructed by:

\[
\tilde{f}(x, y, z) = \sum_{x=0}^{N_x} \sum_{y=0}^{N_y} \sum_{t=0}^{N_t} A_{nml} \times V_{nm}(r, \theta)
\]

(10)

3) Spatio-temporal Tchebytchev moments

The discrete polynomial Tchebytchev 1D on the point \( x \) and order \( n \) is defined as follows:

\[
t_1(i) = (1-N)_i, t_2(n_i-x, 1+n; 1-N; 1), n, x, y \in \{0,1,2,...,N-1\}
\]

(11)

And \( F_2 \) is the generalized hypergeometric function:

\[
F_2(a_1, a_2, a_3, b_1, b_2; z) = \sum_{k=0}^{\infty} \frac{(a_1)_k(a_2)_k(a_3)_k z^k}{(b_1)_k(b_2)_k k!}
\]

(12)

The symbol \( (a)_k \) is the Pochhammer symbol given by:

\[
(a)_k = a(a+1)...(a+k-1)
\]

(13)

Then we describe the extension \( (2D + t) \) of Tchebytchev moments with order \( (n + m + 1) \) of the function \( f \). These moments are obtained from the weighted Tchebytchev polynomials defined by:

\[
T_{nml} = \frac{1}{\rho(n,N)\rho(m,M)\rho(l,nb)} \sum_{n=0}^{N} \sum_{m=0}^{M} \sum_{l=0}^{nb} R_{nml} f(x,y,t)
\]

(14)

Where

\[
R_{nml} = t_n(x) t_m(y) t_t(t)
\]

(15)

The normalization constant defined by:

\[
\rho(n,N) = (2n)! \left( \frac{N+n}{2n+1} \right)
\]

(16)

The projection of a function \( 2d+t \) \( f(x, y, t) \) can be completely reconstructed using the Tchebytchev moments as follows:

\[
f(x,y,t) = \sum_{x=0}^{N_x} \sum_{y=0}^{N_y} \sum_{t=0}^{N_t} T_{nml} f(x,y,t)
\]

(17)

\( f(x,y,t) \) is a discrete binary volume. The parameters \( N \), \( M \) and \( nb \) are correlated with the dimensions of the volume in the binary video.

C. Silhouettes reconstruction by interesting points and orthogonal moments

1) Interest points detection

Among all the features that can be extracted from videos, spatio temporal interest points are particularly interesting because they are low-level features simple and robust that allow a good characterization of moving objects. We used the SIFT detector (Scale Invariant Feature Transform) proposed by David Lowe to detect points that are extrema in each volume extracted silhouette.

2) Orthogonal moments computing

We apply Krawtchouk, Zernike and Tchebytchev moment expressions presented in the previous section (3.2) on detected interest points and their neighbors. Each calculated vector describes patches around the interest detected points and support the volume reconstruction:

a) Spatio temporal Tchebytchev moments

Tchebytchev moment vectors \( (2D + t) \) of order \( (n + m + 1) \) of interest points founded \( f(i,j,z) \) and their neighbors is obtained from the weighted Tchebytchev polynomials defined as follows:
Thus, the reconstruction volumes of silhouettes can be achieved using Tchebytchev moments of each interest points as follows:

\[ f(x, y, t) = \sum_{i=0}^{N_x} \sum_{j=0}^{N_y} \sum_{t=0}^{T-1} T_{n}(i) T_{m}(j) t_{f}(z) \]  

(20)

b) Spatio temporal Krawtchouk moments

Krawtchouk moments \((2D + t)\) of order \((n + m + l)\) for each interest point obtained \(I(i, j, z)\) from the weighted Krawtchouk polynomials are defined as follow:

\[ \tilde{C}_{nm} = \sum_{i=0}^{N_x} \sum_{j=0}^{N_y} \sum_{t=0}^{T-1} K_{n,m,l}(i) I(i, j, z) \]  

(21)

Where:

\[ K_{n,m,l} = K_{n}(i; p_{i}, N_{i}) K_{m}(j; p_{j}, N_{j}) K_{l}(z; p_{z}, N_{z}) \]  

(22)

Silhouettes volumes can be completely reconstructed using spatio-temporal Krawtchouk moment vectors of interest points as follows:

\[ f(x, y, t) = \sum_{i=0}^{N_x} \sum_{j=0}^{N_y} \sum_{t=0}^{T-1} K_{n,m,l} \tilde{C}_{nm} \]  

