# Consistent Optical Flow Maps for full and micro facial expression recognition

Benjamin Allaert, Ioan Marius Bilasco, Chaabane Djeraba

# Consistent Optical Flow maps for full and micro facial expression recognition

Benjamin Allaert[1], Ioan Marius Bilasco[1] and Chabane Djeraba[1]
[1]*Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL - Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France*
*benjamin.allaert@ed.univ-lille1.fr, {marius.bilasco, chabane.djeraba}@univ-lille1.fr*

Keywords: Facial expression, Micro-expression, Optical Flow

Abstract: A wide variety of face models have been used in the recognition of full or micro facial expressions in image sequences. However, the existing methods only address one family of expression at a time, as micro-expressions are quite different from full-expressions in terms of facial movement amplitude and/or texture changes. In this paper we address the detection of micro and full-expression with a common facial model characterizing facial movements by means of consistent Optical Flow estimation. Optical Flow extracted from the face is generally noisy and without specific processing it can hardly cope with expression recognition requirements especially for micro-expressions. Direction and magnitude statistical profiles are jointly analyzed in order to filter out noise and obtain and feed consistent Optical Flows in a face motion model framework. Experiments on CK+ and CASME2 facial expression databases for full and micro expression recognition show the benefits brought by the proposed approach in the filed of facial expression recognition.

## 1 INTRODUCTION

Automatic facial expression analysis has attracted great interest over the past decade in various domains. Facial expression recognition has been widely studied in computer vision. Recent methodologies for static expression recognition have been proposed and obtain good results for acted expression. However, in order to cope with the natural context challenges like face occlusions, non-frontal poses, expression intensity and amplitude variations must be addressed.

Challenges like illumination variation, face occlusions, non-frontal poses have been addressed in fields other then expression recognition. Several research results were also published on this topic primarily based on face alignment. Although the methodology is more mature, it is far from being fully robust. This topic attracts still many researches and discussions.

In the following we focus on challenges brought by supporting a wide range of facial movement amplitudes when producing a full or micro expression. In case of full expression the underlying facial movement and the induced texture deformation can be clearly differentiated from the noise that can appear when analyzing the face properties. However, as the amplitudes are much smaller in micro-expressions attention must be paid to small changes encoding.

Automatic micro-expression recognition algorithms have recently received growing attention in the literature (Yan et al., 2014; Liu et al., 2015; Wang et al., 2014b; Wang et al., 2014c). Micro-expressions are quite different from full-expression recognition. They are characterized by rapid facial movements having low intensity. Micro-expressions typically involve a fragment of the facial region. Therefore, previous work that were suitable for full-expression recognition may not work well for micro-expressions. In other words, it seems difficult to find a common methodology for analyzing full and micro expression in an accurate manner.

Dynamic texture is an extension of texture characterization to the temporal domain. Description and recognition of dynamic textures in facial expression recognition have attracted growing attention because of their unknown spatial and temporal extent. Impressive results have recently been achieved in dynamic texture synthesis using the framework based on a system identification theory which estimates the parameters of a stable dynamic model (Wang et al., 2014b; Wang et al., 2014c). However, the recognition of dynamic texture is a challenging problem compared with the static case (Péteri and Chetverikov, 2005). Indeed, for real videos the stationary dynamic textures must be well-segmented in space and time and it

is difficult to define a metric in the space of dynamic models.

In facial expression recognition, Optical Flow methods are popular as Optical Flow estimation is a natural way to characterize the local dynamics of a temporal texture (Fortun et al., 2015). The use of Optical Flow reduces dynamic texture analysis to the analysis of a sequence of instantaneous motion patterns viewed as static textures. Optical Flow are recently used to analyze full-expression (Liao et al., 2013; Su et al., 2007; Lee and Chellappa, 2014) and micro-expression (Liu et al., 2015). Good performances were obtain in both cases. However, the usage of the Optical Flow is still questioned because the accuracy drops in the presence of motion discontinuities, large displacements or illumination changes. Recent Optical Flow algorithms (Revaud et al., 2015; Chen and Koltun, 2016; Bailer et al., 2015) evolved to better deal with noise and motion discontinuities employing complex filtering requiring high computation time. Still, these algorithms were designed for generic Optical Flow computations and are not adapted to facial morphology and physical constraints.

In this paper, we investigate the effectiveness of using a facial dedicated filtered dense Optical Flow in order to recognize the full-expressions (anger, fear, disgust, happiness, sadness, and surprise) and micro-expressions (positive, negative, surprise) in near-frontal-view recordings. In section 2, we discuss existing work related to static approaches for full and micro expression recognition. In section 3, we present our approach for extracting the coherent movement on the face in different locations from dense Optical Flow method. We filter the noise considering the facial movement hypothesis (local coherency and propagation). Next, we explore the characterization of the coherent Optical Flow into a facial model formulation in section 4 and discuss several strategies for encoding the facial movement for full and micro expressions recognition. Experimental results are discussed in section 5. Finally, the conclusion and future perspectives are given in section 6.

## 2 RELATED WORK

Most of the 2D-feature-based methods are suitable for the analysis of near frontal facial expressions in presence of limited head motions and intense expressions. In order to provide the reader with an overview of approaches challenging these limitations, we present the recent facial alignment method and how the approach of facial expression recognition is adapted to the different intensity of expression.

The face is usually detected and aligned in order to establish the correspondence of major facial components such as eyes, nose, mouth across different face images and to reduce variations in scale, rotation, and position. Alignment based on eyes is the most popular strategy since eyes are the most reliable facial components to be detected and suffer little changes in presence of expressions. Assuming the face region is well aligned, histogram-like features are often computed from equal-sized facial grids. However, apparent misalignment can be observed and it is primarily caused by variations in face pose and facial deformation, as well as the diversity in human face geometry. Recent studies use the facial landmarks to define a facial region that increase robustness to facial deformation during expression. Jiang et al. (Jiang et al., 2014) define a mesh over the whole face with an Active Shape Model (ASM), and extract features from each of the regions enclosed by the mesh. Han et al. (Han et al., 2014) use an Active Apparent Model (AAM) to transform a facial grid and improve feature extraction for recognizing facial Action Units (AUs).

