International Journal of Engineering

Journal Homepage: www.ije.ir

# Pain Facial Expression Recognition from Video Sequences Using Spatio-temporal Local Binary Patterns and Tracking Fiducial Points

I. Firouzian*[a], N. Firouzian[b], S. M. R. Hashemi[a], E. Kozegar[c]

[a] *Computer Engineering & IT Department, Shahrood University of Technology, Shahrood, Iran*
[b] *Department of Strategic Management, Bank Melli Iran, Tehran, Iran*
[c] *Faculty of Engineering, East Guilan, University of Guilan, Guilan, Iran*

PAPER INFO

*Paper history:*
Received 10 November 2017
Received in revised form 23 December 2017
Accepted 04 Janurary 2018

*Keywords*:
*Facial Expression*
*Tracking Fiducial Points*
*Spatio-temporal*
*Local Binary Patterns*
*Pain Expression*
*Video Sequences*

ABSTRACT

Monitoring the facial expressions of patients in clinical environments is a necessity in addition to vital sign monitoring. Pain monitoring of patients by facial expressions from video sequences eliminates the need for another person to accompany patients. In this paper, a novel approach is presented to monitor the expression of face and notify in case of pain using tracking fiducial points of face in video sequences and spatio-temporal Local Binary Patterns (LBPs) for eyes and eyebrows. The motion of eight fiducial points on facial features such as mouth, eyes, eyebrows are tracked by Lucas-Kanade algorithm and the movement angles are recorded in a feature vector which along with the spatio-temporal histogram of LBPs creates a concatenated feature vector. Spatio-temporal LBPs boost the proposed algorithm to capture minor deformations on eyes and eyebrows. The feature vectors are then compared and classified using the Chi-square similarity measure. Experimental results show that leveraging spatio-temporal LBPs improves the accuracy by 12% on STOIC database.

***doi****: 10.5829/ije.2020.33.05b.38*

## 1. INTRODUCTION

Facial expression recognition is an active research area that covers different subjects such as Human Computer Interaction (HCI), Smart Environments and medical applications. Recognition of facial expressions is a hard task due to several limitations such as lighting conditions, facial deformations, facial occlusions or facial hair.

Since lots of efforts have been done on prototypic facial expressions, the focus of this paper is mainly on pain facial expression. Applications of pain facial expression is primarily related to medical applications. It is of considerable value in certain circumstances where patients are not unable to communicate pain verbally (e.g. newborns, individuals with severe cognitive impairments like autism).

The task of facial expression recognition can be divided into three main steps of face detection, facial fea--ture extraction, facial feature classification. Face detection is a process of locating a face in an image for further processing. Facial feature extraction is a method applied to represent the facial expressions in terms of feature vectors. Facial feature classification which is the step in which extracted features are classified into expression classes.

Related works in the field of pain facial expression is currently limited to Zakia Hammal et al. [1], Littlewort et al. [2], Monwar et al. [3]. Therefore, we make a quick review of the methods in facial expression recognition area as a whole, not just pain facial expression.

Lots of methods have been presented in literature for facial expression recognition such as Neural Network [4, 5], Support Vector Machine (SVM) [6], Bayesian Network (BN) [7], and rule-based classifiers [8-10]. In addition to these methods, some descriptors have also been proposed to discriminate a face from others. Gabor wavelet, Local Binary Pattern (LBP), Haar-like descriptors, shape descriptors [11-15]. These techniques,

*Corresponding Author Email: iman.firouzian@shahroodut.ac.ir (I. Firouzian)

Please cite this article as: I. Firouzian, N. Firouzian, S. M. R. Hashemi, E. Kozegar, Pain Facial Expression Recognition from Video Sequences using Spatio-temporal Local Binary Patterns and Tracking Fiducial Points, International Journal of Engineering (IJE), IJE TRANSACTIONS B: Applications Vol. 33, No. 5, (May 2020) 1038-1047

methods and descriptors regardless of their accuracy are time-consuming and they are not recommended for real-time applications. Hence, a simple, yet powerful real-time approach is required for pain facial expression.

The proposed method, extracts facial features and places fiducial points on the face according to the result of facial feature extraction process. Displacements of fiducial points across different frames carry the information about which facial expression is taking place. A vector of displacement directions is created for each fiducial point.

Since the selection of fiducial points in the facial features of eyes and eyebrows is fairly difficult, and the range of motions of fiducial points for the two facial features is also not significant, the associated feature vectors do not carry enough information to distinguish facial expression from each other [16]. To resolve the issue, a feature vector of spatio-temporal LBP histograms is applied on eyes and eyebrows, and is added to the feature vector of fiducial point displacements.

The feature vectors of displacement directions are classified using the Chi-square similarity measure. One advantage of our proposed method is that the whole facial expression information turns into a few sequences that are easily comparable by similarity measures.

The rest of the paper is organized as follows. Section 2 briefly reviews existing facial expression methods, techniques and descriptors. In Section 3, our proposed algorithm is fully explained. Section 4 presents the results and finally this paper is concluded in Section 5.

## 2. RELATED WORKS

There are some descriptors in literature for facial expression recognition. Gabor wavelet is the most frequently used descriptor (e.g. Gwen et al. [15]), which its time complexity is considerable. Some texture descriptors have also been tested in facial expression recognition domain. The LBP on arbitrarily gridded sub-regions (e.g. Shan et al. [14]) and Haar-like descriptors extended to variant rectangles from Kim et al. [12] are of texture descriptor samples. Some authors proposed to use shape descriptors such as Irene et al. [13] who considered shape information from facial grids based on a set of landmarks, or Zhu et al. [11] who computed moment invariants on several manually annotated faces areas. However, most of these descriptors support static images and do not consider the temporal information of facial action units.

To exploit the temporal information of facial action units, different techniques were presented for facial expression recognition from image sequences. There have been several attempts to track and recognize facial expressions over time [17, 18].

