The Utility of Facial Analysis Algorithms in Detecting Melancholia

Matthew P. Hyett, Abhinav Dhall, and Gordon B. Parker

Abstract Facial expressions reliably reflect an individual's internal emotional state and form an important part of effective social interaction and communication. In clinical psychiatry, facial affect is routinely assessed, and any identified deviations from normal affective range and reactivity may signal the presence of a potential psychiatric disorder. An example is melancholic depression or 'melancholia' where facial immobility and non-reactivity are viewed as sensitive diagnostic indicators of the illness. However, affect in depressive disorders such as melancholia, and indeed psychiatric conditions more broadly, is largely assessed by clinicians, without biological or computational quantification. While such clinical assessment provides useful qualitative descriptors of illness features, the inherent subjectivity of this approach raises concerns regarding diagnostic reliability, and may hinder communication between clinicians. Methodological advances and algorithm development in the field of affective computing have the potential to overcome such limitations through objective characterization of facial features. Among these methods are implicit face analysis techniques, which are based on local spatio-temporal descriptors such as the space-time interest points and Bag-of-Words framework, and explicit face analysis techniques based on deformable model fitting methods such as Constrained Local Models and Active Appearance Models. In this chapter we overview these approaches and discuss their application toward detection and diagnosis of depressive disorders, in particular their capacity to delineate melancholia from the residual non-melancholic conditions.

M.P. Hyett (✉) • G.B. Parker
School of Psychiatry, University of New South Wales, Sydney, NSW, Australia 2052

Black Dog Institute, Prince of Wales Hospital, Randwick, NSW, Australia 2031
e-mail: m.hyett@unsw.edu.au

A. Dhall
Human-Centred Technology Research Centre, University of Canberra, University Drive, Bruce, Canberra, ACT, Australia 2617

Research School of Computer Science, Australian National University, Acton ACT, Australia 0200

M. Kawulok et al. (eds.), *Advances in Face Detection and Facial Image Analysis*,
DOI 10.1007/978-3-319-25958-1_13

1 Introduction

Clinical psychiatry has long utilized so-called mental state signs to assist in the diagnosis of mental disorders. In assessing the presence of a psychiatric disorder, particular attention is given to appearance, behavior, speech, thought, perception, insight into illness and, of key relevance to this chapter, mood and facial affect [1]. The latter contributes to delineating depressive illnesses such as melancholia from non-melancholic depression, and is typically assessed by clinicians (e.g., psychiatrists). However, such clinical observation is, by virtue of its non-standardized nature, subjective and thus may not reliably capture the presence or absence of specific disorders. As more objective means of quantifying facial affect and emotion emerge, driven largely by advances in affective computing, the diagnosis of disorders characterized by specific affective signs—such as melancholia—will be made more valid. Prior to examining the potential utility of these emerging technologies in depressive illness, we broadly overview conceptual models of emotion and depression classification, with an emphasis on the relevance of affect in such disorders.

2 Emotion and Affect

The subjective affective experience of an individual, variably referred to as one's 'passions', 'emotions', 'feelings', or 'moods' have, to this day, largely eluded definition. Over a century ago, James [2] suggested that emotions evoke a certain "phenomenal quality"; referring to the notion that emotions are "*sensation-like* mental states" [3], and were in effect seen as 'intuitions' in response to emotion-eliciting events. Others (e.g., [4]) challenged James's notion that emotions were reflex-like in nature (e.g., fear *in response to* a bear in the woods), and instead positioned emotion as an appraisal-based process (e.g., feeling happy or sad towards an object). Throughout the 1960s, appraisal theories of emotion dominated [5], but even to this day there continues to be an ongoing philosophical debate over what constitutes an emotional state [6]. The fundamental biological and psychological processes underlying the above theories of emotion generation and/or appraisal are referred to as the 'evolutionary core' [3]—that is, a set of discrete emotion mechanisms aligned to one's "basic" emotions.

There have been several dominant theories in the field of psychology that offer insight into such basic emotions. Beginning with McDougall [7], emotion was defined as a small set of adaptive processes, so-called emotional *instincts*, and included the behaviorally well-defined emotions of flight, repulsion, curiosity, pugnacity, self-abasement, self-assertion, and the parental instinct [8]. Most modern variants of discrete emotion theory correspond at least partially to McDougall's model. For instance, Tomkins [9–11], who is credited with the development of "affect theory", considered there to be nine independent affects: two positive

(enjoyment/joy and interest/excitement), one neutral (surprise/startle), and six negative affects (anger/rage, disgust, dissmell [similar to distaste], distress/anguish, fear/terror, and shame/humiliation). Likewise, Izard [12] considered fear, anger, shame, contempt, disgust, guilt, distress, interest, surprise, and joy to be primary emotions, and with various combinations of these giving rise to tertiary emotions (e.g., such as affection). While the basic emotions were initially believed to be largely subjective states of mind, Ekman [13] postulates that seven basic emotions correspond near-universally to observable facial expressions, namely anger, disgust, fear, happiness, sadness, surprise, and contempt. Such categorical definitions of emotion have received substantial empirical support, but there is also a question as to whether affect can be conceptualized dimensionally [14]; that is that emotions exist along interrelated continuums of, for example, arousal and valence [15].

It has been suggested that discrete (i.e., categorical) emotion theory provides a more accurate index of the current, momentary experience of emotion, while dimensional perspectives may be most relevant to temporal emotional experience, such as mood states [16]. Rather than being mutually exclusive, however, it is likely that both viewpoints are pertinent to affective states such as depression. The work of Ekman in particular [17], which highlights the importance of emotional expressions, is especially relevant for quantifying affect. Indeed, 'affect display' refers to the externally displayed (i.e., observable) affect of an individual, through facial, vocal or gestural means. When affect is in line with the subjective mood state of an individual, it is termed 'congruent affect', but when subjective states and affect is misaligned it is referred to as 'incongruent affect'. As will become apparent throughout this chapter, affect in depression is typically congruent with the individual's self-reported, subjective mood state; for instance, subjective low mood or sadness is often identifiable in the depressed patient through observation of a flat and/or non-reactive affect.