(23)

c) Spatio temporal Zernike moments

spatiotemporal Zernike moments \(A_{nm}^{l}\) of interest points founded \(I(i, j, t)\) are expressed by:

\[ A_{nm}^{l} = \frac{m+1}{\pi} \sum_{\gamma} \sum_{\gamma} U(t, \mu, \gamma) S(m, n) I(i, j, t) \]  

(24)

Thus, the volume of silhouettes can be reconstructed by:

\[ \tilde{f}(x, y, z) = \sum_{i} \sum_{j} \sum_{z} A_{nm}^{l} \times V_{nm}(r, \theta) \]  

(25)

IV. EXPERIMENTAL RESULTS

A. Weizmann Dataset:

The reconstruction process was evaluated on a publicly available benchmark dataset of human action: \#Weizmann dataset0 (Fig.2). This dataset composed by 90 low resolution videos \((180 * 144, 50\) fps) where 9 persons performs 10 actions (bend, jack, jump, jump in place, run, side, skip, walk, wave1hand and wave2hand).

B. Experimentation process:

Experimentation begins by segmenting each video of Weizmann dataset in the space-time volume. Then, the associated silhouettes are extracted. We make two methods of silhouettes volumes reconstruction: in the first method, spatio-temporal Krawtchouk, Tchebytchev and Zernike moments are calculated for each silhouettes volumes to characterize video.

The second method involves, started by extracting the interesting points of each silhouette. Then calculating the moment vectors for patches around the interest detected points. In order to have a convincing study of the orders in these two methods, we conducted several tests that predict the reconstruction quality. In fact, we tested several different orders \((3, 15, 30, 45, 60, 75, 90, 120)\) and, every time we increased the order, we tested and observed the obtained results to choose the optimal order. So that we describe adequately our videos.

C. Experimental results and discussion

1) Silhouettes reconstruction by orthogonal moments

Reconstruction results by this first method are given in the following figure (Fig.2), with a moments order varied from 3 to 120. We note, visually, that the lower orders \((3, 15 \text{ and } 30)\) do not well describe silhouette volumes. From this following figure we can infer that the spatio-temporal Krawtchouk moments is beginning to improve at the order 45. Casting a glance at the volume resulting from order 75, we notice that it is the best order to get a good reconstruction result similar to the original volume. However, for the results of spatio-temporal Tchebytchev moments reconstruction, we notice that from the order 90, silhouettes volume is reconstructed with a good quality compared to the lower order. Zernike moments are calculated only for square images. This, brought us to perform a series of silhouettes scaling to a common size of 103 * 103 pixels. After a long test of result, we can say that from the order 30, Zernike gives an acceptable volumes reconstruction. But this volumes starts to lose his quality with a poor performance from an order higher than 60.

![Figure 2: Reconstruction of 'bend' action with Krawtchouk, Tchebytchev and Zernike spatio temporal moment by method1](image)

After achieving the reconstruction by different moments, we need to do a comparative study between moments to determine which among them has a high description rate. In this case, we applied two performance metrics: CPU time and PSNR quality (the quality of each reconstructed volume is measured in terms of peak signal-to-noise ratio).

The following table (TAB.1) describes execution time for the reconstruction of silhouettes volumes between Krawtchouk, Tchebytchev and Zernike moments tested with
different orders. In fact, we recorded the execution time for each volume then we rebuilt their average. From the order 60, we notice that there is a big difference between the execution time of Krawtchouk and those of Zernike and Tchebytchev. Thus, Krawtchouk execution time is much faster (3.2452 seconds) than Tchebytchev (27.4522 seconds). Zernike moments present a too slow execution time compared to the two other moments that is just 2 hours, which is not only practical in the real time application, but also not logical to rebuild a single video in two hours time, as in our case the dataset contains 90 videos of different actions.

### TABLE 1. CPU TIME OF RECONSTRUCTION PROCESS BY METHOD 1 (in seconds)

<table>
<thead>
<tr>
<th>Order</th>
<th>Krawtchouk</th>
<th>Tchebytchev</th>
<th>Zernike</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.9228</td>
<td>5.4521</td>
<td>10.9851</td>
</tr>
<tr>
<td>15</td>
<td>2.3045</td>
<td>9.7452</td>
<td>103.4521</td>
</tr>
<tr>
<td>30</td>
<td>2.5012</td>
<td>12.5124</td>
<td>524.4569</td>
</tr>
<tr>
<td>45</td>
<td>3.4521</td>
<td>20.8239</td>
<td>1527</td>
</tr>
<tr>
<td>60</td>
<td>3.2452</td>
<td>27.4522</td>
<td>6324.75</td>
</tr>
<tr>
<td>75</td>
<td>2.8795</td>
<td>30.6521</td>
<td>-</td>
</tr>
<tr>
<td>90</td>
<td>3.4578</td>
<td>39.5126</td>
<td>-</td>
</tr>
<tr>
<td>120</td>
<td>4.2145</td>
<td>45.4521</td>
<td>-</td>
</tr>
</tbody>
</table>