Monwar et al. [3] proposed a video based technique for facial expression recognition of acted pain sequences in a 2-alternative forced choice classification (pain vs. painless). Black and Yacoob [19] used a local parameterized model of image motion obtained from optical flow analysis. They utilized a planar model for rigid facial motion and an affine plus-curvature model for non-rigid motion. Tian et al. [4] presented a Neural Network based approach to recognize facial action units in image sequences. Essa and Pentland [20] first locate the nose, eyes and mouth. Then, from two consecutive normalized frames, a 2D spatio-temporal motion energy representation of facial motion is used as a dynamic face model. Hidden Markov Models (HMMs) have been widely used to model the temporal behaviors of facial expressions from image sequences [7, 21]. Ira et al. [7] proposed a multi-level HMM classifier which automatically segments a long video sequence to different expressions segments without resorting to heuristic methods of segmentation. But HMMs can't deal with dependencies in observation. Dynamic Bayesian Networks (DBNs) recently were exploited for sequence-based expression recognition [22-24]. Tian et al. [4] proposed a feature-based method, which uses geometric and motion facial features and detects transient facial features. The extracted features (mouth, eyes, brows and cheeks) are represented with geometric and motion parameters. The furrows are also detected using a Canny edge detector to measure orientation and quantify their intensity. The parameters of the lower and upper face are then fed into separate neural networks trained to recognize Action Units (AUs). Kaliouby et al. [23] proposed a system for inferring complex mental states from videos of facial expressions and head gestures, where a multi-level DBN classifier was used to model complex mental states as a number of interacting facial and head displays. Firouzian et al. [25] introduced a feature extraction method for real-time responses and the method is employed in this paper. Zhang and Ji [24] explored the use of multisensory information fusion technique with DBNs for modeling and understanding the temporal behaviors of facial expressions in image sequences. Chang et al. proposed a probabilistic video-based facial expression recognition method based on manifolds [26]. Shan et al. introduced a Bayesian temporal model to combine this information, based on static information (LBP histogram on gridded small regions) from each frame in the sequence, so as to compute the posterior probability. Inspired by this probabilistic blending method, José et al. [27] built a real-time system based on deformation space and maximum posterior probability but the recognition rate is not competitive. In order to represent both static and dynamic information, Guoying et al. [28] connected the LBP features on three orthogonal planes to represent the video, taking the same approach for motion than the one used for appearance, regardless of their obvious different texture pattern. Note that the authors annotated manually the eye's position in the sequences. Some other authors explored expression recognition on non-frontal face images by using SIFT features [29], hybrid features of LBP and Gabor [30] or variable-intensity template. Bartlett et al. [6] performed systematic comparison of different techniques including AdaBoost, SVM and LDA for facial expression recognition, and best results were obtained by selecting a subset of Gabor filters using the AdaBoost and then training SVM on the outputs of the selected filters. Pantic and Rothkrantz adopted rule-based reasoning to recognize action units and their combination [9]. Other methods that use head pose information for classification in addition to facial expression was proposed in [31], Werner et al. [32]. Hashemi et al. [33] also evaluated various face identification methods. They proposed a method for a fast and real-time face detection and recognition on live cameras. Lucey et al. used 3D parameters derived from Active Appearance Model (AAM) along with facial expressions to detect pain and assess the level of pain based on the PSPI scale [31]. Werner et al. proposed a method for fully automatic detection of pain based on facial expressions and head pose information [32]. In this method, depth and color features are extracted at the frame level and a time window descriptor is calculated from them. Despite repeated claims of a breakthrough in mass media, the suitability of these automated computer vision systems for clinical use is still not given. Some of the available systems have developed impressive solutions for mapping the face [1, 31]. A large body of methods has been proposed to automatically assess pain using behavioral (e.g., facial expression [25, 34-38] and crying [33, 34]) or physiological (e.g., changes in vital signs [35, 36] and cerebral hemodynamic changes [37, 38]).

O'Neill et al. [39] videotaped one hundred infants during their routine vaccination appointment and facial configurations were graphed in 5-second epochs for 1-min post-vaccination and subsequently analyzed for facial expression by time effects using Repeated Measures ANOVAs at each age. Virrey et al. evaluated the facial expression datasets as a basis for building and evaluating the largely unmapped facet of pain expressions [40]. Their study provides the summary of different characteristics of expressions that are relevant and justifiable indicators of pain. They also tested a preliminary platform with accuracy rate of 85.66% using the collected FER datasets as testing inputs. Zamzmi et al. [41] presented an end-to-end N-CNN for automatic recognition of neonatal pain for an unconstrained condition. They proposed CNN designed and trained from the scratch to detect neonatal pain for the first time. The proposed N-CNN method achieved 91% average accuracy and 0.93 AUC.

## 3. PROPOSED METHOD

The proposed method of pain facial expression recognition is performed on video sequences of clinical context. The method classifies the video sequences into one of the eight facial expression classes: happiness, surprise, disgust, anger, fear, sadness, neutral and pain.

The proposed method tracks fiducial points on face along the time. Since the tracking approach significantly depends on appropriate fiducial point selection, the approach fails to capture regional smooth movements. Facial expression variations of eyes and eyebrows in different facial expressions have regional smooth movements which tracking fiducial points does not singly suffice. To resolve the issue, spatio-temporal LBPs are employed to capture regional smooth variations. The proposed method includes followings steps:

1. Facial feature extraction from the first frame.
2. Localization of fiducial points on the first frame.
3. Tracking fiducial points on the remaining video frames.
4. Theta angle computation according to displacements of fiducial points.
5. Spatio-temporal LBP histogram extraction from "eyes" and "eyebrows" facial features.
6. Classification of concatenated feature vectors including feature vector of theta angles associated to fiducial points and feature vector of LBP histograms of eyes and eyebrows into one of eight facial expressions using the Chi-square similarity measure.