2.1 What Precisely Is an Affective Disorder?

Historical descriptions of disordered affective states date back over 2000 years. Hippocrates described the existence of "melancholia" as a disease-like state arising from an excess of black bile (this being one of four 'humors', or bodily fluids, that were thought to directly influence health and temperament). Current day formulations of melancholia position it as the prototypical depressive disease, principally of biological and genetic origin [18], which presents with characteristic clinical features such as psychomotor slowing (i.e., physical slowing, concentration impairment), anergia, anhedonia, diurnal mood variation (mood worse in the morning), early morning wakening, and appetite and weight loss. Despite evidence for the existence of melancholia as a distinct condition, it was largely abandoned in psychiatric circles in 1980 with the introduction of the Diagnostic and Statistical Manual of Mental Disorders (DSM), which popularized 'symptom-centric' models of affective illness. Its third edition (DSM-III) [19] brought three psychiatric

disorders under the diagnostic umbrella of the 'affective disorders', including major depression (subsuming melancholia), bipolar disorder (depression and/or hypomania or mania) and dysthymia (chronic low mood, defined as having fewer symptoms than major depression). Under the DSM framework, symptom reports by patients guide diagnostic decisions, which in isolation are thought to contribute to misdiagnosis, thus impacting on appropriate management [18]. Hence, while the affective disorders began with melancholia as observable illness states, they are now characterized by the severity of symptoms. It is argued that this has contributed to an era of insipidity in depression research, and is of limited clinical import. Here, we argue that there is scope for departure away from dimensional accounts of depression (e.g., major depression), back to more refined categorical depressive subtypes (e.g., melancholic vs. non-melancholic conditions).

3 The Case for Objectively Informed Classification of Depression Subtypes

The classification of depression has long been contentious, with numerous typologies being proposed since the beginning of the twentieth century. So-called 'simple typologies' conceptualized depression as comprising anywhere between one and five categories. The father of modern psychiatry, Emil Kraepelin, distinguished between *dementia praecox* (now schizophrenia) and manic-depressive insanity [20]. All major affective disturbances, namely mania and depression, were thought by Kraepelin to be part of the same illness (manic-depressive insanity). In 1926, British psychiatrist Edward Mapother at the Maudsley Hospital in London claimed [21] that there was only one form of depression, and did not distinguish between depressive subtypes—and hence saw depression as varying along a continuum. The influential theories of Kraepelin and Mapother laid the foundations for a move toward unitary models of depression. Sir Aubrey Lewis, an eminent psychiatrist at the same institute as Mapother, also viewed all depressions as essentially the same (and hence proposed a single category, "depressive illness"), but conceded that there are likely differences between individuals regarding its causation: specifically depressions that are more hereditary versus those caused by environmental factors [22].

The "separatists" opposed the unitary view of depression from the mid-twentieth century, arguing that depression could be classified into different types [23]. Such categorical views principally saw the diagnosis and classification of depression according to a "binary" model—differentiating those of 'constitutional origin' (i.e., caused from within the individual) versus those that were reactions to environmental stressors. There have been several examples in the literature that support the existence of different depressive subtypes, which is typically achieved by identifying differences between groups on some metric (e.g., clinical or self-report ratings of an illness variable). Kiloh and Garside [24] quantified the independence

of neurotic and endogenous (synonymous with melancholic) depressives through analysis of reported symptoms and clinical variables. Features such as psychomotor retardation and concentration difficulties correlated with the diagnosis of endogenous depression, whereas 'reactivity of depression' (by which it is assumed to mean reactivity of affect), irritability and variability of illness were correlated with neurotic depression. Similarly, a 'point of rarity' was identified between endogenous and neurotic depressives in a series of papers from the Newcastle school [25, 26]. Here, clinical ratings of depressive symptoms and signs, along with personality and anxiety, were shown to differentiate endogenous and neurotic depressive subgroups. Again, psychomotor change was more prevalent in the endogenous group than the neurotic group. Despite strong support for the notion that endogenous depression is a categorically distinct entity (one either has it or does not), and that neurotic depression consisted of symptoms varying dimensionally [23], it was abandoned under the DSM system in favor of a dimensional approach to diagnosis.

'Melancholia' was retained in DSM-III (and subsequent iterations of DSM), albeit in much-diluted form, as a 'specifier' diagnosis [19]. The retaining of a melancholic specifier allows for diagnosis of major depression with melancholic features, but has been criticized for its focus on symptom expression (a severity framework), while also explicitly disregarding differing aetiological contributions [18]. In the DSM, melancholia is diagnosed by the presence of additional symptoms of anhedonia (reduced interest or reactivity to previously pleasurable events) and psychomotor slowing, amongst others, which are present in almost all individuals with clinically significant depression [27]. Observable (not just reported) psychomotor disturbance has been proposed as a specific diagnostic marker of melancholia, which aligns with some of the earliest definitions of the disorder, from classical antiquity, where "symptoms … were not part of the concept" ([6] p. 298). We therefore developed a clinician-rated scale (the CORE) to measure psychomotor disturbances in depressed patients [28]. The tool allows rating of 18 clinical signs (observable features) across the domains of non-interactiveness (including features such as emotional non-reactivity and inattentiveness), retardation (including slowed motor movements and facial immobility), and agitation (including facial apprehension and motor agitation). Clinically diagnosed melancholia (still arguably the "gold standard" in diagnosing the condition) was associated with "substantial" CORE scores. In these studies a score of >8 defined "substantial" psychomotor disturbances [27], with such scores being representative of those with melancholia but not those with non-melancholic depression. Despite the high sensitivity in detecting melancholia, such systems—much like clinical diagnosis—require extensive training and exposure to appropriate clinical populations (i.e., in services where those with melancholia are likely to present) to be of any benefit. Several investigators, including our own research team, have thus sought to clarify biological correlates of melancholia in the hope that any identified perturbations will eventually assist in its diagnosis.