In our experiments, we recorded the PSNR for each reconstructed video. Then, we calculated the average of these PSNRs. This average described the reconstruction quality of the entire database actions. We plotted as well, the video quality (normalized PSNR) according to the different orders of Krawtchouk, Tchebytchev and Zernike moments used for reconstruction. Test results are presented in the following figure (Fig.3), from which important conclusions can be drawn: Increasing each time the order of Krawtchouk moment, the quality of reconstructed volumes is improving gradually. The quality of volumes reconstructed by Zernike is increasing and getting better, but while reaching the order 30, volumes lose their qualities. Similarly for Tchebytchev, from the order 75, PSNR decreases to reach the order 90. Then, from this order, the quality remains constant even if we increase orders.

2) **Silhouettes reconstruction by interesting points and orthogonal moments**

We present in this section the results obtained by calculating Krawtchouk, Tchebytchev and Zernike moment for each patches around the interest detected points. The results of reconstruction are given in the following figure (Fig.4) with a moments order varied from 3 to 160. The results presented with Krawtchouk moments shows that from the order 30, volumes begins to rebuild. Reaching around the order 90, result is improved and volumes are well reconstructed and becomes similar to the originals. As for the obtained results of Tchebytchev moments, we note that the orders superior or equal to 120 gives a well-reconstructed silhouette volume with some quality degradation. Calculation using Zernike moment requires a too long execution time. This is why we calculated this moment just for the orders 3,15,30, 45 and 75. At the order 15, the volumes begins to rebuild. But after increasing the order, we note that reconstruction does not give the expected result, and poor results are emerged. This is explained by their characteristic sensitivity to noise levels with the greatest.

![Figure 3: Reconstructed video quality according to the orders of Krawtchouk, Tchebytchev and Zernike moments by method1](image)

![Figure 4: Reconstruction of bend action with Krawtchouk, Tchebytchev and Zernike spatio temporal moment by method2](image)

<table>
<thead>
<tr>
<th>Order</th>
<th>Krawtchouk</th>
<th>Tchebytchev</th>
<th>Zernike</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>7.8141</td>
<td>8.5021</td>
<td>300</td>
</tr>
<tr>
<td>15</td>
<td>8.1901</td>
<td>9.5005</td>
<td>1500</td>
</tr>
<tr>
<td>30</td>
<td>8.4553</td>
<td>10.0605</td>
<td>4500</td>
</tr>
<tr>
<td>45</td>
<td>8.2681</td>
<td>9.9793</td>
<td>7200</td>
</tr>
<tr>
<td>75</td>
<td>9.0621</td>
<td>10.7017</td>
<td>9400</td>
</tr>
<tr>
<td>90</td>
<td>9.0621</td>
<td>11.0933</td>
<td>-</td>
</tr>
<tr>
<td>120</td>
<td>10.4878</td>
<td>12.4209</td>
<td>-</td>
</tr>
<tr>
<td>160</td>
<td>12.2155</td>
<td>14.3413</td>
<td>-</td>
</tr>
</tbody>
</table>

**Figure 4. Reconstruction of bend action with Krawtchouk, Tchebytchev and Zernike spatio temporal moment by method2**

**TABLE 2. CPU TIME OF RECONSTRUCTION PROCESS BY METHOD 2 WITH: KRAWTCHOUK, TCHEBYTCHEV AND ZERNIKE MOMENTS (in seconds)**
We plotted the videos quality (normalized PSNR) in function of the number of orders of Zernike, Tchebytchev and Krawtchouk polynomials (Fig.5). It shows that Krawtchouk moments PSNR quality reconstruction is much better than Tchebytchev and Zernike moments.

![Figure 5. Reconstructed video quality according to the orders of Krawtchouk, Tchebytchev and Zernike moments by method2](image-url)

V. CONCLUSION

In this paper, we made a comparative study between three spatio temporal orthogonal moments. We use for this purpose two different methods that helps to study the suitability for the action description process of moving objects in video. Results interpretation confirms the superiority description rate by using spatio-temporal Krawtchouk moments taking into account two criteria: lower computation time and a good reconstruction quality. The work will deal with a more realistic datasets for different object types to be generalized.

REFERENCES