In this next subsection, the facial feature extraction method is proposed. Note that facial feature extraction methods are usually time-consuming. The facial feature extraction method is not the main contribution but yet provides a simple way to achieve a desirable result within lesser time.

### 3. 1. Facial Feature Extraction

A preprocessing procedure is initially applied on video frames. The contrast of frames are first normalized and a Gaussian filter is then applied to reduce the probable noises. The applied Gaussian filter is a low-pass filter with window size of 3×3 pixels and standard deviation of 2. Then, vertical and horizontal edges are detected. Edge is actually a sudden change in image intensity values. A face image has vertical and horizontal edges because of its concavities and convexities; the biological features help in localization of eyes, eyebrows and mouth. There are lots of methods and algorithms in the literature which are designed to detect edges; one of which is Sobel operator which is a fast operator with low complexity in edge detection. Sobel operator leverages a mask for edge detection; if input image is not considerably rotated, the Sobel operator leverages two masks (drawn in Figure 1) to detect vertical and horizontal edges. The resultant image from the edge detection process is a binary image which pixels of edges are valued by one. Figure 2 shows results of the edge detection process. The input image was taken from the STOIC database [42].

**3. 2. Histogram Analysis** Once edges are detected in vertical and horizontal axis, two vertical and horizontal histograms are computed. The obtained vertical and horizontal histograms are as shown in Figure 3.

**3. 2. 1. Candidate Locations for Facial Features** In this subsection, candidate locations for eyes, eyebrows, mouth are localized using the horizontal and vertical histograms. An algorithm is designed by which the horizontal histograms are analyzed and regions of the image with continuous horizontal histograms are specified. Once regions of the image with continuous horizontal histograms are localized, vertical histogram from each of the regions is computed. Sub-regions with continuous vertical histograms are also extracted. Having computed and merged the information of vertical and horizontal histograms, candidate locations for the two eyes, two eyebrows and mouth are obtained. Since a set of small regions might also be introduced as candidate regions, it is recommended to ignore the regions by a specific size condition. By ignoring candidate region by the width or length less than 5 pixels, the candidate regions of a sample image would look like as in Figure 4.

Once candidate locations are localized for facial features, some undesirable incorrect locations are also detected. In order to refine the candidate locations and to obtain the exact locations for eyes, eyebrows and mouth, geometrical relationships among facial features is used as presented in [4]. Since locations and sizes of facial features, like eyes and mouth is not considerably variant across different human faces, the geometrical relationship is employed. In [4], the probable locations for facial features is determined as depicted in Figure 5.

In Figure 5, the two rectangles in the upper part of face designate probable geometrical locations for eyes and eyebrows and the rectangle at the lower part designates the probable geometrical location for mouth and nose. Since the probable location of nose is not necessary in this paper, we remove the associated rectangle. Therefore, the geometrical relationships operate like a refinement mask for candidate locations in Figure 4 to obtain the exact locations. If the mask is applied directly on a test image, the obtained image is shown in Figure 6.

Two results are achieved by this stage:

- Candidate locations for two eyes, two eyebrows and mouth as drawn in Figure 4.
- Geometrical relationships for facial features regarding the face size.

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} -1 & 2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

(A) (B)

**Figure 1.** Two masks of Sobel operator

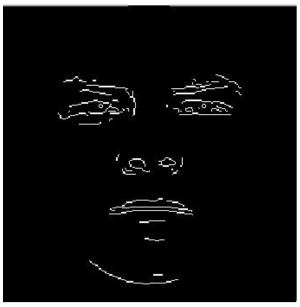

**Figure 2.** Sample output of edge detection process. The input image is taken from STOIC database.

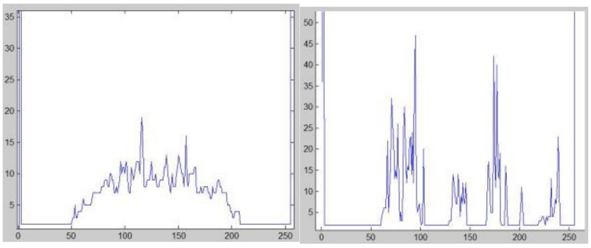

**Figure 3.** From left to right; horizontal and vertical histograms obtained from the detected edge image in Figure 2

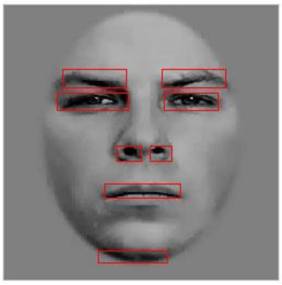

**Figure 4.** Candidate locations for facial features obtained from horizontal and vertical histograms

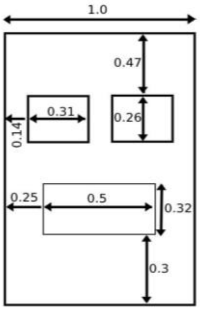

**Figure 5.** Geometrical relationships among facial features [4]

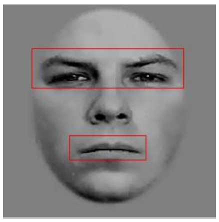

**Figure 6.** Candidate locations for eyes, eyebrows and mouth

Having merged the two results, the exact locations of facial features are obtained. The result of merging is shown in Figure 7.

### 3. 3. Fiducial Points Localization And Tracking

Subsequent to localization of facial features, fiducial points must be placed according to the obtained regions. Fiducial points are the points which should be tracked by Lucas-Kanade algorithm in the next frames of the movie. The eight fiducial points are shown in Figure 8. Four fiducial points from the mouth are placed in center sides of the rectangle designating mouth candidate location. Two fiducial points from the eyes are placed in inner corners of the rectangles designating eye candidate locations. The two fiducial points from the eyebrows are also placed in inner vertices of the rectangles designating eyebrows candidate locations.

