

以樣板圖樣為基礎之人臉表情辨識

賴智錦 林致圻 吳信融 林伯豪

國立高雄大學電機工程學系

cclai@nuk.edu.tw, {ssn4010302, wux4030, borhaur}@gmail.com

摘要

自動化的人臉表情分析是一項極具挑戰性的研究問題，而且該議題對人機互動、機器人控制、駕駛者的狀態監控、與資料驅動動畫等諸多應用有著相當重要的影響。自動化人臉表情辨識主要是擷取人臉表情影像中較具有辨識能力的特徵，再透過分類器進行分類以判定該影像中的表情為何。本論文提出一個植基於臉部紋理與邊緣資訊的人臉表情辨識方法，該法是找出臉部適合表示表情的特徵，再計算出影像中各像素點的特徵值，最後利用支持向量機進行這些特徵的分類，藉此判定該影像所呈現的表情。實驗結果顯示，本論文的方法可以獲得不錯的辨識效果。

關鍵詞：人臉表情辨識；臉部特徵資訊；支持向量機。

1. 前言

人臉的特徵是十分複雜且變化多端的，人與人之間的交流，表情對於訊息的傳達是非常重要的。隨著科技不斷進步，人們需要以臉部的表情與機器進行溝通以達到一些目的，例如：影像照護系統，就是以臉部表情變化判斷病患的生理是否發生異常狀況？或是機器人視覺系統，則是透過臉部表情判斷使用者的生理、心理等狀況，再進行相對應的處理程序。

現今許多科技產品強調具備人工智慧，增強了機器與人類之間的互動；若能使電腦自動辨別人類的情緒反應，更可拉近電腦與人類的關係，於是臉部表情辨識(facial expression recognition, FER)的研究在學術界越來越受到重視。透過分析臉部表情的不同，將各種表情特徵量化並分類，就可以使電腦學習如何辨別人類的情緒狀態，進一步提供資訊給電腦系統做出該有的對應與互動。

若以臉部影像特徵作為電腦辨識表情的依據，目前主要有兩種特徵擷取方式。第一種

方式是基於幾何特徵擷取[1-6]，以臉部動作編碼系統(Facial Action Coding System, FACS)[7]為導向，該編碼系統可以用來描述臉部特定區域個別移動的狀況；例如：Kotsia and Pitas[6]提出人臉表情的變化可利用網格狀模型進行參數化設定，網格狀模型主要由許多節點所組成，這些節點的選定即是以 FACS 中的動作單元(Action Units)來決定。由於需先將臉部特徵手動標示節點作為參考基準，因此這些節點的位置將決定辨識結果的優劣。第二種是基於外觀特徵擷取[8-12]，主要是針對區域紋理進行處理；例如：Bashyal and Venayagamoorthy [12]利用 Gabor 小波轉換擷取肌肉紋理，手動選擇重要的特徵部位，最後以學習向量量化(Learning Vector Quantization, LVQ)進行表情分類，雖然此法準確度高，但計算過程複雜，需耗費較多時間。

基於外觀特徵的方法主要依據表情的紋理及像素值上的差異來進行分析。Zhang 等人[13]證明了 Gabor 小波表示法在表情辨識上比使用幾何特徵的方法效果更佳。Donato 等人[14]在其研究中，則探討了主成分分析(Principal Component Analysis, PCA)、獨立成分分析(Independent Component Analysis, ICA)、區域特徵分析(Local Feature Analysis, LFA)、線性判別分析(Linear Discriminant Analysis, LDA)、與 Gabor 小波表示法對於臉部表情辨識的優劣，結果顯示 Gabor 小波表示法與 ICA 的效能最佳。為了解決 Gabor 小波表示法在計算上需要花費許多時間與儲存空間的問題，區域二元圖樣(Local Binary Patterns, LBP)逐漸被使用在影像資訊的表達上[15]。Viola and Jones 則證實，LBP 對於光線的變化及低解析度影像有較高的強韌性，其辨識效果更優於 Gabor 小波表示法[16]。

在本論文中，我們提出一個方法將表情影像透過雙向區域樣板圖樣(Bidirectional Local Template Pattern, B-LTP)[17]計算出各個像素點的特徵值後，藉此表示為人臉表情影像之特徵，再以支持向量機(Support Vector Machine,