Thanks to recent databases (Yan et al., 2014; Li et al., 2013), the demand for computer vision techniques to improve the performance of micro-expression recognition is increasing. Automatic micro-expression recognition algorithms have recently received attention, but there is still a considerable gap to fill for improving the recognition accuracy. Recent works usually used spatiotemporal local binary pattern (LBP) for micro-expression analysis (Wang et al., 2014c; Wang et al., 2014b; Wang et al., 2014a; Yan et al., 2014). Huang et al. (Huang et al., 2016b) proposed spatiotemporal completed local binary pattern (STCLQP) and obtained promising performances with regard to similar state-of-the-art methods. The reason may be that STCLQP provides more useful information for micro-expression recognition, as STCLQP extracts jointly information characterizing magnitudes and orientations. Huang et al. (Huang et al., 2016a) criticized the fact that state-of-the-art spatiotemporal LBP features are extracted from the global face regions and hence, they ignore the discriminative information between two micro-expression classes. To overcome this problem, they propose a discriminative spatiotemporal LBP based on an improved integral projection. Recently, Liu et al. (Liu et al., 2015) built a feature for micro-expression recognition based on a robust Optical Flow method and extract a Main Directional Mean Optical-flow (MDMO). They showed that the magnitude is more discriminant than the direction when working with micro-expression and they achieve better performance than spatiotemporal LBP approach.

Some approaches employ dense Optical Flow for full expression recognition and perform well in several databases. Su et al. (Su et al., 2007) propose to uniformly distribute 84 feature points over the three automatically located rectangles instead of extracting precise facial features (eyebrows, eyes, mouth). They select the facial regions which contribute more towards the discrimination of expressions. Lee et al. (Lee and Chellappa, 2014) design sparse localized facial motion dictionaries from dense motion flow data of facial expression image sequences. The proposed localized dictionaries are effective for local facial motion description as well as global facial motion analysis. Liao et al. (Liao et al., 2013) improve the existing feature extraction result by learning expression-specific spatial weighting masks. The learned spatial weighted masks correspond to the human attention to discriminate between expressive faces, and determine the importance of facial regions. The weighted masks can significantly increase the performance of facial expressions recognition and intensity estimation on several databases.

Any change made in the facial region has important side-effects on the Optical Flow (Fortun et al., 2015). So it's important that the Optical Flow is computed from the original image, even though motion discontinuities and large displacements may influence the extraction process. Evaluation benchmarks like MPI Sintel benchmark (Butler et al., 2012) propose new challenges allowing the development of Optical Flow capable to cope with the problems encountered in natural interaction situation (e.g. occlusion, large displacements). Despite constant advances (Revaud et al., 2015), handling these issues in a unique method still remains an open problem and demands high computational time. As generally, the robust Optical Flow prediction approaches are based on smoothing motion propagation, when facial movement is considered, they allow to keep a consistency in the local face area, but discriminant information may be lost.

Inspired by the success of simple dense Optical Flow approach, we explore magnitude and direction constraints in order to extract the relevant movement on the face. Considering the smoothing of motion of recent Optical Flow approach, a simple Optical Flow combined with magnitude constraint seems adequate for reducing the noise induced by lighting changes and small head motions. In the next section, we propose a filtering Optical Flow method based on consistent local motion propagation to keep only the pertinent motion during facial expression. The section 4 explores the construction of facial region which generates discriminating features used to separate the six basic expressions effectively and micro-expressions.

# 3 FACIAL RELATED FILTERING

The facial characteristics (skin smoothness, skin reflect and elasticity) involves dealing with the inconsistency and the noise induced by motion discontinuities, as well as, illumination changes while extracting directly the Optical Flow on the face.

Instead of explicitly computing the global Optical Flow field, the Optical Flow constraint equation is used in a specific facial area defined in relation with the facial action coding system in order to keep only the pertinent motion of the face. The pertinent motion is defined as the Optical Flow extracted from regions where the intensity of moving pixels reflects natural facial movements characteristics. We consider a natural facial movement to be uniform during motion if it is characterized by continuity over neighboring pixels, as well as, by continuous diminution of its intensity over neighboring regions. The filtering operation of Optical Flow is divided into several stages and they are illustrated in Figure 1. The Farnebäck algorithm (Farnebäck, 2003) is used to compute fast dense Optical Flow (A). It is not the most accurate algorithm but it ensures that motion is not disaggregated by smoothing and the computation time is low. Based on the Farnebäck flow field, we determine the consistent facial motion from the facial regions having high probability of movement (RHPM) (B). Each RHPM analyze their neighbors behavior in order to estimate the propagation of the motion on the whole face (C). The filtered Optical Flow field is computed from the coherent motion in each RHPM (D).

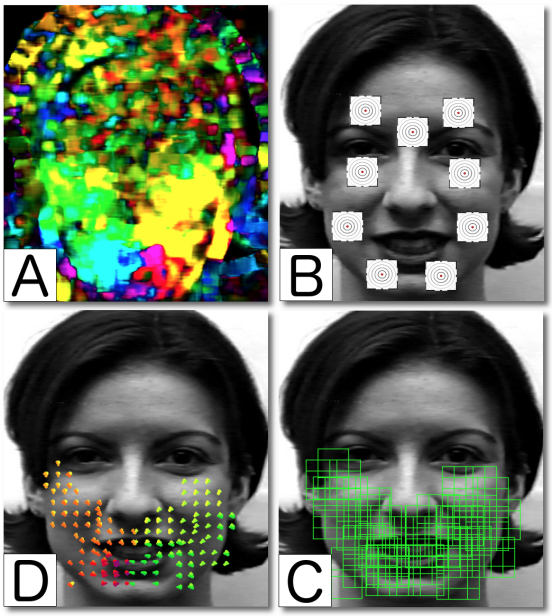


Figure 1: All stages of the proposed method.

Next, we present in detail our approach for extracting the coherent movement in different locations on the face from dense Optical Flow method by filtering the noise on the basis of the facial movement hypothesis assuming local coherency and propagation.

## 3.1 RHPM local coherency

In order to cope with the noise and to filter the Optical Flow information, we start by analyzing the direction distribution within each local region in order to keep only the reliable flow. The proposed method is illustrated in Figure 2.