Since determination of locations of fiducial points in every frame of the video is necessary and applying the facial feature extraction algorithms (even the least time consuming ones) on every frame of the video is not reasonable, the facial feature extraction algorithm is employed merely on the first frame and the fiducial points are tracked using the Lucas-Kanade algorithm [43]. In the following Lucas-Kanade algorithm, w is the warping function, I is the image and P is the parameter of warping function:

```
Iterate:
    1. Warp I with W([x y];P) => I(W([x y];P))
    2. Compute error image T(x) – I(W([x y];p))
    3. Warp gradient of I to compute ∇I
    4. Evaluate Jacobian
    $\nabla I \frac{\partial w}{\partial P}$
    5. Compute Hessian matrix
    $\sum(\nabla I \frac{\partial w}{\partial P})^T (\nabla I \frac{\partial w}{\partial P})$
    6. Compute
    $\sum\left(\nabla I \frac{\partial w}{\partial P}\right)^T (T(x,y) - I(W([x,y];P)))$
    7. Compute Δp
    8. Update w(x;p) ← w(x;p) o w(x;Δp)
Until Δp is negligible.
```

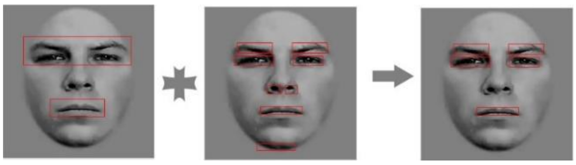

**Figure 7.** Left: Geometrical relationships for facial features. Middle: Facial features obtained from histograms. Right: Merging of the two

When the points are tracked in the next frames of the movie, the positions of fiducial points in all frames are obtained as shown in Figure 9. Two vectors including a vector of absolute displacement values and a vector of displacement directions are obtained using the positions of fiducial points. Hammal et al. [1] used distances between fiducial points but the distance between points would lose a set of spatial information. Therefore, in this paper, a vector of absolute displacement values along with a vector of displacement directions is considered.

### 3. 4. Spatio-temporal LBP

Since the selection of fiducial points in the facial features from eyes and eyebrows is fairly difficult and the range of motions of fiducial points for the two facial features is also not significant, the associated feature vectors does not carry enough information to distinguish various facial expressions. To resolve the issue, a feature vector of spatio-temporal LBP histograms applied on eyes and eyebrows is added to the feature vector of fiducial point displacements.

The basic LBP is defined to process spatial data. An extension to basic LBP leverages spatio-temporal data to analyze the dynamic texture over time. To this end, the 3D-LBP was presented by Zhao et al. [44] to handle the

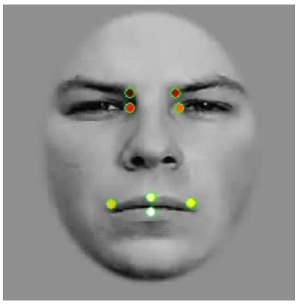

**Figure 8.** Eight fiducial points extracted from first frame. The points are to be tracked in next frames

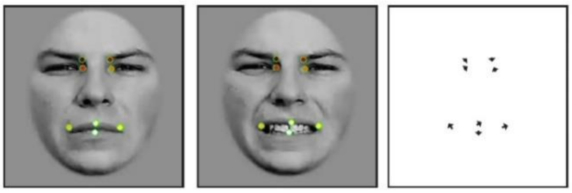

**Figure 9.** left & middle: two consecutive frames of the movie with designated fiducial points. right: vectors of displacements of fiducial points

problem. As shown in Figure 10, The main idea of the 3D-LBP lies in computation of texture volume in (X, Y, T) space, wherein X and Y are spatial components and T is a temporal component representing the index of the frame. First of all, adjacency of each pixel in the 3D space is defined. Then, volume textons are defined similar to the LBP in spatial domain and then associated histograms are extracted. Therefore, the 3D-LBP merges the appearance and the appearance changes in dynamic textures. To simplify the 3D-LBP computations, an operator based on common occurrence of local binary patterns is defined on three orthogonal surfaces known as LBP-TOP.

In LBP-TOP, LBPs associated to three orthogonal surfaces XY, XT, YT are combined such that only the common occurrence of the three surfaces are considered. A video file is usually considered as a stack of XY surfaces along the T axis. XT and YT surfaces also contain information about spatio-temporal changes. In LBP-TOP method, the number of histogram bins is only equal to 3×2P, where P is the number of facial features. This approach reduces the complexity and makes the progress of adjacent cells simpler. LBP-TOP uses three orthogonal surfaces which have the central pixel in common. LBP-TOP reduces the size of feature vector because of the combination of the three surfaces. Since, the change direction of the texture is unknown, not only the direction along the time axis is investigated, but also all circular adjacent points are investigated. Sample images of the three surfaces XY, XT and YT are shown in Figure 11.

The LBP codes obtained from XY, XT and YT are shown as XY-LBP, XT-LBP and YT-LBP respectively. The final histogram is obtained from the concatenation of the three. The process is shown is Figure 12. In this approach, the face expression changes are coded by XY-LBP, XT-LBP, YT-LBP. In face expression recognition, the radius along the time axis is preferably different from the radius in the spatial domain. The radius size along the time axis significantly depends on the frame rate. For example, for a video file with 300×300 resolution and 12 frames per second, an eight pixel radius may suffice, but for larger resolutions or lower frame rates, different radiuses for temporal and spatial domain is required. The difference in radius sizes makes the sampling shape change from circular to ellipsoidal.