Throughout the late 1970s and early 1980s there were many investigations into disturbances of hypothalamic-pituitary-adrenal (HPA) axis function in depressive illness [29]. The HPA axis plays a central role in regulating homeostasis of many physical systems, including the metabolic, reproductive, immune, and central

nervous systems [30]. Insights into the function of this system in depression has been achieved by challenging patients with the synthetic corticosteroid, dexamethasone (DEX), and then observing fluctuations in plasma cortisol—referred to as the DEX suppression test or DST. It was seen as a watershed for psychiatry when Carroll and colleagues [31] reported that 48 % of those with primary depression were DEX 'non-suppressors', compared to only 2 % of other psychiatric patients. Carroll et al. [32] subsequently demonstrated that the DST had utility in detecting melancholic depression, with sensitivity of 67 % and specificity of 96 %. This measure hence appeared to be highly informative in depressive subtyping (i.e., nearly all non-melancholic patients were correctly classified as not having melancholia). Despite such promising early findings, the DST lost favor as a diagnostic tool in the mid-1980s after the DSM-III was introduced—given it lacked sensitivity in later studies using broader diagnostic criteria [33]. Since then, disruptions across other cognitive and biological systems have been identified with specificity to the melancholic phenotype, and include working memory impairments and disturbances in sleep architecture [34]. Our team also recently identified a neurobiological signature for melancholia [35]—which was not observed in non-melancholic depression—involving disrupted integration of brain regions supporting interoception and attention.

While such research has been of key importance in understanding the underlying causes of melancholia, their use as diagnostic tools will continue to be limited given their invasiveness, relatively high cost, and difficulty of access (e.g., in rural areas). Facial imaging has the potential to overcome these barriers and become an important tool in the diagnosis of melancholia. In the following sections we overview methodological advances in facial imaging research, and highlight the utility of the methods in contributing to an objectively-informed diagnostic tool for depressive disorders.

4 Methodological Considerations for Quantifying Facial Affect in Depression

Significant advances have been made for inferring affect using facial imaging over the past two decades [36–40]. In this section we discuss the general methodological approaches of affect recognition based on facial analysis. A typical facial analysis system has the following main components: face and fiducial points (facial parts location) detection, feature extraction, and classification. Figure 1 depicts the main components of such a system. Once the image has been captured, face detection algorithms, such as the popular Viola-Jones (VJ) detector [41], are used to locate the face. Next, fiducial points are inferred using parametric models such as the widely used Active Appearance Models (AAM) [42]. Once the location of facial parts such as the eyes and mouth are known, facial features are computed. Features can be extracted either on a holistic level or on individual parts of the face. The

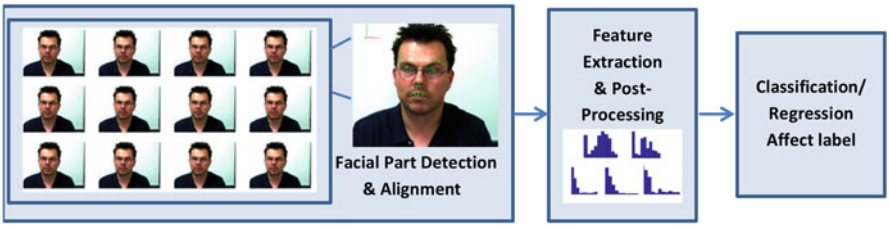

Fig. 1 A typical FAR system

individual features computed in the latter case are concatenated to construct the final feature vector. Further, dimensionality reduction methods such as Principal Component Analysis (PCA) can be applied to the feature vector. This provides a compact and less noisy feature representation capable of higher discrimination (e.g., between features). The classifier then categorizes a given face into differing affective states (such as depressed/non-depressed, various emotion types, and continuous valence/arousal labels etc). The components of a facial affect recognition (FAR) system are discussed in detail in the following sections.

4.1 Face and Fiducial Points Detection

One of the most widely used face detectors is the classic VJ face detector [41]. It is based on a cascade-boosting framework [43] in which haar-like features are extracted using an integral image. The use of an integral image leads to near real-time execution of the face detection method. The cascade classifier scans an image using a sub-window at different scales and localizes the regions that are labeled as faces by the classifier. The Adaboost algorithm is also applied for selecting discriminative haar-features during the training phase of the classifier. The cascade classifier consists of various weak classifiers, which together act as a strong classifier. The use of weak classifiers early on allows fast rejection of patches that do not resemble faces. The open source computer vision library, OpenCV, contains an implementation of the VJ object detector. For multi-view face detection, multiple models are learned for different head poses [44]. These select features using forward feature selection before training the cascade classifier.

There are several other facial detection methods, including energy-based models which infer the face location and head pose simultaneously [45], and vector boosting-based methods where a tree representation is proposed for dividing the face space into smaller subspaces [46]. The power of non-rigid deformable models stems from the low-dimensional representation of the shape and texture of a face they provide. One of the earliest deformable model methods is the Active Contour Model [47]. The Active Shape Model (ASM) algorithm [48] models the shape of an object and has been used extensively in face tracking. During the training process, the

landmark points of all input samples are aligned into a common co-ordinate frame using Procrustes analysis. Following this, the model is computed by applying PCA over the shapes. The shape of the model can be controlled/changed via parameters of deformation. Then, fitting of the model can be performed on a new image using an iterative method, which calculates the best match for the model boundary and hence decides the new location for the model points.