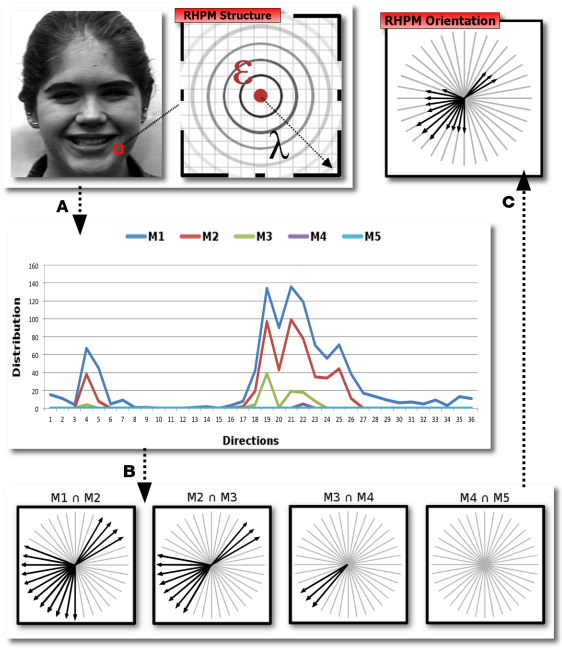


Figure 2: The process of consistent local motion characterization in RHPM

Each region with a high probability of movement contains local Optical Flow information for each pixel : a direction and a magnitude. Each RHPM is defined by a center ε(x,y) called epicenter and a local propagation value λ which define the size of the area under investigation around the epicenter.

In order to measure the consistency of the Optical Flow in terms of directions, we analyze the direction distribution into the RHPM for several layers of magnitude (Figure 2-A). We assume that the motion on the face spread progressively due to the skin elasticity. Furthermore, we have constructed 5 normalized histograms ($M = (M^1, M^2, ..., M^5)$) that represent the direction distribution over 36 bins (of 10° each) for different magnitudes ranges. The magnitude ranges vary according to the characteristics of the data to be processed. We have kept only 5 magnitudes, since they are sufficient to reflect the consistency of movement in facial motion.

Afterwards, the intersection of direction for each pair of consecutive magnitudes is computed to estimate motion overlap between two consecutive magnitudes (Figure 2-B). We build a feature ρ, which represent the intersection between two magnitude histograms by

$$\rho_i^k = \begin{cases} 1, & \text{if } M_i^k > 0 \text{ and } M_{i+1}^k > 0 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where i = 1,2,...,5 is the index of magnitudes and k is the number of bins. The vector ρ is composed only of 0 (no match is found relevant to the bin k) and 1 (histograms have a common occurrence into the bin k). To to cope with the discretization problems where close angles can be spread over different bins, we extend the direction distribution limits by one bin. If no direction is found for all feature vector ρ, the RHPM is considered as being locally incoherent. After extracting the occurrences feature vector ρ for each pair of magnitudes, the union of all ρ vectors provide the main directions.

$$\Psi = \sum_{i=1}^{n} \rho^i \quad (2)$$

The number of occurrences for each direction within Ψ range is from 0 (low intensity) to 4 (high intensity) and characterize the importance of each direction (Figure 2-C). If no common directions between the four feature vector ρ are found, the RHPM is considered as being locally incoherent.

Despite the fact that a RHPM is considered as coherent, the filtering of local motion has not yet been completed. Indeed, if we consider a natural facial movement to be uniform during motion then the local facial motion should spread to other region neighbors. The analysis of the movement propagation in the RHPM neighborhood is explained further.

## 3.2 RHPM neighborhood propagation

Facial muscles action ensures that a local motion spreads to neighboring regions until motion exhaustion. Motion is subject to changes that could affect direction and magnitude in any location. However, intensity of moving facial region tends to remain constant during facial expression. Therefore, a pertinent motion computed in a RHPM appears, eventually with a lower or upper intensity, in at least one neighboring region.

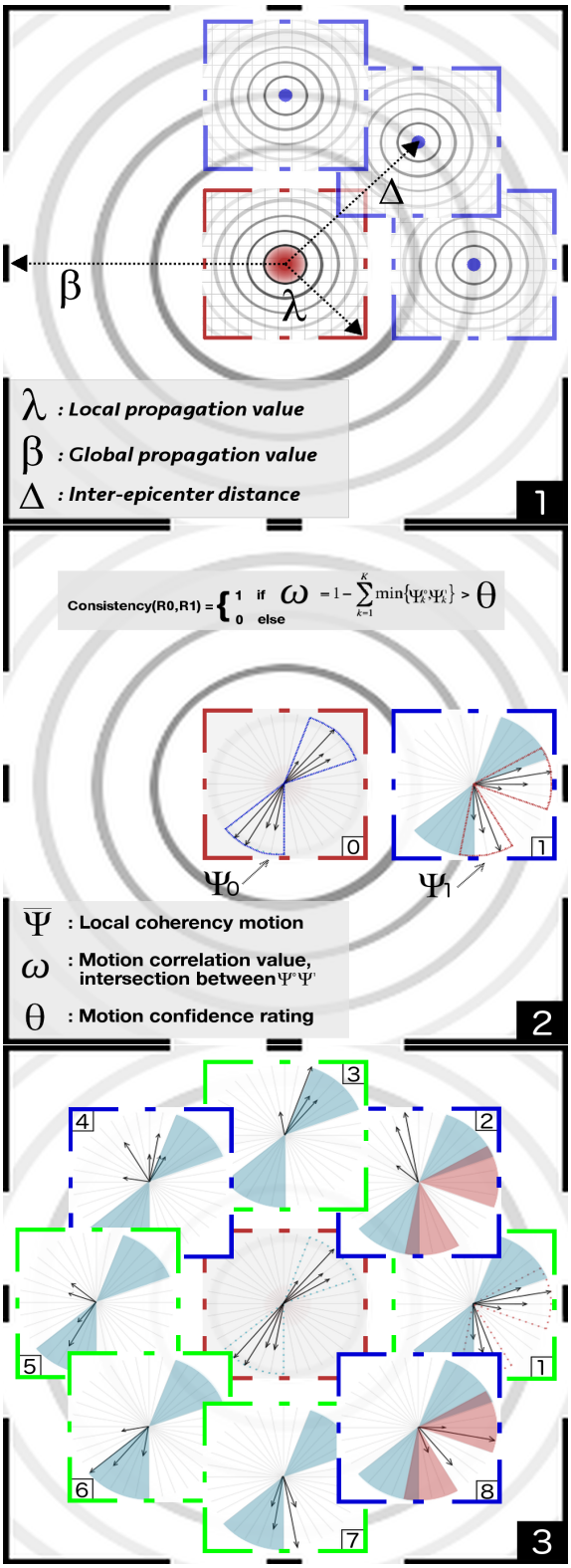


Figure 3: Estimate the motion propagation in the direct neighborhood of the specific RHPM (red square).