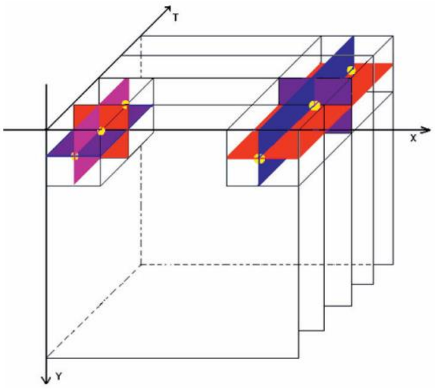

**Figure 10.** Three orthogonal surfaces in facial feature expression

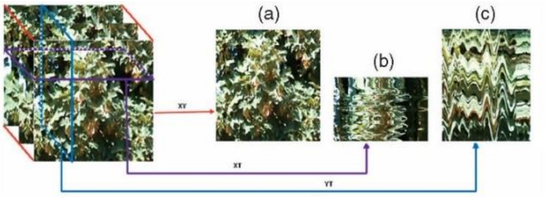

**Figure 11.** Sample images of the three surfaces: a) the XY surfaces; b) the XT surface which captures the horizontal movements; c) the vertical movements in temporal space

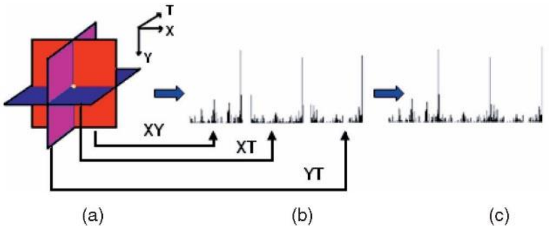

**Figure 12.** Three orthogonal surfaces in facial feature expression along with feature vector concatenation

In general, radiuses in X, Y and T axes as well as the number of adjacent points in XY, XT and YT surfaces may not be of the same size, and they are represented by $R_X, R_Y, R_T, P_{XY}, P_{XT}, P_{YT}$. Suppose components of the central pixel $g_{t_c,c}$ would be $(x_c, y_c, t_c)$, components of $g_{XY,P}$ are computed by $(x_c - R_x\, sin(2\pi p/P_{XY}), y_c + R_Y cos(2\pi p/P_{XY}), t_c)$, components of $g_{XT,P}$ are computed by $(x_c - R_x\, sin(2\pi p/P_{XT}), y_c, t_c - R_T cos(2\pi p/P_{XT}))$, and components of $g_{YT,P}$ computed by $(x_c,\ y_c - R_Y\, Cos(2\pi p/P_{YT}), t_c - R_T sin(2\pi p/P_{YT}))$.

Consider a facial feature expression as $\times Y \times T$ $(x_c \in \{0, \dots, X-1\}, y_c \in \{0, \dots, Y-1\},\ t_c \in \{0, \dots, T-1\})$. In calculating $LBP-TOP_{P_{XY};\,P_{XT};\,P_{YT};\,R_X;\,R_Y;\,R_T}$ distribution for facial feature expression, the central part is only considered because a sufficiently large neighborhood cannot be used on the borders in this 3D space. A histogram of the facial feature expression can be defined as:

$$H_{i,j} = \sum_{i=0,\dots,n_j-1,\ \ j=0,1,2} I\{f_j(x,y,t) = i\}, \tag{1}$$

in which $n_j$ is the number of different labels produced by the LBP operator in the $jth$ plane $(j = 0 : XY, 1 : XT\ and\ 2{:}\,YT), f_i(x,y,t)$ expresses the LBP code of central pixel $(x,y,t)$ in the $jth$ plane and

$$I\{A\} = \begin{cases} 1, \ if\ A\ is\ true; \\ 0, if\ A\ is\ false. \end{cases}$$

when the facial features to be compared are of different spatial and temporal sizes, the histograms must be normalized to get a coherent description:

$$N_{i,j} = \frac{H_{i,j}}{\sum_{k=0}^{n_j-1} H_{k,j}} \tag{2}$$

In this histogram, a description of facial feature is effectively obtained based on LBP from three different planes. The labels from the XY plane contain information about the appearance, and, in the labels from the XT and YT planes, co-occurrence statistics of motion in horizontal and vertical directions are included. These three histograms are concatenated to build a global description of facial feature with the spatial and temporal features.

**3. 5. Vectors of Displacement Directions** As shown in Figure 13, vectors of displacement directions are the key to distinguish among different kinds of facial expressions. Vectors of absolute displacement values just carry the velocity information and the information just represents how fast a facial expression is changed and would not help classifying facial expressions.

Since each facial expression has minor differences across different faces, fiducial points displacements are approximately equal across different subjects. Therefore, a vector of theta angles is created that would look like ($\Theta_1$, $\Theta_2$, … , $\Theta_n$). Each vector of displacement directions (theta angles) of each facial expression is unique and follows a specific pattern. Therefore, facial expression recognition is performed through classifying vectors of displacement directions. In the classification phase, training set is prepared as pivotal data. Pivotal data consists of 8 sequences of theta angles for each of the facial expressions (pain, neutral, happiness, surprise, disgust, anger, fear and sadness). Each of the sequences is associated to a fiducial point. Each sequence of pivotal data is obtained by taking the average of corresponding theta angle vectors of all training set subjects for each facial expression.

When a test facial expression video is prepared, vectors of displacement directions associated to each fiducial point are computed. The numbers of the vectors are equal to the numbers of fiducial points. To classify the test facial expression, the test vectors are compared with pivotal trained vectors. The comparison can is performed using Chi-square similarity measure.

As shown in Figure 14, vectors of theta angles from one of the mouth feature points across all frames of video for two distinct subjects are drawn. The two drawn sequences are related to happy facial expression. As it is shown in Figure 14, the values of two sequences are both ascending and follow a pattern. This property convey a discriminant feature for vector of thetas.

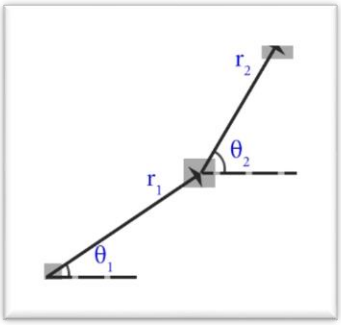

**Figure 13.** Displacement of a specific pixel can be defined by a displacement direction angle (Θ) and a displacement absolute value (r)

Since the variation speed of facial expressions is different even in a same person, time-scale translation is applied to the vector of displacement directions (theta angles). Therefore, the similarity measure is expected to support time-scale translation. In the next subsection, an appropriate similarity measure is introduced which satisfies the criteria.