SVM)進行這些特徵之辨識。為了驗證所提方法的效能，我們將使用 Cohn-Kanade (CK) [18] 和 Extended Cohn-Kanade (CK+) [19]兩個表情影像資料庫作為測試對象。實驗結果顯示，利用本文所提的方法，可得到不錯的辨識效果。

本論文架構如下：第二節回顧本論文所使用之表情辨識相關技術，第三節介紹本論文所提出的方法，第四節是相關實驗結果與討論，第五節為結論與未來研究方向。

2. 背景介紹

2.1 雙向區域模板圖樣

受到樣板直方圖(histogram of template, HOT)與中心對稱區域二元圖樣(center-symmetric local binary patterns, CS-LBP)的觀念啟發，Xu 等人[17]提出了雙向區域樣板圖樣(bidirectional local template patterns, B-LTP)的觀念，並已成功地應用在行人偵測的研究主題。

B-LTP 主要由四個基本樣板所構成，這些樣板主要是考慮在某一區域裡，中心點像素和對角像素點之間的紋理資訊和梯度資訊的差異；因此，可視為是影像中某種區域結構特徵的表現。此外，這些樣板亦可反應影像裡的邊緣資訊，與突出區域的紋理資訊。有此四個基本樣板，使用者便可以定義出諸多的數學公式，藉此分別用於計算影像中的梯度與紋理資訊。此四個樣板的範例，如圖 1 所示。

針對影像中紋理的部分，B-LTP 可以下述公式予以定義：

$$B-LTP_1 = S(g(P_i), g(P), g(P_{i+4})) \times 2^i, \quad (1)$$

$$S(g(P_i), g(P), g(P_{i+4})) = \begin{cases} 1, & (g(P_i) \geq g(P) \cap g(P) \geq g(P_{i+4})) \cup \\ & (g(P_{i+4}) \geq g(P) \cap g(P) \geq g(P_i)) \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

$$B-LTP_2 = g(P) \geq g(P_i) \cap g(P) \geq g(P_{i+4}), \quad (3)$$

$$B-LTP_3 = \arg \max_k \{g(P)_k + g(P_i)_k + g(P_{i+4})_k\}, \quad (4)$$

其中， $g(P)$ 表示像素 P 的灰階值。至於影像中梯度大小的資訊，則以下述公式考慮之：

$$B-LTP_4 = S(Mag(P_i), Mag(P), Mag(P_{i+4})) \times 2^i, \quad (5)$$

$$B-LTP_5 = Mag(P) \geq Mag(P_i) \cap g(P) \geq g(P_{i+4}), \quad (6)$$

$$B-LTP_6 = \arg \max_k \{Mag(P)_k + Mag(P_i)_k + Mag(P_{i+4})_k\}, \quad (7)$$

其中， $Mag(P)$ 表示像素 P 的梯度大小。

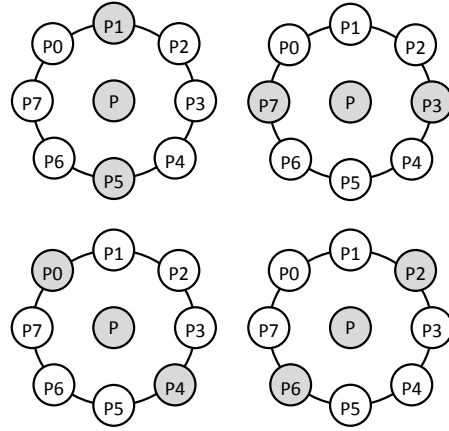


圖 1. B-LTP 的四個基本樣板[17]，其中 P 為影像中某一像素點， $P_i (i = 0, \dots, 7)$ 為其八鄰點。

以此方式，對於影像中紋理資訊的部分，我們可以得到一個維度大小是 24 直方圖；其中，前 16 個數值是由 $B-LTP_1$ 的數值決定，接著 4 個數值則是由 $B-LTP_2$ 的條件決定，最後的 4 個數值則是由 $B-LTP_3$ 決定。梯度大小的計算方式與紋理的部分雷同，所以亦可產生維度同為 24 的直方圖。最後將梯度大小的部分和紋理部分的直方圖串接在一起，可得到維度是 48 的直方圖。然而，在文獻[17]中，作者並未對梯度大小的計算提供詳細的說明或採行的方法；因此，在本論文中，我們將以 Laplacian 運算子[20]作為計算影像中邊緣資訊的依據。