Facial motion analysis consists in estimating the motion propagation in the direct neighborhood of the specific RHPM. We propose a method to find the local facial motions that best discriminate expressions and corresponds to the regional importance of the expressive faces. The propagation analysis is illustrated in Figure 3. Next, we explain the process steps : how to locate RHPM Neighbouring regions (Figure 3-1); how to calculate the consistency between two region (Figure 3-2) and how to estimate the global consistent motion around the RHPM (Figure 3-3).

When an RHPM is locally coherent, we must verify that the motion has expanded into a neighboring RHPM. The propagation motion analysis is illustrated in Figure 3-1. The neighboring RHPM regions (represented by blue square) are regions with a high probability of propagation (RHPP). It is expected to measure a consistent motion between a region and its neighborhood. Eight RHPP are generated around the RHPM. All these regions are at a distance $\Delta$ from the RHPM epicenter. The bigger distance between two epicenter, the less coherence the overlapping area may exhibit. $\lambda$ is the size of the area under investigation around the epicenter. Finally, $\beta$ characterize the number of direct propagation from the epicenter that is carried out by the propagation analysis. The impact of these three parameters on the quality of filter will be detailed in the section 4.

Each RHPP is analyzed in order to evaluate the local coherency of the initial RHPM as illustrated in Figure 3-2. As an outcome of the process, each locally consistent RHPP is characterized by a directional vector $\Psi$ containing 36 bins (10° wide) of different magnitudes. Here the magnitudes correspond to the number of occurrences of a given orientation at different movement intensity scales ($M^1$ to $M^5$). The RHPM is considered to be consistent with its RHPP if a confidence rating $\omega$ exceeds a fixed percentage threshold $\theta$. $\omega$ is computed as follow :

$$\omega = 1 - \sum_{k=1}^{k} \min(\Psi_k^0, \Psi_k^1) \qquad (3)$$

where $\omega$ correspond to the intersection between two neighboring region directional vector $\Psi^0$ and $\Psi^1$ and k = 1,2,...,36 is the index of the bin. Next, recursively, for each inter-coherent RHPP we conduct the same inter-region coherency measurements as long as at least one nearly created RHPP is inter-region coherent with neighbor the previous one. The recursive process ends when the value $\beta$ is reached.

The motion propagation after one iteration is given in Figure 3-3. RHPP are represented with green borders if the motion is coherent with the RHPM. Otherwise, RHPP are represented with blue borders.

When the motion between two neighbors region is considered as coherent, a binary coherency map is updated in order to keep track of the evaluation procedure and avoid cycles. However, local region that are marked as non inter-region coherent, may be re-evaluated as coherent with an other RHPP in subsequent propagation. This is especially true in presence of skin wrinkles or furrows because motion discontinuities appears. This is the case for the RHPP n°8 in Figure 3-3. In the first iteration the distribution is not consistent directly with the original RHPM (corresponds to the blue area into the RHPP n°8). However, in the next iteration, this direction region is considered consistent with the RHPP n°1 (correspond to the red area into the RHPP n°8) which itself is consistent with the original RHPM.

Finally, each distribution vector ($\Psi$) corresponding to the RHPPs that have direct or indirect connections to the original RHPM (e.g. at least once motion is consistent between 2 neighbor regions) characterize the global region motion. If the motion propagation between all neighbors is inconsistent, the propagation motion is no more explored and that means that there are no more pertinent motions into the region. The global region motion is extracted by applying the following formula

$$\eta = \sum_{i=1}^{n} \Psi^i \tag{4}$$

Where n is the number of consistent regions (the RHPM and all consistent RHPR). $\eta$ is a histogram over 36 bins, which contains, for each bin the sum of each intensity of coherent RHPP. The maximum value for each bin correspond to the number of consistent regions $n$ multiplied by the high intensity of motion in $\Psi$, that is 4. Therefore, at this stage of the process, we are able to calculate the coherent propagation motion defined by an oriented histogram $\eta$ from a specific location.

In the next, we study the impact of RHPM location on the face. More specifically, we show that the intensity of expression (full or micro) plays a key role in the positioning of RHPM and, in the same time, it impacts the way to extract the consistent motion on the face.

### 3.3 Impact of RHMP location in face

The intensity of full-expression is more accentuate and the motion propagation covers a large facial area. If one RHPM is randomly placed in this area, then the motion consistency will always be respected and retrieved. However, with micro-expression, the motion propagation covers a restricted facial area. As the motions are less intense, the motion propagation is discontinued in large face areas, and that causes a disruption of movement between facial areas relatively close. This means that the locations of RHPM need to placed attentively at specific location when micro-expressions are under study. The Figure 4 shows the similar consistent motion extracted from a happy sequence computed from three different locations (columns 1,2 and 3) further and further away from the lips corner. The distribution of motion is computed from each epicenter (red point). Although, in the line 1 (full expression), the location of each epicenter is different, the distributions present large overlaps (column 4). However, in the line 2 (micro-expression), the distribution corresponding to column 3 is completely different. This shows the importance of locating in an adequate manner the epicenter while working with micro-expression.

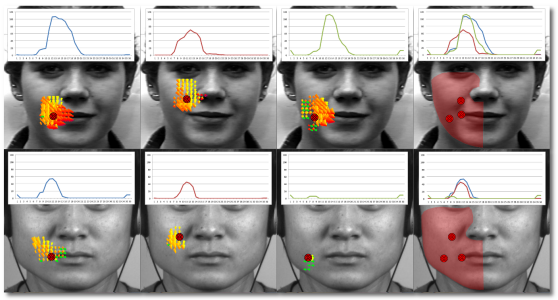

Figure 4: Consistent motion from a happy sequence computed from different locations in the same region

The next step consists in finding the local facial motions that accurately characterize the coherent motion on the face and best discriminate expressions. This informs about the importance of each region for the expression recognition process, and where to place the RHPM in the face to extract the consistent motion for full and micro expressions.

## 4 EXPRESSION RECOGNITION

In the following, we explore the integration of the coherent Optical Flow into a facial model formulation and discuss several strategies for considering discriminant local regions on the face. The first step consists in detecting pertinent motions which generates discriminant features to separate the six basic expressions effectively and the micro-expressions. Next, a vector is constructed which express the relationships between facial region of motion and full and micro expressions.