4.2 Active Appearance Models

The Active Appearance Models (AAM) are an extension of the ASM. However, they not only model the shape but also consider the grey-level appearance. During the training process the grey-level appearance is modeled by warping each training sample image using a triangulation algorithm that aligns it to the mean shape. The grey-level information is then sampled and PCA is applied to the samples. Hence, this model can then be fitted to a new image using an optimization algorithm that uses the difference between intensities of the learnt model and the reference image. There are a number of AAM fitting algorithms that can be broadly classified into two classes: generative and discriminative fitting. In generative fitting (*Fixed Jacobian* [42], *Project Out Inverse Compositional* [49], *Simultaneous Inverse Compositional* [50]), minimization/maximization of some measure of fitness between the model's texture and warped image region is applied to the image. In discriminative fitting (*Iterative Error Bound Minimization* [51], *Haar-like Feature Based Iterative Discriminative Method* [52]) a relationship is learned between the features and the parameters, by using the features extracted from parameter settings, which are perturbed from their optimal setting in each image. The disadvantage of AAM is their limited generalizability to unobserved subjects. AAM's can be classified as subject-dependent or subject-independent. Subject-dependent AAM is best suited to scenarios when the train and test images have the same subjects. Subject-independent modeling is for situations where the subjects in the train and test set are different. In one of the earliest works in automatic depression analysis, McIntyre et al. [53] and Cohn et al. [54] used subject-dependent AAM for facial part detection. In such studies, the facial points were used as feature descriptors for learning a classifier.

Constrained Local Models (CLM) [55] are an extension of the AAM algorithms. The texture is divided into blocks. This helps in generalization and better subject-independent performance. Subject-dependent AAM methods perform better than subject-independent CLM [56]. However, the current state-of-art descriptors compensate for small errors introduced by subject-independent CLM [56]. Another limitation of both AAM and CLM is their requirement of large volumes of labeled data representing different scenarios, such as illumination, pose and expression during training. However, despite the advantages, labeling fiducial points is a manually laborious and erroneous task.

4.3 Pictorial Structure

In the Pictorial Structure (PS) framework [57], an object is represented as a graph with n vertices $V = \{v_1, \ldots, v_n\}$ for the parts and a set of edges E, where each $(v_i, v_j) \in E$ pair encodes the spatial relationship between parts i and j. For a given image I, PS learns two models. The first one learns the evidence of each part as an *appearance model*, where each part is parameterized by its location (x, y), orientation θ, scale s, and foreshortening. All of these parameters (together referred to as D) are learned from exemplars and produce a likelihood model for I. The second model learns the kinematic constraints between each pair of parts in a prior *configuration model. L* is the parts configuration. Given the two models, the posterior distribution over the whole set of part locations is:

$$p\left(L \left| I, D \right.\right) \alpha \, p\left(I \left| L, D \right.\right) \tag{1}$$

where $p(I|L, D)$ measures the likelihood of representing I in a particular configuration and $p(L|D)$ is the kinematic prior configuration. A major problem of this framework is the low contribution of the occluded parts, resulting in either erroneous or missed detection of these parts, leading to inaccurate pose estimation. Everingham and colleagues [58] proposed a PS based fiducial point detector, which is initialized using the VJ face detector. Recently, Zhu and Ramanan [59] proposed an extension to the PS framework by adding mixtures representing different face poses. This latter study performed face and fiducial point detection and head pose inference in the one framework. The face detector performs better than the VJ face detector. The disadvantage of the PS based method as used by Everingham et al. [58] is that it requires initialization from a face detector like VJ. However, Zhu and Ramanan [59] overcome this limitation by using multiple pose as mixture detectors.

Selecting the appropriate face and fiducial point detector is problem driven. For example, in the case of affect analysis, it is desirable that the system should generalize over subjects. Joshi et al. [60] used the PS framework of Everingham and colleagues [58] for fiducial points detection. Even though the Mixture of Pictorial Structures (MoPS) framework performs better than the PS method, our own research group [60] prefer the use of PS. We argue that the inference time for MoPS is substantially longer than that obtained in the Everingham et al. [58] study, which matters when analyzing long duration depression video clips. Furthermore, such work can utilize spatio-temporal descriptors (Local Binary Pattern in Three Orthogonal Planes (LBP-TOP) [61]) that compensate for algorithm error. On the other hand, applications such as facial performance transfer [62] require accurate fiducial points. Asthana and colleagues [62] use subject dependent AAM models as they are more accurate as compared to CLM and subject-independent AAM.

4.4 Facial Descriptors

Once the face and facial parts location is identified and aligned, feature descriptors are computed for extracting information for learning classifiers. FAR techniques can be segregated on the basis of the type of descriptors used, broadly classified as either geometric- or appearance-based features. Geometric features [63–65] correspond to facial points and to the location of different facial parts. Appearance features generally correspond to face texture information [61, 66, 67]. Furthermore, facial feature descriptors can be divided on the basis of temporal information (i.e., spatio-temporal descriptors [61] and frame-based descriptors [67]). A popular method for modeling geometric features is based on Facial Animation Parameters (FAP), defined in the MPEG-4 video-coding standard. Lavagetto and Pockaj [68] present a method for synthesizing facial animations using FAP and Facial Definition Parameters (FDP). Similarly, Sebe and colleagues [69] use the Piecewise Beizier Volume Deformation (PBVD) tracker for tracking facial parts. The motion information between two consecutive frames is measured using template matching.

Various classifiers can also be compared, allowing insight to be gained as to their utility in depression classification. Asthana et al. [70] compute geometric features by fitting AAM models on input faces. They compared various AAM fitting techniques and experimented on a Cohn-Kanabe (CK+) database. One of the limitations of this work is that it required manual initialization of facial parts. Dhall and colleagues [63] propose the use of the geometric descriptor algorithm, "Emotion Image" (EI). This feature constructs a visual map based on an undirected map derived from a facial points detector. EIs of two faces (images) is compared using the Structural Similarity Index Metric (SSIM) of Wang et al. [71], allowing computation of their similarity, and is applied to the problem of expression based album creation. The discriminative ability of EI is dependent on fiducial point detection quality, which may introduce some errors when the fiducial points detection is not very accurate. Valstar and Pantic [72] showed that the performance of geometric features is similar to that of the appearance features. However, the limitation of geometric features comes from their dependence on accurate facial parts location information. Facial parts detection is relatively accurate on lab-controlled scenario data; however, it is still an open problem for images in real-world conditions. If there is an error in the facial parts detection, the error generally propagates in the geometric feature representation. Chew et al. [56] argue that appearance descriptors are able to compensate error produced by facial parts detectors to some extent. Popular appearance descriptors are described below.