2.2 Laplacian 運算子

Laplacian 運算子的概念源自於在一張影像裡，相同顏色的區域內，像素之間的灰階值不會有太大變化；然而，在兩不同顏色的邊緣交接處，其灰階值就會顯示出大的變化，這種大幅度的變化就是我們用以檢測邊緣的重要依據。

對於擁有雙變數的函數 $f(x, y)$ 的 Laplacian 運算子，其定義為：

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \quad (8)$$

為了能夠將其應用在影像處理上，此方程式需以離散形式表達。假設現有兩個變數，則在 x 方向及 y 方向上的二階偏微分可以下方公式表示：

$$\frac{\partial^2 f}{\partial x^2} = f(x+1, y) + f(x-1, y) - 2f(x, y) \quad (9)$$

$$\frac{\partial^2 f}{\partial y^2} = f(x, y+1) + f(x, y-1) - 2f(x, y) \quad (10)$$

上述公式(8)的二維 Laplacian 運算子可由公式(9)和(10)進行相加而得：

$$\begin{aligned} \nabla^2 f = & [f(x+1, y) + f(x-1, y) \\ & + f(x, y+1) + f(x, y-1)] \\ & - 4f(x, y) \end{aligned} \quad (11)$$

2.3 區域紋理特徵的擷取

為了有效地表示整張人臉影像中細微紋理的組成，本論文採取區塊式方法建構區域紋理特徵。所謂區塊式特徵的擷取與建構是先將人臉影像分割成固定大小且不重疊的子區塊，針對每個子區塊進行 B-LTP 特徵的計算與相關直方圖的建構，最後將全部子區塊的直方圖串接在一起，形成單一且代表該張臉部表情影像的特徵直方圖，其觀念如圖 2 所示。

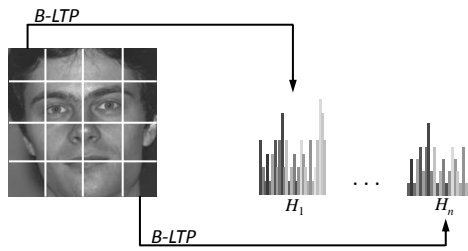


圖 2. 局部紋理特徵擷取示意圖

3. 研究方法

本論文所提的臉部表情辨識方法，乃是利用上述理論擷取臉部具有代表性之紋理特徵，再將這些特徵送交 SVM，以進行各種表情的辨識。

首先我們需將臉部影像分別進行區塊式 B-LTP 的運算。為了求得區域代表性之臉部紋理，在進行 B-LTP 的運算時，我們會把影像切割成 $N \times N$ 個不重疊的子區塊以進行特徵直方圖的運算。最後將各區塊所求取之區塊式 B-LTP 特徵直方圖串接一起，形成一個延伸特徵直方圖，我們以此作為輸入臉部表情影像之特徵向量。

圖 3 是本論文所提方法之系統流程圖，可分為訓練與測試兩部分。在訓練部分，每一張

影像經上述方法提取相關特徵後，利用 SVM 將欲訓練之數種表情樣本進行分類訓練，可獲得一個訓練好的 SVM 模型。在測試部分，將欲測試之影像經上述方法提取特徵後送入訓練好的 SVM 模型，進行預測並獲得辨識結果。

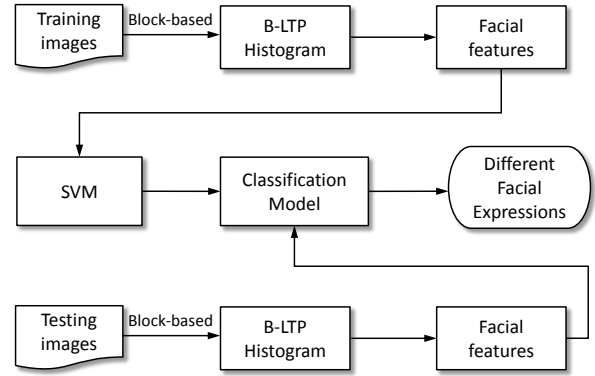


圖 3. 本文所提臉部表情辨識系統流程圖

4. 實驗結果

為了驗證本論文所提方法之效能，我們選擇在表情辨識研究上廣為研究者使用的 Cohn-Kanade (CK) 資料庫和 Extended Cohn-Kanade (CK+) 資料庫進行相關實驗。