## 4.1 Best discriminant facial region

To identify the location on the face with the highest probability of movement, many RHPM are placed at regular intervals on the aligned face and the consistent motion vector of each RHPM is computed. In consequence, a consistent motion mask as well as motion information is extracted for each video sequence. Next, each consistent motion mask is normalized and merged to form a heat map of motion for the underlying expression. The six consistent motion masks for the basic expressions illustrated in Figure 5. They are computed from the sequences available in CK+ database.

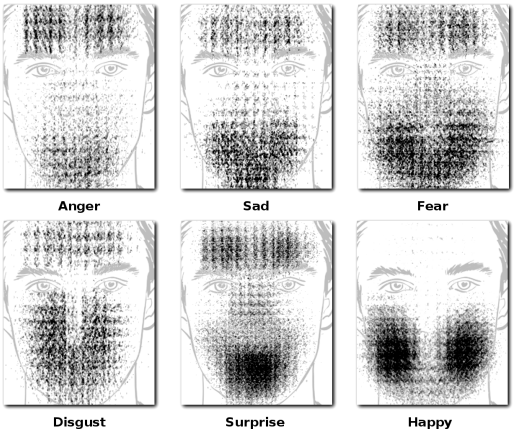


Figure 5: Pertinent motions to separate the six basic expressions effectively on CK+ database.

The extracted mask indicates that the pertinent motions are located below the eyes, in the forehead, around the nose and mouth, as illustrated in Figure 6. Some facial motions are located in the same place during elicitation for several expressions, but they are distinguishable by their intensity, direction and density. For example, Anger and Sad motion mask are similar because the main motion appears around the mouth and the eyebrows. However, when a person is angry, facial motions are convergent (e.g the mouth upwards and the eyebrows downwards) and facial motions are divergent when a person is sad. The facial areas which are active during different facial expressions are extensively studied in (Zhong et al., 2012). The results of the study match with our consistency map.

The same search strategy for finding the best discriminate regions for full expressions in CK+ was used in CASME2 for the micro-expressions. As illustrated in Figure 6, the pertinent motions are located near the eyebrows and the lips corner. If we compare that with the full expression motion maps, we see that the propagation distances are highly reduced. It is important to note that the motion map corresponding to the "other" class that is very close to others class, which doesn't facilitate good recognition decisions.

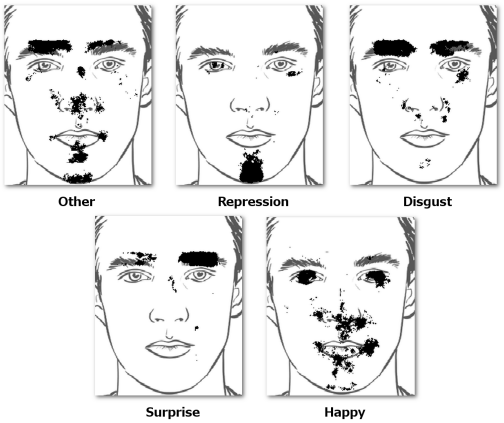


Figure 6: Pertinent motions to separate the micro-expressions on CASME2 database.

At this stage, the main facial regions of motion are accurately identified. We now construct a vector which express the relationships between facial region of motion and full and micro expressions.

## 4.2 Facial motion descriptor

We use the facial landmarks to define a facial region that increase facial deformation robustness during expression. Similarly to Jiang et al. (Jiang et al., 2014), the facial landmarks are used to define a mesh over the whole face, and a feature vector can be extracted from each of the regions enclosed by the mesh. To extract these facial meshes from face images, the facial landmarks are located with the method proposed by Kazemi et al. (Kazemi and Sullivan, 2014). Next, landmark positions and the geometrical statistics of the face are used to compute a new set of points that allow to define a mesh over the whole face (forehead, cheek). Finally, the best discriminant landmarks points are selected from original landmarks corresponding to the active face regions and specific points are computed in order to set out the mesh boundaries. The partitioning of facial regions of interest (ROIs) is illustrated in the Figure 7. The partitioning of these ROIs is based on the facial motion observed in the previous consistency maps extracted from both full and micro-expressions. The locations of these ROIs are uniquely determined by the landmarks points. For example, the position of the feature point $f_Q$ is the average of positions of two feature points, f10 and f55. The distance between the eyebrows and the forehead feature points ($f_A$,$f_B$,...,$f_F$) correspond to the size of the nose $Distance_{f27,f33}/4$ which makes it possible to

maintain the same distance for optimal adaptation to the size of the face.

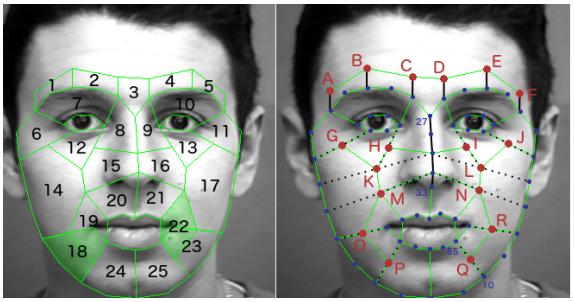


Figure 7: The partitioning of facial regions of interest.

The facial motion mask is computed from these 25 ROIs. The method used to build the feature vector from the facial motion mask is illustrated in the Figure 8. In each frame $f_i$, we consider the filtered Optical Flow inside each ROI $R_i^k$, where $i$ is the index of frames and $k = 1, 2, ..., 25$ is the index of ROIs. Inside each $R_i^k$, a histogram ($\eta$) is computed as defined in equation 4 from the Optical Flow filtered considering the ROI as initial RHPM. Overtime, for each ROI, the histograms are summed as defined in equation 5, which correspond to local facial motion of the entire sequence of facial motion.

$$\zeta(R^k) = \sum_{i=1}^{n} \eta_i^k(R_i^k) \tag{5}$$

Finally, all histograms $\zeta$ are concatenated into one-row vector, which is considered as the feature vector for the full and micro expression $\overline{\zeta} = (\zeta^1, \zeta^2, ..., \zeta^n)$. An example is illustrated in the Figure 8, where all histograms corresponding to the $R_i^1$ and $R_i^{22}$ with $i \in [1, n]$ are summed as defined in equation 5 in $\zeta^1$ and $\zeta^{22}$ respectively then added to $\overline{\zeta}$.