4.4.1 Local Binary Patterns (LBP)

The LBP family of descriptors has been extensively used in computer vision for texture and face analysis [59, 73, 74]. The LBP descriptor assigns binary labels to pixels by thresholding the neighborhood pixels with the central value. Therefore,

for a center pixel p of an image I and its neighboring pixels N_i, a decimal value d is assigned to it:

$$d = \sum_{i=1}^{k} 2^{i-1} I(p, Ni)$$
$$\text{where } I(p, Ni) = \begin{cases} 1 \text{ if } c < N_i \\ 0 \text{ otherwise} \end{cases} \tag{2}$$

4.4.2 Local Binary Pattern-Three Orthogonal Planes (LBP-TOP)

Local Binary Pattern-Three Orthogonal Planes (LBP-TOP) [61] is a popular descriptor in computer vision. It considers patterns in three orthogonal planes: *XY*, *XT* and *YT*, and concatenates the pattern co-occurrences in these three directions. The local binary pattern (LBP-TOP) descriptor assigns binary labels to pixels by thresholding the neighborhood pixels with the central value. Therefore, for a center O_p of an orthogonal plane O and its neighboring pixels N_i, a decimal value d is assigned to it:

$$d = \sum_{O}^{XY,XT,YT} \sum_{p} \sum_{i=1}^{k} 2^{i-1} I\left(O_p, N_i\right) \tag{3}$$

Joshi et al. [75] proposed a LBP-TOP based framework for analyzing depression data. The video clips were divided into temporal slices and LBP-TOP was computed on each time slice. Temporal slicing helps in encoding spatio-temporal changes. Further, a Bag-of-Words (BoW) representation was learnt with LBP-TOP from each temporal slice from an interview-based video (a 'document'). BoW-based representations come from the domain of document processing. A BoW feature represents a document (image/video) as an unordered set of frequencies of words. Li and colleagues [76] were the first to use a BoW for FAR—they fused PHOG- and BoW-based histograms constructed from a dictionary based on Scale Invariant Feature Transform (SIFT). Even though BoW-based vectors can represent the frequency of different stages of an expression, the temporal sequencing information is still missing. To overcome this problem, a data-driven technique was recently proposed to explicitly encode the temporal information using n-grams. Bettadapura et al. [77] performed experiments on human action recognition and activity analysis and showed that adding temporal sequencing information based on their method increases the accuracy of the BoW-based techniques.

The facial descriptor representation modeled on computing features based on the output of the facial parts detection module can be referred to as explicit modeling of affect. Here, an explicit model (face model) is used to localize the facial parts. In a different approach (implicit modeling), Joshi et al. [60] proposed the computation of Space Time Interest Points (STIP) [78] on the upper body of a subject in the video frame in depression. STIP are widely used in the computer vision community

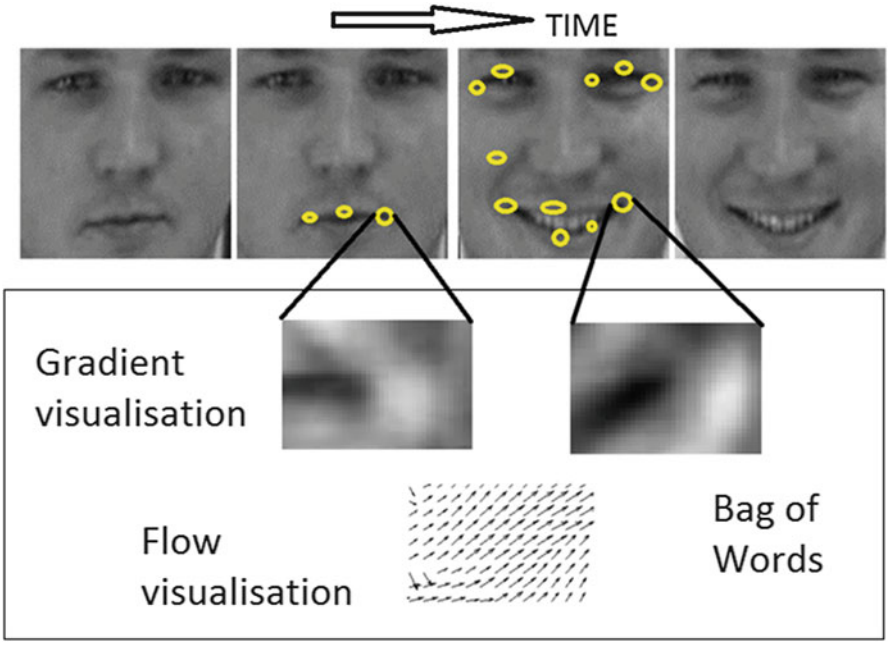

Fig. 2 The figure describes the STIP computation on a video from FEEDTUM database [42]. The blocks represent the gradient information [41] around the generated interest points. The HOG and HOF features represent the local spatio-temporal movements

for human action recognition. STIP are salient spatio-temporal locations in a video, where a change has occurred. These are based on a 3D extension of the 2D Harris interest point detector. Once an interest point is detected, Histogram of Gradients (HOG) and Histogram of Flow (HOF) is computed around the interest point. Joshi et al. [75] computed STIP on the upper body and compared the performance with face area only STIP in depressed subjects. Further, a BOW is learnt on HOG and HOF features generated around the interest points. Figure 2 describes the STIP computation on a sample from the FEEDTUM database [79].