Cohn-Kanade 資料庫目前已被臉部表情研究領域中的學者廣泛採用，此資料庫主要由 97 位年齡介於 18 歲到 30 歲之間的大學生組成，其中 65% 為女性、15% 為美裔非洲人、3% 為亞洲人或拉丁美洲人，共 486 組序列影像。每個人都有二到七種不同的表情(分別是：自然表情(NE)、高興(HA)、恐懼(FE)、厭惡(DI)、驚訝(SU)、悲傷(SA)、憤怒(AN)等)，每組影像序列皆由自然表情(或稱無表情)開始，並以其中一種表情做為結束。我們從資料庫共 486 組影像序列中，選取其中 320 段序列影像，再從每一段影像序列中，選其第一張作為無表情影像，最後三張作為表情影像。

Lucey 等人[19]於 2010 年將原始 CK 資料庫進行擴增，建置了 Extended Cohn-Kanade Facial Expression Database (CK+)。我們使用 CK+ 資料庫中的表情影像部份，是由 123 位不同的人所拍攝之 593 組序列影像所組成，年齡介於 20 到 45 歲，所有序列影像皆是由無表情開始，並以某一種表情結束。由於有部分序列影像如同 CK 資料庫一般有不完整之情形，因此我們從 593 組序列影像中選出 395 組完整的序列影像，將每段序列影像的第一張影像視作

無表情、最後三張作為某類表情影像。

本實驗所使用到 CK 與 CK+ 資料庫，各類表情的影像數量如表 1 所示。

表 1. CK 與 CK+ 之實驗用影像張數

資料庫	表情種類						
	AN	DI	FE	HA	SA	SU	NE
CK	108	120	99	282	126	225	320
CK+	159	156	174	267	183	246	395

本論文採取 10 倍交叉驗證法 (10-fold cross-validation) 進行試驗。10 倍交叉驗證法是將資料集隨機分成十等分，每次選取其中九等分當作訓練樣本，剩餘一等分當作測試樣本，並且重複上述動作 10 次，直到每等分的資料都曾被選為測試樣本後結束。我們將所有表情影像樣本隨機分成十等分，每等分皆有七類表情樣本，再以 10 倍交叉驗證法進行訓練及測試，所求取的平均數值稱為該法的表情辨識率。

4.1 實驗一

為了瞭解使用區域紋理特徵擷取時，區塊數目對於辨識結果的影響，在本實驗我們以 CK 資料庫為測試對象，實驗不同的區塊數目並觀察其辨識效能。相關數據如表 2 所示。

表 2. 不同區塊數目之辨識效能(%) - 以 CK 為例

區塊數目	2×2	3×3	4×4	5×5	7×7	9×9
辨識率	92.50	96.64	96.72	97.73	97.73	97.89

從表 2 的數據中可以發現，隨著區塊數目的增加，可以獲得更多有用的紋理資訊，進而達到提升辨識的效果。當區塊數目為 9 × 9 時，可以得到最高的表情辨識率；因此，我們以其為基準，觀察其七種表情的混淆矩陣，相關數據如表 3 所示。

在表 3 的實驗數據顯示，我們的方法在厭惡與高興兩種表情的辨識結果最佳；其次在憤怒、驚訝、與無表情的辨識上，亦可獲得相當不錯的辨識效果。唯獨在悲傷表情的辨識率，會略低於 95%。由於悲傷表情是屬於五官中較無明顯變化的狀況，因此，有略高的比率會將其誤判為自然表情。從此現象推論，若使用機