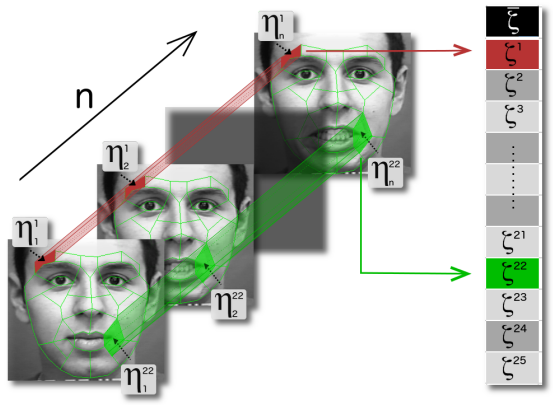


Figure 8: Method for building the feature vector from the facial motion mask.

The features vector size is equal to the number of ROI multiplied by the number of bins, making a total of 900 features values.

# 5 EVALUATION

In this section, we evaluate the performance of our proposed method on two datasets : (the extended Cohn-Kanade database (Lucey et al., 2010) and CASME2 (Yan et al., 2014). We discuss the choice of optimal parameters for the databases and show that only the magnitude intervals must be adapted to accommodate the specificities of intensity of facial expression. Finally, we compare our performance against major state-of-the-art approaches.

## 5.1 Full-expression

CK+ contains 410 facial expression sequences from 100 participants coming from different ethnicities and genders. In these image sequences, the expression starts from a neutral status and ends in the apex status. The number of samples for the following expressions, i.e. anger, sadness, happiness, surprise, fear and disgust are 42, 82, 100, 80, 64 and 45, respectively.

In the experiments, we use LIBSVM (Chang and Lin, 2011) with the Radial Basis Function kernel and the 10 fold cross-validation protocol. This protocol is used by several approaches working on CK+ as it fits better to the size and the structure of the data set. Each expression is classified into one of the six classes : anger, fear, disgust, happiness, sadness, and surprise.

The following experimental results are obtained using $\lambda = 15$, $\beta = 3$, $\Delta = 10$. Initially, we considered the following magnitude intervals in every region: $M1(x)|x \in [1, 10]$, $M2(x)|x \in [2, 10]$, $M3(x)|x \in [3, 10]$, $M4(x)|x \in [4, 10]$, $M5(x)|x \in [5, 10]$. Each interval stops at a maximum of 10, where 10 corresponds to the mean of the max of coherent magnitude estimated from all sequences. The overlap of intervals allow to ensure consistency in each histogram. Small movements around the mouth corners and between the eyes were not always detected and we included the magnitude $M_0$ and delete the magnitude $M_5$ to retain only 5 intervals of magnitudes for the corresponding regions $R_3, R_{19}, R_{22}$.

We compare the performance of the different ways to position the RHPMs on the face and various data normalization techniques prior to coherent facial motion extraction. RHPM are applied in different localization of the face: a) Block-based RHPM is implemented by partitioning each frame of the video into 10x10 non-overlapping blocks then place an RHPM

in the center of each block (Grid); b) On each facial landmarks (Lands); c) On the center of our facial mask (Mask). Experiments were conducted either on : i) raw data (without normalization); ii) data normalized after applying a face alignment based on eyes from the first picture; iii) or data normalized by means of facial registration based on facial landmarks.

Table 1 compares the performance of the different approaches to position the RHPMs (Grid, Landmarks and Mask) are explored to compute the facial motion and classify expressions. Column 4 (Geom) presents results obtained by considering only the geometric information inferred by a mesh generated from the landmarks positions and column 5 (Mask+Geom) reports on results combining geometrical and motion information.

Table 1: Performance Comparison of Different Approaches.

| Norma. | Grid | Lands | Mask | Geom. | Mask+Geom |
|---|---|---|---|---|---|
| No | 87,31% | 84,14% | 92.68% | 86.58% | 92.92% |
| Eyes | 86,58% | 83,41% | 93.17% | 85.85% | 95.34% |
| Shape | 82,19% | 82,92% | 85,85% | 87.56% | 88.53% |

Alignment based on eyes obtains the best performances in CK+ because only translation and in-plane rotation occur. Finally, the normalization based on shape becomes less efficient with Optical Flow because the feature points are affected by the actions of various expressions. Without normalization, the results are still correct due to limited head motion. This result shows the performance of the consistent motion filter in presence of small head motions. Motion computed around landmarks report worse performance than the other methods, demonstrating the importance of cheek, chin and forehead regions that are not captured directly by landmarks. Assuming the face region is well aligned, the motion calculated from equal-sized facial grids is better than landmarks in presence of facial deformation during expression. Concerning the Mask, it shows better results and prove that facial models are more appropriate to recognize facial expressions.

Appearance based methods are reported to outperform geometry based methods (Whitehill et al., 2008). However, as suggested in (Zhang and Ji, 2005; Kotsia et al., 2008), the combination of the geometric and appearance features can provide additional information to the recognition process. We computed geometric features by exploiting the size and shape of the facial ROIs. Specifically, features that can be extracted from the facial mask include the length and orientation of facial ROI (Geom). A combination of the geometric and appearance features are computed with the improved version of the RHPM features and the geometric features from the facial ROI (Mask+Geom). As highlighted in the Table 1, geometric features are not as competitive as appearance features. However, the combination of the geometric and appearance features slightly increase the recognition rate.

Table 2 compares the performance of the proposed method with the recent state-of-the-art Optical Flow methods on CK+. The performance of the our system is comparable with the other systems as it achieved an average recognition rate of 93.17% with alignment based on eyes and coherent Optical Flow. Nevertheless, the highest recognition rate is obtained using features from the filtered coherent facial motion combined with geometric features.

Table 2: Performance Comparison of Different State-of-the-Art Optical Flow Approaches on CK+ Database. The bold means our proposed methods.

| Method | Measure | Seq. | Exp. | Acc(%) |
|---|---|---|---|---|
| (Liao et al., 2013) | LOSO | 442 | 6 | 92,5% |
| (Su et al., 2007) | train/test | 415 | 5 | 93,27% |
| (Lee and Chellappa, 2014) | 4-fold | ndef | 7 | 86,7% |
| **Coherent Flow + RHPM** | **10-fold** | **410** | **6** | **93.17%** |
| **Coherent Flow + RHPM + Geom.** | **10-fold** | **410** | **6** | **95.34%** |

Our method reported comparable recognition performance with the most competitive Optical Flow approaches. Although we report the best accuracy results for 6-class expressions, Su et al. (Su et al., 2007) achieve higher scores, but they only use 5 classes (Neutral, Happy, Surprise, Sad, Anger). Considering the variations observed in the number of sequences and expression types recognized by the various methods it is difficult to clearly identify the best one. For our experiments we used the original CK+ collection as introduced in (Lucey et al., 2010) and we brought no modification neither to the videos nor the annotations. As presented in Table 1, our approach does not require normalized images in order to obtain good performances. This is not the case for above-cited papers. Indeed, the coherent motion accumulation over time allows ignoring possible disruptions caused by small head motions and illumination variation in some frames. The face normalization brings performances gains in the recognition process because some videos contain a series of disruptions, which causes significant deterioration over time.