Note that the yellow ellipses around the eyes and mouth are the interest points generated. The regions of interest around the interest points are scaled up, and gradient information is shown (as demonstrated previously [80]). It is easy to notice that the two gradient blocks around the same region of interest (i.e., right tip of the mouth) show different gradients due to changes in facial expression. The flow information here describes the motion change around the interest point on the right tip point of the mouth. Hence, STIP captures local movements on a holistic level (full-face), and information is captured implicitly without having to use any output from a facial parts localization method.

4.5 Objective Identification of Melancholia

Many of the above methods have recently been used to facilitate better identification of those with depression. Williamson and colleagues [81] showed that depression was associated with changes in the coordination and movement of facial gestures using facial action units. When used in multivariate modeling, such facial features were predictive of self-reported depressive symptomatology, highlighting the utility of the approach in quantifying the depressive syndrome. There is also evidence to suggest that facial landmark fitting (66 points to the face on each frame) allows robust prediction of affective dimensions (i.e., valence/arousal) and global depression state (measured through self-rated depression scores) [82]. In addition, the predictive capacity of facial imaging in detecting depression appears to increase with the addition of other 'affective computing' data modalities. Pérez-Espinosa and colleagues [83] demonstrated that fusing affective dimensions and audiovisual features (facial imaging) allowed for accurate construction of depression recognition models. Whilst no studies have directly examined the performance of the methods above in classifying different types of depression such as melancholia, we propose that they will likely be of significant benefit in future studies. Indeed, our own research team recently completed recruitment for a large facial imaging study of those with melancholic depression, those with a non-melancholic depression, and healthy controls, to determine whether the above methods could be used to accurately classify these groups. Preliminary analyses using data from this study have been completed [84], and analyses are now underway to determine whether melancholia can be detected on the basis of its unique, and quantifiable, facial features.

5 Summary and Future Trends

In this chapter we have reviewed the role of facial imaging technologies in detecting affective states, and specifically the potential of emerging methods in classifying depressive disorders. Methodological advances and continuing algorithm development in the field of affective computing have the potential to offer unique insights regards depression detection. Objective characterization of facial expressions with specificity to melancholia will be the next step in determining the overall clinical utility of these methods. Based on the literature to date, however, facial imaging technologies are well positioned to contribute to the assessment and monitoring of depressive disorders.

References

1. S. Bloch, B.S. Singh, *Foundations of Clinical Psychiatry*, 2nd edn. (Melbourne University Publishing, Melbourne, 2001)
2. W. James, What is an emotion? Mind **9**, 188–205 (1894)
3. R. Reisenzein, A short history of psychological perspectives on emotion, in *The Oxford Handbook of Affective Computing*, ed. by R. Calvo, S. D'Mello, J. Gratch, A. Kappas (Oxford University Press, Oxford, 2015)
4. M.B. Arnold, *Emotion and Personality* (Columbia University Press, New York, 1960)
5. R.S. Lazarus, *Psychological Stress and the Coping Process* (McGraw-Hill, New York, 1966)
6. G.E. Berrios, *The History of Mental Symptoms: Descriptive Psychopathology Since the Nineteenth Century* (Cambridge University Press, Cambridge, UK, 1996)
7. W. McDougall, *An Introduction to Social Psychology* (Methuen, London, 1908/1960)
8. K.L. Davis, J. Panksepp, The brain's emotional foundations of human personality and the affective neuroscience personality scales. Neurosci. Biobehav. Rev. **35**, 1946–1958 (2011)
9. S.S. Tomkins, *Affect Imagery Consciousness: Volume I: The Positive Affects* (Springer Publishing Company, New York, 1962)
10. S.S. Tomkins, *Affect Imagery Consciousness: Volume II: The Negative Affects* (Springer Publishing Company, New York, 1963)
11. S.S. Tomkins, *Affect Imagery Consciousness: Volume III: Anger and Fear* (Springer Publishing Company, New York, 1991)
12. C.E. Izard, *The Face of Emotion* (Appleton-Century Crofts, New York, 1971)
13. P. Ekman, Facial expression and emotion. Am. Psychol. **48**, 384 (1993)
14. C.E. Izard, Basic emotions, natural kinds, emotion schemas, and a new paradigm. Perspect. Psychol. Sci. **2**, 260–280 (2007)
15. J.A. Russell, A circumplex model of affect. J. Pers. Soc. Psychol. **39**, 1161–1178 (1980)
16. D. Keltner, P. Ekman, Facial expressions of emotions, in *Handbook of Emotions*, ed. by M. Lewis, J. Haviland-Jones, 2nd edn. (Guildford Press, New York, 2000)
17. P. Ekman, W.V. Friesen, *The Facial Action Coding System: A Technique for the Measurement of Facial Movement* (Consulting Psychologists, Palo Alto, 1978)
18. G. Parker, Classifying depression: should paradigms lost be regained? Am. J. Psychiatry **157**, 1195–1203 (2000)
19. APA, *Diagnostic and Statistical Manual of Mental Disorders (DSM-III)*, 3rd edn. (American Psychiatric Association, Washington, DC, 1980)
20. E. Kraepelin, *Manic-Depressive Insanity and Paranoia* (E. & S. Livingstone, Edinburgh, 1921)
21. E. Mapother, Discussion on manic-depressive psychosis. Br. Med. J. **ii**, 872–879 (1926)
22. A. Lewis, Melancholia: a clinical survey of depressive states. J. Ment. Sci. **80**, 277–378 (1934)
23. G. Parker, *The Bonds of Depression* (Angus & Robertson Publishers, Sydney, 1978)
24. L.G. Kiloh, R.F. Garside, The independence of neurotic depression and endogenous depression. Br. J. Psychiatry **109**, 451–463 (1963)
25. M.W.P. Carney, M. Roth, R.F. Garside, The diagnosis of depressive syndromes and the prediction of ECT response. Br. J. Psychiatry **111**, 659–674 (1965)
26. C. Gurney, M. Roth, R.F. Garside, T.A. Kerr, K. Schapira, Studies in the classification of affective disorders. The relationship between anxiety states and depressive illnesses. II. Br. J. Psychiatry **121**, 162–166 (1972)
27. G. Parker, P.D. Hadzi, *Melancholia: A Disorder of Movement and Mood* (Cambridge University Press, New York, 1996)
28. G. Parker, D. Hadzi-Pavlovic, K. Wilhelm et al., Defining melancholia: properties of a refined sign-based measure. Br. J. Psychiatry **164**, 316–326 (1994)
29. C.M. Pariante, S.L. Lightman, The HPA axis in major depression: classical theories and new developments. Trends Neurosci. **31**, 464–468 (2008)
30. J.P. Herman, W.E. Cullinan, Neurocircuitry of stress: central control of the hypothalamo-pituitary-adrenocortical axis. Trends Neurosci. **20**, 78–84 (1997)