**3. 6. Desirable Similarity Measure** A similarity measure is a type of scoring function that assigns a numerical score to a pair of sequences based on proximity between them. A higher score implies a greater similarity. Quantifying similarity by developing a suitable similarity measure is often a difficult task. Currently, there are many measures that can be used for quantifying the similarity [1, 3, 4]. These measures are classified into three different groups in this paper: distance measures, proximity measures, and binary measures. There are translations like amplitude-scale, amplitude-shift, time-shift, time-scale and phase delay. Chi-square similarity measure supports the translations needed to compare the feature vectors of the proposed algorithm:

$$\mathrm{x}^2(\mathrm{S,M}) = \sum\nolimits_i \frac{(Si-Mi)^\wedge 2}{Si+Mi} \tag{3}$$

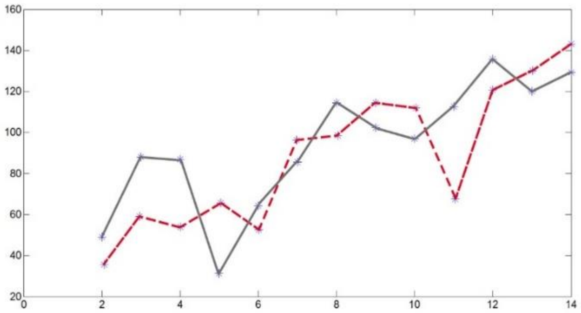

**Figure 14.** Vector of thetas of one of the mouth fiducial points across frames of video. One Subject is drawn as a dashed line and the other is drawn as a straight line

in which S and M are two feature vectors. Note that some local regions of face is more informative than others in facial expression classification. For example, the focus of feature vector for facial expression is mainly on eyes and mouth. Therefore, all subregions are weighted, due to the associated significance. The weighted $X^2$ is defined as follows:

$$x^2(S, M) = \sum_{i,j} wj \frac{(Sij-Mij)\text{^}2}{Sij+Mij} \quad (4)$$

in which S and M are two feature vectors and Wj is the associated weight to region j.

## 4. RESULTS

Since the focus of the paper is on pain facial expression, there is only one database in the literature that satisfies criteria and that is STOIC database. The STOIC database was developed and validated by Roy et al. from the Université de Montréal [4]. The STOIC database is a dynamic facial expression database validated by human observers.

Lots of efforts have been done in generating databases containing six basic facial expressions (happiness, surprise, disgust, anger, fear and sadness) but still lack of databases in pain facial expression. In this paper, performance of proposed method on STOIC database is evaluated.

STOIC database includes 80 video files from 5 men and 5 women in 8 different facial expressions including pain, happiness, surprise, disgust, anger, fear, sadness and neutral. Facial expressions in the database are acted facial expressions. Size of frames of the video is 256×256 pixels. Each video file starts with neutral facial expression and face gradually changes to the asked facial expression. Some sample frames of the database are shown in Figure 15.

As explained in Section 3.3, vectors of displacement directions should be compared with some pivotal data to represent how much the vectors are deviated from pivotal data. Pivotal data consists of 8 sequences of thetas for each of the facial expressions (pain, neutral, happiness, surprise, disgust, anger, fear and sadness); each of the sequences is related to a specific point.

Each sequence of pivotal data is obtained by taking the average of corresponding theta vectors of three male subjects and three female subjects for each certain facial expression. Therefore, six subjects formed our training set. Results of applying the proposed method on test set is brought to you in a table format (see Tables 1 and 2).

Results show that only one subject (DF5pa) is mistakenly reported as "Happy", while it is actually expressing the pain facial expression. It should be noted that this database is complicated and hard to recognize because of its generality in facial expressions. Although,

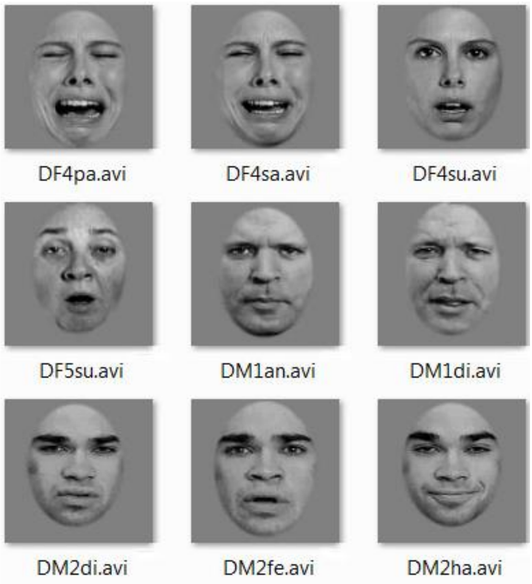

**Figure 15.** Extracted sample frames from video files of STOIC database

**TABLE 1.** Results of neutral facial expression

| Video Filename | Expression | Gender | Validation |
|---|---|---|---|
| **DM1ne** | Neutral | Male | True |
| **DM2ne** | Neutral | Male | True |
| **DM3ne** | Neutral | Male | True |
| **DM4ne** | Neutral | Male | True |
| **DM5ne** | Pain | Male | False |
| **DF1ne** | Neutral | Female | True |
| **DF2ne** | Neutral | Female | True |
| **DF3ne** | Neutral | Female | True |
| **DF4ne** | Neutral | Female | True |
| **DF5ne** | Neutral | Female | True |

**TABLE 2.** Results of pain facial expression

| Video Filename | Expression | Gender | Validation |
|---|---|---|---|
| DM1pa | Pain | Male | True |
| DM2pa | Pain | Male | True |
| DM3pa | Pain | Male | True |
| DM4pa | Pain | Male | True |
| DM5pa | Pain | Male | True |
| DF1pa | Pain | Female | True |
| DF2pa | Pain | Female | True |
| DF3pa | Pain | Female | True |
| DF4pa | Pain | Female | True |
| DF5pa | Happy | Female | False |

the accuracy is achieved 90% in both neutral and pain facial expressions.