We have shown that our approach obtains good performances in CK+, where the expressions are acted and there is little or no disruptions (head motion, lightning variation). Next, we present the performance of our method to recognize micro-expressions.

## 5.2 Micro-expression

The CASME2 database contains 246 spontaneous micro-expressions from 26 subjects, categorized into five classes: happiness (32 samples), disgust (63 samples), surprise (25 samples), repression (27 samples) and Others (99 samples).

In the experiments, we use leave-one-subject-out (LOSO) cross validation protocol because it is more strict than leave-one-out (LOO) cross validation and matches better the structure of the data (different number of videos par subject). In this protocol, the samples from one subject are used for testing, the rest for training. We use the LIBSVM (Chang and Lin, 2011) with the Radial Basis Function and the grid search method, where the optimal parameter is provided using ten-fold cross validation.

Table 3 shows a comparison to some other approaches for micro-expression using the recognition rates given in each paper. It should be noted that the results are not directly comparable due to different experimental setups (number of expression classes and number of sequences), but they still give an indication of the discriminating power of each approach. Our method outperforms the other methods in almost all cases. The best results are obtained using the same mask and parameters as for full-expression recognition ($\lambda = 15, \beta = 3, \Delta = 10$) except for the division of magnitudes defined here as follows : $M1(x)|x \in [0.1,5]$, $M2(x)|x \in [0.2,5]$, $M3(x)|x \in [0.3,5]$, $M4(x)|x \in [0.4,5]$, $M5(x)|x \in [0.5,5]$. The geometric information was not considered here, as the landmarks locations are mostly stables throughout the sequence. It should be noted that the Optical Flow is not calculated from two consecutive frame but on two frame intervals. Indeed, the time lapses between two frame in CASME2 is so small (recorded with high-speed camera (at 200 fps)) and combined with the low expression intensity it is difficult not make a distinction between the noise and the true facial motion. No magnitude consistency can be found in local region with our method when consecutive frames are processed. Hence, we are considering the entire sequence, but this is frequent in the literature as other authors summarize videos in fewer frames (Wang et al., 2014c; Huang et al., 2016a; Huang et al., 2016b).

In order to better understand the limitation of our approach, we build a confusion matrix. Looking at the confusion matrix in Figure 9, Happiness is often confound with Other class. It may be explained by the fact that Other class includes some confused micro-expressions similar to others as illustrated in Figure 6. If the recognition process is re-evaluated on a four classes basis (Happy, Disgust, Surprise, Repression), the performance is improved by 11.57%, which corresponds to an accuracy of 76.92%. This proves that the Other class does not stand out clearly from others. In (Liu et al., 2015), the repression and the other sequences are combined in a single class, which reduces the chances of falsely classification of Happiness to Repression class. This new organization reported a gain of 1.02% with our method. Moreover, (Liu et al., 2015) reported on removing 11 samples in the recognition process due to mis-estimates of the facial features in the first frame of the video.

Table 3: Performance comparison with the state-of-the-art methods on CASME2 database. Results in bold correspond to our method.

| Method | Measure | Class | Acc(%) |
|---|---|---|---|
| Baseline (Yan et al., 2014) | LOO | 5 | 63.41% |
| LBP-SIP (Wang et al., 2014c) | LOO | 5 | 67.21% |
| LSDF (Wang et al., 2014b) | LOO | 5 | 65.44% |
| TICS (Wang et al., 2014a) | LOO | 5 | 61.76% |
| MDMO (Liu et al., 2015) | LOSO | 4 | 67.37% |
| STCLQP (Huang et al., 2016b) | LOSO | 5 | 58.39% |
| STLBP-IIP (Huang et al., 2016a) | LOSO | 5 | 62.75% |
| DiSTLBP-IPP (Huang et al., 2016a) | LOSO | 5 | 64.78% |
| **Coherent Flow + RHPM** | **LOSO** | **5** | **65.35%** |

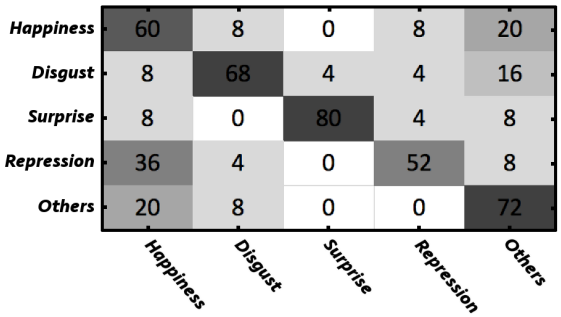


Figure 9: The confusion matrix for five micro-expression categorizations on CASME2 database.

In the annotations provided by CASME2, we can see that each class is characterized by specific action units. More specifically, if we analyze the average action units distribution of each class, we obtain the following division : Happy (AU6, AU12), Surprise (AU1, AU2), Disgust (AU4, AU7, AU9), Repression (AU15,AU17) and Others (AU4). If we are referring strictly to action units instead of the provided expression annotations, all sequences from the Other category including AU4 may be considered as Disgust sequences. In doing so, we obtain a new distribution of the 246 spontaneous micro-expressions: happiness (34 samples), disgust (128 samples), sur-

prise (25 samples), repression (27 samples) and Others (32 samples). Based on the action units, the recognition rate is improved by 7.09%, which corresponds to an accuracy of 72.44%. A synthesis on different CASME2 configurations is illustrated in the Table 4.

Table 4: Performance Comparison of Different data Segmentation on CASME2

| Details of segmentation | classes | Seq. | Accuracy |
|---|---|---|---|
| Based on original data | 5 | 246 | 65.35% |
| Combine Repression and Other classes | 4 | 246 | 66.37% |
| Based on Action Units | 5 | 246 | 72.44% |
| Delete the Other class | 4 | 147 | 76.92% |

The results obtained on the original CASME2 and the reorganized variants show the good performances for micro-expressions recognition. Our method outperforms the other state-of-the-art methods in almost all cases. We have discussed about issues related to the ambiguous annotations of the Others category in CASME2, which further reduces the recognition rate and it may make sense to use the action units to annotate the database. These results were obtained by employing the same method used for recognizing full-expressions, except for, smaller magnitude intervals that were considered in order to fit better to low magnitudes in micro-expressions.