31. B.J. Carroll, G.C. Curtis, J. Mendels, Neuroendocrine regulation in depression. II. Discrimination of depressed from nondepressed patients. Arch. Gen. Psychiatry **33**, 1051–1058 (1976)
32. B.J. Carroll, M. Feinberg, J.F. Greden et al., A specific laboratory test for the diagnosis of melancholia. Standardization, validation, and clinical utility. Arch. Gen. Psychiatry **38**, 15–22 (1981)
33. E. Shorter, M. Fink, *Endocrine Psychiatry: Solving the Riddle of Melancholia* (Oxford University Press, New York, 2010)
34. G. Parker, M. Fink, E. Shorter et al., Issues for DSM-5: whither melancholia? The case for its classification as a distinct mood disorder. Am. J. Psychiatry **167**, 745–747 (2010)
35. M.P. Hyett, M.J. Breakspear, K.J. Friston, C.C. Guo, G.B. Parker, Disrupted effective connectivity of cortical systems supporting attention and interoception in melancholia. JAMA Psychiatry **72**, 350–358 (2015)
36. I. Cohen, N. Sebe, L. Chen, A. Garg, T.S. Huang, Facial expression recognition from video sequences: temporal and static modelling. Comput. Vis. Image Underst. **91**, 160–187 (2003)
37. B. Fasel, J. Luettin, Automatic facial expression analysis: a survey. Pattern Recogn. **36**, 259–275 (2003)
38. M. Pantic, L.J.M. Rothkrantz, Facial action recognition for facial expression analysis from static face images. IEEE Trans. Syst. Man Cybern. B Cybern. **34**, 1449–1461 (2004)
39. E. Sariyanidi, H. Gunes, A. Cavallaro, Automatic analysis of facial affect: a survey of registration, representation and recognition. IEEE Trans. Pattern Anal. Mach. Intell. **37**, 39–58 (2015)
40. Z. Zeng, M. Pantic, G.I. Roisman, T.S. Huang, A survey of affect recognition methods: audio, visual, and spontaneous expressions. IEEE Trans. Pattern Anal. Mach. Intell. **31**, 39–58 (2009)
41. P.A. Viola, M.J. Jones, Rapid object detection using a boosted cascade of simple features, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2001) pp. 511–518
42. T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, in *Proceedings of the European Conference on Computer Vision (ECCV)* (Springer,1998), pp. 484–498
43. Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, in *Computational Learning Theory*, 23–27, Springer (Springer, 1995), pp. 23–27
44. M. Jones, P. Viola, Fast multi-view face detection, in *Mitsubishi Electric Research Lab TR-20003-96,* vol. 3 (MERL, 2003), p. 14
45. M. Osadchy, Y.L. Cun, M.L. Miller, Synergistic face detection and pose estimation with energy-based models. J. Mach. Learn. Res. **8**, 1197–1215 (2007)
46. C. Huang, H. Ai, Y. Li, S. Lao, Vector boosting for rotation invariant multi-view face detection, in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* 446–453. (IEEE, 2005), pp. 446–453
47. M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models. Int. J. Comput. Vis. **1**, 321–331 (1988)
48. T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham, Active shape models-their training and application. Comput. Vis. Image Underst. **61**, 38–59 (1995)
49. S. Baker, I. Matthews, Equivalence and efficiency of image alignment algorithms, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2001), pp. 1090–1097
50. S. Baker, R. Gross, I. Matthews, *Lucas-Kanade 20 Years on: A Unifying Framework: Part 3* (Carnegie Mellon University, RI, USA, 2003)
51. J. Saragih, R. Göcke, Iterative error bound minimisation for AAM alignment, in *Proceedings of the International Conference on Pattern Recognition (ICPR)* (IEEE, 2006), pp. 1192–1195
52. J. Saragih, R. Göcke, A Nonlinear discriminative approach to AAM fitting, in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (IEEE, 2007), pp. 1–8
53. G. McIntyre, R. Goecke, M. Hyett, M. Green, M. Breakspear, An approach for automatically measuring facial activity in depressed subjects, in *Proceedings of Affective Computing and Intelligent Interaction (ACII)* (Springer, 2009), pp. 1–8