A comparative study with the proposed method is the study of Hammal et al. [1] which is based on a dynamic fusion process of facial and context information for the automatic classification of facial expressions. They evaluated their proposed model on STOIC database and achieved 84.5% accuracy on the database. There are two reasons which explains the difference in comparative results. The first reason is that using spatio-temporal LBPs captures the swarm modifications of pixels in a region. The technique does not depend on tracking a set of limited points, but it senses the modification of pixels in a region. The other reason is that we set the threshold radius of pain facial expression to 23 and set the threshold radius of other seven facial expression to 19.

## 5. CONCLUSION

In this paper, a novel method is presented with a low time complexity but yet a powerful method for capturing a face in real-time mode and analyze the facial expression in a clinical context application. The focus of the paper is mainly on pain facial expression. In the proposed method, facial features are first extracted and consequently fiducial points out of facial feature regions are extracted. Fiducial points are tracked using the Lucas-Kanade algorithm in different frames points. The displacement directions of fiducial points creates a feature vector. Displacement directions are then compared with pivotal data. The pivotal data is obtained using Chi-square similarity measure on training set data and test data. To boost the proposed method, spatio-temporal Local Binary Patterns is applied to eyes and eyebrows. The proposed method is tested on STOIC database. STOIC is the only database that meets the requirements. Experimental results showed that leveraging spatio-temporal LBPs improves the obtained accuracy by 12% on STOIC database. Due to lack of database for pain facial expression, it is recommended to work on creating a database containing pain facial expression from different views and from more subjects.

## 6. REFERENCES

1. Hammal, Z. and Kunz, M., "Pain monitoring: A dynamic and context-sensitive system", ***Pattern Recognition***, Vol. 45, No. 4, (2012), 1265-1280.
2. Littlewort, G.C., Bartlett, M.S. and Lee, K., "Faces of pain: Automated measurement of spontaneousallfacial expressions of genuine and posed pain", in Proceedings of the 9th international conference on Multimodal interfaces. (2007), 15-21.
3. Monwar, M.M. and Rezaei, S., "Pain recognition using artificial neural network", in 2006 IEEE International Symposium on Signal Processing and Information Technology, IEEE., (2006), 28-33.
4. Y.-I., T., T., K. and J.F., C., "Recognizing action units for facial expression analysis", ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, Vol. 23, (2001), 97-115.
5. Tian, Y.-l., "Evaluation of face resolution for expression analysis", in 2004 Conference on Computer Vision and Pattern Recognition Workshop, IEEE. (2004), 82-82.
6. Bartlett, M.S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I. and Movellan, J., "Recognizing facial expression: Machine learning and application to spontaneous behavior", in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), IEEE. Vol. 2, (2005), 568-573.
7. Ira, C., Nicu, S., Ashutosh, G., Lawrence S., C. and Thomas S., H., "Facial expression recognition from video sequences: Temporal and static modeling", ***Computer Vision and Image Understanding***, Vol. 91, No., (2003), 160-187.
8. M, P. and L.J.M, R., "Expert system for automatic analysis of facial expressions", ***Image and Vision Computing***, Vol. 18, (2000), 881-905.
9. M., P. and L.J.M., R., "Facial action recognition for facial expression analysis from static face images", ***IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)***, Vol. 34, (2004), 1449-1461.
10. M., P. and I., P., "Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences", ***IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)***, Vol. 36, (2006), 433-449.
11. Y., Z., L.C., D.S. and C.C., K., "Using moment invariants and hmm in facial expression recognition", ***Pattern Recognition Letters***, Vol. 23, (2002), 83-91.
12. Bumhwi, K., Sang-Woo, B. and Minho, L., Improving adaboost based face detection using face-color preferable selective attention. 2008, Springer Berlin Heidelberg.88-95.
13. Irene, K., Stefanos, Z. and Ioannis, P., "Texture and shape information fusion for facial expression and facial action unit recognition", ***Pattern Recognition***, Vol. 41, (2008), 833-851.
14. Caifeng, S., Shaogang, G. and Peter W., M., "Facial expression recognition based on local binary patterns: A comprehensive study", ***Image and Vision Computing***, Vol. 27, (2009), 803-816.
15. Gwen C., L., Marian Stewart, B. and Kang, L., "Automatic coding of facial expressions displayed during posed and genuine pain", ***Image and Vision Computing***, Vol. 27, (2009), 1797-1803.
16. Zahedi, M. and Firouzian, I., "Pain facial expression recognition among 8 facial expressions using similarity measures", in Machine Vision and Image Processing (MVIP), Zanjan, Iran., (2013).
17. Y., Y. and L.S., D., "Recognizing human facial expressions from long image sequences using optical flow", ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, Vol. 18, (1996), 636-642.
18. I.A., E. and A.P., P., "Coding, analysis, interpretation, and recognition of facial expressions", ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, Vol. 19, (1997), 757-763.
19. Black, M.J. and Yacoob, Y., "Recognizing facial expressions in image sequences using local parameterized models of image motion", ***International Journal of Computer Vision***, Vol. 25, No. 1, (1997), 23-48.
20. Essa, I.A. and Pentland, A.P., "Coding, analysis, interpretation, and recognition of facial expressions", ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, Vol. 19, No. 7, (1997), 757-763.