# 6 CONCLUSIONS

In the paper, we have shown that the coherent movement extracted from dense Optical Flow method by considering the facial movement hypothesis achieves state-of-the-art performance on both facial full-expression and micro-expression databases. The magnitude and direction constraints are estimated in order to reduce the noise induced by lighting changes and small head motions over time. The proposed approach adapts well on both full-expressions (CK+) and micro-expressions (CASME2). The only adjustment concerning the magnitude intervals is actually related to the nature of expression. The other parameters common to both experiences have been selected empirically and deserve specific attention in future experiments.

Our current approach is used only in near-frontal-view recordings where the presence of occlusions, fast head motion and lightning variation is practically zero. The next step consist in adapting our method to the domain of spontaneous facial expression recognition. To address this situation, a normalization method will be necessarily used. However, it must be kept in mind that any change made in the facial picture has important side-effects on the Optical Flow. Despite the wealth of research already conducted, no method is capable of dealing with all issues at a time. We believe that the normalization approaches based on facial components or shape are not adapted to Optical Flow as facial deformation will impact Optical Flow computation by inducing motion distortion. So rather than considering the normalization in the field of facial components, efforts should instead be focused on the Optical Flow domain.

# REFERENCES


Bailer, C., Taetz, B., and Stricker, D. (2015). Flow fields: Dense correspondence fields for highly accurate large displacement optical flow estimation. In *ICCV*, pages 4015–4023.

Butler, D. J., Wulff, J., Stanley, G. B., and Black, M. J. (2012). A naturalistic open source movie for optical flow evaluation. In *ECCV*, pages 611–625. Springer.

Chang, C.-C. and Lin, C.-J. (2011). Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27.

Chen, Q. and Koltun, V. (2016). Full flow: Optical flow estimation by global optimization over regular grids. *CVPR*.

Farnebäck, G. (2003). Two-frame motion estimation based on polynomial expansion. In *Scandinavian conference on Image analysis*, pages 363–370. Springer.

Fortun, D., Bouthemy, P., and Kervrann, C. (2015). Optical flow modeling and computation: a survey. *Computer Vision and Image Understanding*, 134:1–21.

Han, S., Meng, Z., Liu, P., and Tong, Y. (2014). Facial grid transformation: A novel face registration approach for improving facial action unit recognition. In *ICIP*, pages 1415–1419.

Huang, X., Wang, S., Liu, X., Zhao, G., Feng, X., and Pietikainen, M. (2016a). Spontaneous facial micro-expression recognition using discriminative spatiotemporal local binary pattern with an improved integral projection. *CVPR*.

Huang, X., Zhao, G., Hong, X., Zheng, W., and Pietikäinen, M. (2016b). Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns. *Neurocomputing*, 175:564–578.

Jiang, B., Martinez, B., Valstar, M. F., and Pantic, M. (2014). Decision level fusion of domain specific regions for facial action recognition. In *ICPR*, pages 1776–1781.

Kazemi, V. and Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. In *CVPR*, pages 1867–1874.

Kotsia, I., Zafeiriou, S., and Pitas, I. (2008). Texture and shape information fusion for facial expression and facial action unit recognition. *Pattern Recognition*, 41(3):833–851.

Lee, C.-S. and Chellappa, R. (2014). Sparse localized facial motion dictionary learning for facial expression recognition. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3548–3552.

Li, X., Pfister, T., Huang, X., Zhao, G., and Pietikäinen, M. (2013). A spontaneous micro-expression database: Inducement, collection and baseline. In *FG*.

Liao, C.-T., Chuang, H.-J., Duan, C.-H., and Lai, S.-H. (2013). Learning spatial weighting for facial expression analysis via constrained quadratic programming. *Pattern Recognition*, 46(11):3103–3116.

Liu, Y.-J., Zhang, J.-K., Yan, W.-J., Wang, S.-J., Zhao, G., and Fu, X. (2015). A main directional mean optical flow feature for spontaneous micro-expression recognition. *Affective Computing*.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. (2010). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *CVPR Workshops*, pages 94–101.

Péteri, R. and Chetverikov, D. (2005). Dynamic texture recognition using normal flow and texture regularity. In *Iberian Conference on Pattern Recognition and Image Analysis*, pages 223–230.

Revaud, J., Weinzaepfel, P., Harchaoui, Z., and Schmid, C. (2015). Epicflow: Edge-preserving interpolation of correspondences for optical flow. In *CVPR*, pages 1164–1172.

Su, M.-C., Hsieh, Y., and Huang, D.-Y. (2007). A simple approach to facial expression recognition. In *Proc. of the WSEAS Conference on Computer Engineering and Applications*, pages 456–461.

Wang, S., Yan, W.-J., Li, X., Zhao, G., and Fu, X. (2014a). Micro-expression recognition using dynamic textures on tensor independent color space. In *ICPR*, pages 4678–4683.

Wang, S.-J., Yan, W.-J., Zhao, G., Fu, X., and Zhou, C.-G. (2014b). Micro-expression recognition using robust principal component analysis and local spatiotemporal directional features. In *ECCV Workshop*, pages 325–338.

Wang, Y., See, J., Phan, R. C.-W., and Oh, Y.-H. (2014c). Lbp with six intersection points: Reducing redundant information in lbp-top for micro-expression recognition. In *Asian Conference on Computer Vision*, pages 525–537.

Whitehill, J., Bartlett, M., and Movellan, J. (2008). Automatic facial expression recognition for intelligent tutoring systems. In *CVPR Workshops*.

Yan, W.-J., Li, X., Wang, S.-J., Zhao, G., Liu, Y.-J., Chen, Y.-H., and Fu, X. (2014). Casme ii: An improved spontaneous micro-expression database and the baseline evaluation. *PloS one*, 9(1).

Zhang, Y. and Ji, Q. (2005). Active and dynamic information fusion for facial expression understanding from image sequences. *PAMI*, 27(5):699–714.

Zhong, L., Liu, Q., Yang, P., Liu, B., Huang, J., and Metaxas, D. N. (2012). Learning active facial patches for expression analysis. In *CVPR*, pages 2562–2569.