器進行臉部影像的表情辨識時，如果輸入的表情特徵不夠清楚或明顯時，將會造成一定程度的誤判情形。

表 3. 七種表情辨識率(%)之混淆矩陣 - 以 CK 為例

	AN	DI	FE	HA	SA	SU	NE
AN	96.30	0	0	0	0	0	3.70
DI	0	99.17	0	0	0	0	0.83
FE	2.02	0	97.98	0	0	0	0
HA	0	0	0	99.65	0	0	0.35
SA	0	0	0	0	92.86	0	7.14
SU	0	0	1.33	0	0	98.22	0.44
NE	0.94	0	0	0.63	0.31	0	98.12

4.2 實驗二

為了得知我們的方法與其他方法在效能上的差異，我們實作文獻上以區域紋理為基準或變形之方法，並以 CK 資料庫中七類表情影像進行實驗，結果如表 4 所示。實驗數據顯示，我們的方法確實比其他現有的方法，可以獲得更令人滿意的辨識效果。

表 4. 本論文與其他方法之辨識率比較

方法	辨識率(%)
LBP [15]	88.9
LDP [21]	93.4
LDPv [22]	93.1
Ours	97.9

4.3 實驗三

除了以 CK 為實驗對象之外，我們還以 CK+ 作為測試的對象，以驗證當影像張數增加或取相對象的不同時，我們的方法是否依舊可以獲得不錯的辨識效果。如同實驗一的目的，我們先驗證不同的區塊數目，對於辨識效果的影響性，實驗數據如表 5 所示。

表 5. 不同區塊數目之辨識效能(%) - 以 CK+ 為例

區塊數目	2×2	3×3	4×4	5×5	7×7	9×9
辨識率	88.99	95.76	97.78	97.85	98.10	98.86

從表 5 的數據中我們可以發現，隨著區塊數目的增加，由各區塊內的紋理資訊進行組合後，將可獲得更具識別性的紋理特徵，進而提

升辨識的效果。當區塊數目為 9×9 時，可以得到最高的表情辨識率；因此，我們以其為基準，觀察其七種表情的混淆矩陣，相關數據如表 6 示。

在表 6 的實驗數據顯示，我們的方法在恐懼與開心的表情辨識結果最佳；至於在其他表情的辨識，亦可獲得相當不錯的辨識效果。由此可知，隨著各類表情影像的數量增加，可獲取更多的分類資訊；對於分類效果的影響，有正面的幫助。

表 6. 七種表情辨識率(%)之混淆矩陣 - 以 CK+ 為例

	AN	DI	FE	HA	SA	SU	NE
AN	96.86	0.63	0	0	0	0	2.51
DI	0.64	99.36	0	0	0	0	0
FE	0	0	100	0	0	0	0
HA	0	0	0	100	0	0	0
SA	1.64	0	0	0	96.72	0	1.64
SU	0	0	0	0	0	99.19	0.81
NE	0	0	0	0	1.01	0	98.99

4.4 實驗四

為了得知我們的方法與其他方法在效能上的差異，我們實作文獻上以區域紋理為基準或變形之方法，並以 CK+ 資料庫中七類表情影像進行比較，結果如表 7 所示。實驗數據顯示，我們的方法確實比其他現有的方法，可以獲得更令人滿意的辨識效果

表 7. 本論文與其他方法之辨識率比較

方法	辨識率(%)
LGP [23]	90.2
PHOG [24]	95.3
Census Transform [25]	96.8
Ours	98.9

5. 結論

在進行臉部表情辨識時，不僅是要提升辨識率，如何有效縮短機器運算的時間、在不降低辨識率的條件下減少特徵維度的等，都是一個良好的表情辨識方法需要考慮的因素。本論文提出的方法是將表情影像透過區塊式雙向區域樣板圖樣計算出各個像素點的特徵值後，並將其所表示之直方圖當作擷取出來的特徵，形成一個特徵直方圖，再以此作為該張表情影像之特徵資訊。我們的方法是將影像分割成許多區塊以提高表情的辨識效能，由於本文

所提之方法是以整張人臉影像進行特徵值擷取，某些較不具辨識效果的區塊亦會一併進行計算，未來將朝向將人臉中較具有鑑別度的部分，例如：眼睛四周、嘴角、或眉毛等區域選取出來，再進行特徵值擷取的動作，以達到提高辨識效果的目的。