21. M., Y., B., B. and R., S., "From facial expression to level of interest: A spatio-temporal approach, IEEE.
22. J., H. and J.J., L., "Value directed learning of gestures and facial displays, IEEE.
23. R., E.K. and P., R., "Real-time inference of complex mental states from facial expressions and head gestures, IEEE.
24. , Y.Z. and , Q.J., "Active and dynamic information fusion for facial expression understanding from image sequences", ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, Vol. 27, No., (2005), 699-714.
25. Firouzian, I. and Firouzian, N., "Face recognition by cognitive discriminant features", ***International Journal of Nonlinear Analysis and Applications***, Vol. 11, No. 1, (2020), 7-20.
26. Chang, Y., Hu, C. and Turk, M., "Probabilistic expression analysis on manifolds", in Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., IEEE. Vol. 2, (2004), II-II.
27. José M., B., Enrique, M. and Luis, B., "Recognising facial expressions in video sequences", ***Pattern Analysis and Applications***, Vol. 11, (2007), 101-116.
28. Guoying, Z. and Matti, P., "Boosted multi-resolution spatiotemporal descriptors for facial expression recognition", ***Pattern Recognition Letters***, Vol. 30, (2009), 1117-1127.
29. Zheng, Z., Zheng, Z. and Tiantian, Y., Expression recognition based on multi-scale block local gabor binary patterns with dichotomy-dependent weights. 2009, Springer Berlin Heidelberg.895-903.
30. Stephen, M. and Richard, B., "The effects of pose on facial expression recognition, British Machine Vision Association. (2009).
31. Lucey, P., Cohn, J.F., Prkachin, K.M., Solomon, P.E., Chew, S. and Matthews, I., "Painful monitoring: Automatic pain monitoring using the unbc-mcmaster shoulder pain expression archive database", ***Image and Vision Computing***, Vol. 30, No. 3, (2012), 197-205.
32. Werner, P., Al-Hamadi, A., Niese, R., Walter, S., Gruss, S. and Traue, H.C., "Towards pain monitoring: Facial expression, head pose, a new database, an automatic system and remaining challenges", in Proceedings of the British Machine Vision Conference. (2013), 1-13.
33. Hashemi, S.M.R. and Faridpour, M., "Evaluation of the algorithms of face identification", in 2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI), IEEE. (2015), 1049-1052.
34. Bartlett, M.S., Littlewort, G.C., Frank, M.G. and Lee, K., "Automatic decoding of facial movements reveals deceptive pain expressions", ***Current Biology***, Vol. 24, No. 7, (2014), 738-743.
35. Werner, P., Al-Hamadi, A. and Niese, R., "Comparative learning applied to intensity rating of facial expressions of pain", ***International Journal of Pattern Recognition and Artificial Intelligence***, Vol. 28, No. 05, (2014), 1451008.
36. Sikka, K., Ahmed, A.A., Diaz, D., Goodwin, M.S., Craig, K.D., Bartlett, M.S. and Huang, J.S., "Automated assessment of children's postoperative pain using computer vision", ***Pediatrics***, Vol. 136, No. 1, (2015), e124-e131.
37. Liu, D., Peng, F., Shea, A. and Picard, R., "Deepfacelift: Interpretable personalized models for automatic estimation of self-reported pain", ***arXiv preprint arXiv:1708.04670***, (2017).
38. Gruss, S., Geiger, M., Werner, P., Wilhelm, O., Traue, H.C., Al-Hamadi, A. and Walter, S., "Multi-modal signals for analyzing pain responses to thermal and electrical stimuli", ***JoVE (Journal of Visualized Experiments)***, No. 146, (2019), e59057.
39. O'Neill, M.C., Ahola Kohut, S., Pillai Riddell, R. and Oster, H., "Age-related differences in the acute pain facial expression during infancy", ***European Journal of Pain***, Vol. 23, No. 9, (2019), 1596-1607.
40. Virrey, R.A., Liyanage, C.D.S., Petra, M.I.b.P.H. and Abas, P.E., "Visual data of facial expressions for automatic pain detection", ***Journal of Visual Communication and Image Representation***, Vol. 61, (2019), 209-217.
41. Zamzmi, G., Paul, R., Goldgof, D., Kasturi, R. and Sun, Y., "Pain assessment from facial expression: Neonatal convolutional neural network (n-cnn)", in 2019 International Joint Conference on Neural Networks (IJCNN), IEEE. (2019), 1-7.
42. Roy, S., Roy, C., Éthier-Majcher, C., Fortin, I., Belin, P. and Gosselin, F., "Stoic: A database of dynamic and static faces expressing highly recognizable emotions", ***J. Vis***, Vol. 7, (2007), 944.
43. Lucas, B.D. and Kanade, T., "An iterative image registration technique with an application to stereo vision", (1981).
44. Pietikäinen, M., Hadid, A., Zhao, G. and Ahonen, T., Local binary patterns for still images, in Computer vision using local binary patterns. 2011, Springer.13-47.

## Persian Abstract

### چکیده

نظارت مداوم حالت چهره بیماران در محیط های بیمارستانی، علاوه بر نظارت علائم حیاتی بیماران، یک ضرورت محسوب می‌شود. نظارت بیماران از طریق تشخیص حالت چهره از توالی‌های ویدئویی، نیاز به حضور همراه بیمار را از میان می‌برد. در این مقاله، یک رویکرد نوینی برای نظارت حالت چهره ارائه می شود که در صورت احساس درد در چهره بیمار، هشدار می دهد. رویکرد پیشنهادی این مقاله از طریق رهگیری نقاط محوری در چهره درون فایل های ویدئویی و همچنین از الگوهای دودویی محلی فضایی-زمانی برای بخش های چشم‌ها و ابروها بهره می‌برد. زوایای حرکت نقاط محوری در چهره در تصاویر ویدئویی و همچنین هیستوگرام‌های الگوهای دودویی محلی فضایی-زمانی، بردار ویژگی نهایی را می‌سازند. بردارهای ویژگی توسط معیار شباهت ***Chi-square*** درنهایت با یکدیگر مقایسه می‌شوند. نتایج آزمایشات نشان می دهد که بکارگیری الگوهای دودویی محلی فضایی-زمانی بهبود، به متد رهگیری نقاط محوری، موجب بهبود ۱۲ درصدی می‌شود.