54. J.F. Cohn, T.S. Kreuz, I. Matthews, et al. Detecting depression from facial actions and vocal prosody, in *Proceedings of Affective Computing and Intelligent Interaction (ACII)*, (Springer, 2009), pp. 1–7
55. J.M. Saragih, S. Lucey, J. Cohn, Face alignment through subspace constrained mean-shifts, in *Proceedings of the IEEE International Conference of Computer Vision (ICCV)*, September (IEEE, 2009), pp. 1034–1041
56. S.W. Chew, P. Lucey, S. Lucey et al., In the pursuit of effective affective computing: the relationship between features and registration. IEEE Trans. Syst. Man Cybern. B Cybern. **42**, 1006–1016 (2012)
57. P.F. Felzenszwalb, D.P. Huttenlocher, Pictorial structures for object recognition. Int. J. Comput. Vis. **61**, 55–79 (2005)
58. M. Everingham, J. Sivic, A. Zisserman, Hello! My name is: Buffy" – automatic naming of characters in TV Video, in *Proceedings of the British Machine and Vision Conference (BMVC)* (Springer, 2006), pp. 899–908
59. X. Zhu, D. Ramanan, Face detection, pose estimation, and landmark localization in the wild, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2012), pp. 2879–2886
60. J. Joshi, A. Dhall, R. Goecke, M. Breakspear, G. Parker, Neural-net classification for spatio-temporal descriptor based depression analysis, in *Proceedings of the International Conference on Pattern Recognition (ICPR)* (IAPR, 2012), pp. 2634–2638
61. G. Zhao, M. Pietikainen, Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans. Pattern Anal. Mach. Intell. **29**, 915–928 (2007)
62. A. Asthana, M. de la Hunty, A. Dhall, R. Goecke, Facial performance transfer via deformable models and parametric correspondence. *IEEE Trans. Vis. Comput. Graph.***18**, 1511–1519 (2012)
63. A. Dhall, A. Asthana, R. Goecke, Facial expression based automatic album creation, in *Proceedings of the Neural Information Processing. Models and Applications (ICONIP)*, (Springer, 2010), pp. 485–492
64. M. Pantic, L. Rothkrantz, Expert system for automatic analysis of facial expression. Image Vision Comput. J. **18**, 881–905 (2000)
65. L. Zhang, D. Tjondronegoro, V. Chandran, Evaluation of texture and geometry for dimensional facial expression recognition, in *Proceedings of the International Conference on Digital Image Computing Techniques and Applications (DICTA)* (IEEE, 2011), pp. 620–626
66. A. Dhall, A. Asthana, R. Goecke, T. Gedeon, Emotion recognition using PHOG and LPQ features, in *Proceedings of the IEEE Conference Automatic Faces & Gesture Recognition workshop FERA* (IEEE, 2011), pp. 878–883
67. K. Sikka, T. Wu, J. Susskind, M. Bartlett, Exploring bag of words architectures in the facial expression domain, in *Proceedings of the European Conference on Computer Vision and Workshops (ECCVW)* (Springer, 2012), pp. 250–259
68. F. Lavagetto, R. Pockaj, The facial animation engine: towards a high-level interface for the design of MPEG-4 compliant animated faces. IEEE Trans. Circuits Syst. Video Technol. **9**, 277–289 (1999)
69. N. Sebe, I. Cohen, T. Gevers, T.S. Huang, Emotion recognition based on joint visual and audio cues, in *Proceedings of the International Conference on Pattern Recognition (ICPR)*, 2006
70. A. Asthana, J. Saragih, M. Wagner, R. Goecke, Evaluating AAM fitting methods for facial expression recognition, in *Proceedings of the IEEE International Conference on Affective Computing and Intelligent Interaction (ACII)* (Springer, 2009), pp. 598–605
71. Z. Wang, A.C. Bovik, H.R. Sheikh, S. Member, E.P. Simoncelli, S. Member, Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**, 600–612 (2004)
72. M. Valstar, M. Pantic, Fully automatic facial action unit detection and temporal analysis, in *Proceedings of the International Conference on Computer Vision and Pattern Recognition Workshop (CVPRW)*, (IEEE, 2006), pp. 149–149
73. T. Ojala, M. Pietikinen, T. Menp, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**, 971–987 (2002)

74. V. Ojansivu, J. Heikkilä, Blur insensitive texture classification using local phase quantization, in *Proceedings of the Image and Signal Processing (ICISP)* (Springer, 2008), pp. 236–243
75. J. Joshi, R. Goecke, M. Breakspear, G. Parker, Can body expressions contribute to automatic depression analysis? in *Proceedings of the International Conference on Automatic Face and Gesture Recognition (FG)* (IEEE, 2013), pp. 1–7
76. Z. Li, J.-I. Imai, M. Kaneko, Facial-component-based bag of words and PHOG descriptor for facial expression recognition, in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC)* (IEEE, 2009), pp. 1353–1358
77. V. Bettadapura, G. Schindler, T. Plötz, I. Essa, Augmenting bag-of-words: data-driven discovery of temporal and structural information for activity recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2013), pp. 2619–2626
78. I. Laptev, T. Lindeberg, Space-time interest points, in *International Conference on Computer Vision (ICCV)* (IEEE, 2003), pp. 432–439
79. F. Wallhoff, *Facial Expressions and Emotion Database*, Technical Report (2006)
80. C. Vondrick, A. Khosla, T. Malisiewicz, A. Torralba, HOGgles: Visualizing object detection features, in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (IEEE, 2013), pp. 1–8
81. J.R. Williamson, T.F. Quatieri, B.S. Helfer, G. Ciccarelli, D.D. Mehta, Vocal and facial biomarkers of depression based on motor incoordination and timing, in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge (AVEC)* (ACM, 2014), pp. 65–72
82. R. Gupta, N. Malandrakis, B. Xiao, et al. Multimodal prediction of affective dimensions and depression in human-computer interactions, in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge (AVEC)* (ACM, 2014), pp. 33–40
83. H. Pérez-Espinosa, H.J. Escalante, L. Villaseñor-Pineda, M. Montes-y-Gómez, D. Pinto Avedaño, V. Reyez-Meza, Fusing affective dimensions and audio-visual features from segmented video for depression recognition, in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge* (ACM, 2014), pp. 49–55
84. J. Joshi, R. Goecke, S. Alghowinem et al., Multimodal assistive technologies for depression diagnosis and monitoring. J. Multimodel User Interfaces **7**, 217–228 (2013)