致謝

本研究承蒙行政院科技部計畫補助，特此致謝。研究計畫編號：MOST 104-2221-E-390-018-MY2。

參考文獻

- [1] M. Pantic and L. Rothkrantz, "Expert system for automatic analysis of facial expression," *Image and Vision Computing*, vol. 18, no. 11, pp. 881-905, 2000.
- [2] Y. Tian, T. Kanade, and J. Cohn, "Recognizing action units for facial expression analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 97-115, Feb. 2001.
- [3] R. E. Kaliouby and P. Robinson, "Real-time inference of complex mental states from facial expressions and head gestures," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2005, pp. 181-200.
- [4] M. Pantic and L. J. M. Rothkrantz, "Facial action recognition for facial expression analysis from static face images," *IEEE Trans. Systems, Man, and Cybernetics -- Part B: Cybernetics*, vol. 34, no. 3, pp. 1449-1461, 2004.
- [5] M. Pantic and I. Patras, "Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences," *IEEE Trans. Systems, Man, and Cybernetics -- Part B: Cybernetics*, vol. 36, no. 2, pp. 433-449, 2006.
- [6] I. Kotsia and I. Pitas, "Using geometric deformation features and support vector machines," *IEEE Trans. Image Processing*, vol. 16, no. 1, pp. 172-187, 2007.
- [7] P. Ekman and W. V. Friesen, *Facial Action Coding System*, Consulting Psychologists Press, College Avenue, Palo Alto, CA, 1978.
- [8] Z. Zhang, M. J. Lyons, M. Schuster and S. Akamatsu, "Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using

- multi-layer perceptron," in *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*, 1998, pp. 454-459.
- [9] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski, "Classifying facial actions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974-989, 1999.
- [10] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: machine learning and application to spontaneous behavior," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2005, vol. 2, pp. 568-573.
- [11] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Trans. Neural Networks*, vol. 13, no. 6, pp. 1450-1464, 2002.
- [12] S. Bashyal and G. K. Venayagamoorthy, "Recognition of facial expressions using Gabor wavelets and learning vector quantization," *Engineering Applications of Artificial Intelligence*, vol. 21, no. 7, pp. 1056-1064, 2008.
- [13] Z. Zhang, M. J. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron," in *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*, 1998, pp. 454-459.
- [14] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski, "Classifying facial actions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974-989, 1999.
- [15] C. Shan, S. Gong, and P. W. McOwan, "Robust facial expression recognition using local binary patterns," in *Proc. IEEE Int'l Conference Image Processing*, 2005, vol. 2, pp. 370-373.
- [16] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137-154, 2004.
- [17] J. Xu, N. Jiang, and S. Goto, "Pedestrian detection based on bidirectional local template patterns," in *Proc. 20th European Signal Processing Conference*, 2012, pp. 400-404.
- [18] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*, 2000, pp. 46-53.
- [19] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn-Kanade dataset (CK+): A complete facial expression dataset for action unit and emotion-specified expression," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshop*, 2010, pp. 94-101.
- [20] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, Addison-Wesley Publishing Company, 1992.
- [21] T. Jabid, Md. H. Kabir, and O. Chae, "Robust facial expression recognition based on local directional pattern," *ETRI Journal*, vol. 32, no.5, pp. 784-794, 2010.
- [22] Md. H. Kabir, T. Jabid, and O. Chae, "A local pattern variance (LDPv) based face descriptor for human facial expression recognition," in *Proc. 2010 7th IEEE Int'l Conf. Advanced Video and Signal Based Surveillance*, 2010, pp. 526-532.
- [23] M. S. Islam, "Local gradient pattern - A novel feature representation for facial expression recognition" *Journal of AI and Data Mining*, vol. 2, no .1, pp. 33-38, 2014.
- [24] R. A. Khan, A. Meyer, H. Konik, and S. Bouakaz, "Human vision inspired framework for facial expressions recognition," in *Proc. 19th IEEE Int'l Conf. Image Processing*, 2012, pp. 2593-2596.
- [25] S. E. Kahou, P. Froumenty, and C. Pal, "Facial expression analysis based on high dimensional binary features," in *Computer Vision – ECCV 2014 Workshops*, Part II, LNCS 8926, pp 135-147, 2015.