

Invited Talk: Coupling Deformable Models and Learning Methods for Nonverbal Behavior Analysis: Applications to Deception, Multi-cultural Studies and ASL^{*}

Dimitris Metaxas

Distinguished Prof. of CS, Director CBIM Center, Rutgers University
`dmm@cs.rutgers.edu`

Abstract. Based on recent advances in deformable model tracking theory, we have developed a novel system for real-time facial and gesture tracking and action recognition. In particular, our face tracker by using deformable statistical models that encode facial shape variation and local texture distribution, it robustly tracks 79 facial landmarks, which correspond to facial components such as the eyes, eyebrows, nose, and mouth. The model initializes automatically, tolerates partial occlusions, detects and recovers from lost track. Moreover, it handles head rotations of -90° to 90° in any direction by using manifold embedding methods. During online tracking, the model dynamically adapts to the facial shape of the current subject and temporal filters stochastically smooth the target's position. Tracked landmarks are then used by our learning modules for feature extraction and event recognition. In order to speed up convergence to the optimal landmark configuration, the system employs multi-resolution model fitting. To further reduce computational complexity, we track landmarks in successive frames using a Sum of Squared Differences point tracker and running the relatively "expensive" step of face search only periodically to prevent any error accumulation. This scheme allows us to have a measure of tracking success (confidence) for each landmark, so that we can detect early on if we are beginning to drift from the target, in which case we immediately invoke the deformable fitting algorithm to self-correct the result. Similarly, we have developed a skin blob tracker for tracking the orientation, position, velocity and area of head and hand blobs, which is automatically initialized with a generic skin color model, dynamically learning the specific subject's color distribution online for adaptive tracking. Detected blobs are filtered online, both in terms of shape and motion, using eigenspace analysis and temporal dynamical models to prune false detections.

We apply this framework to three different recognition applications. First, we use the tracked facial landmarks to crop the face region and extract appearance features, which are used to learn models that detect universal facial expressions (i.e., sadness, anger, fear, disgust, happiness

^{*} Collaborators: Nicholas Michaels, Fei Yang, Peng Yang (Rutgers), Carol Neidle (BU).

and surprise). In particular, our method utilizes the relative intensity ordering of facial expressions (i.e. neutral, onset, apex, offset) found in the training set to learn a ranking model (Rankboost) for recognition and intensity estimation, which improves our average recognition rate (87.5% on the CMU benchmark database). Second, we use the tracked landmarks and blobs to compute derived features (e.g., features characterizing posture openness, asymmetrical facial expressions, etc.) and recognized gestures (e.g., head touching, hands together, eye blinking, etc.). Using these features with discriminative learning methods, we train subject-specific models to detect when subjects from various cultures are deceptive or not in an interview scenario of a mock crime (12 responses per subject) as well as to identify cultural gestures. Using Leave-One-Out-Cross-Validation (LOOCV) we achieved average deception detection accuracy (percentage of correctly tagged responses) of 81.6% for 147 subjects. Third, we apply our tracking and learning methods to track signers of American Sign Language (ASL) and recognize gestures and expressions which have grammatical meaning. In particular, by tracking eyebrow movements, eye aperture changes, head tilts, head nods and head rotations, we can recognize wh-question markers, topic markers, and negation, using generative temporal models and spectral embedding methods to reduce feature dimensions and uncover the manifold separation (average accuracy 84% in continuous sequences).

References

1. Yang, P., Liu, Q., Metaxas, D.N.: Exploring Facial Expressions with Compositional Features. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR (2010)
2. Yang, P., Zhong, L., Metaxas, D.N.: Ranking Model for Facial Age Estimation. In: 20th International Conference on Pattern Recognition, ICPR (2010)
3. Yang, P., Liu, Q., Metaxas, D.N.: Boosting Encoded Dynamic Features for Facial Expression Recognition. *Pattern Recognition Letter* (2009)
4. Yang, P., Liu, Q., Metaxas, D.N.: RankBoost with l1 regularization for Facial Expression Recognition and Intensity Estimation. In: Proc. of Int. Conf. on Computer Vision, ICCV (2009)
5. Neidle, C., Michael, N., Nash, J., Metaxas, D.N.: A Method for Recognition of Grammatically Significant Head Movements and Facial Expressions, Developed through use of a Linguistically Annotated Video Corpus. In: Proceedings of the Workshop on Formal Approaches to Sign Languages, Held as Part of the 21st European Summer School in Logic, Language and Information, Bordeaux, France, July 20-31 (2009)
6. Michael, N., Metaxas, D.N., Neidle, C.: Spatial and Temporal Pyramids for Grammatical Expression Recognition of American Sign Language. In: Proceedings of the Eleventh International ACM SIGACCESS Conference on Computers and Accessibility, Philadelphia, PA, October 26-28 (2009)
7. Burgoon, J.K., Jensen, M.L., Twyman, N.W., Meservy, T.O., Metaxas, D.N., Michael, N., Nunamaker Jr., J.F.: Automated Kinesic Analysis for Deception Detection. In: 43rd Hawaii International Conference on Systems Sciences (HICSS 2010) (Proceedings of the Credibility Assessment and Information Quality in Government and Business Symposium) (2010)

8. Michael, N., Neidle, C., Metaxas, D.N.: Computer-based recognition of facial expressions in ASL: from face tracking to linguistic interpretation. In: Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, LREC, Malta, May 22-23 (2010)
9. Michael, N., Dilsizian, M., Metaxas, D., Burgoon, J.K.: Motion Profiles for Deception Detection Using Visual Cues. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part VI. LNCS, vol. 6316, pp. 462–475. Springer, Heidelberg (2010)
10. Huang, J., Zhang, T., Metaxas, D.N.: Learning with Structured Sparsity. In: The 26th International Conference on Machine Learning, ICML 2009, Montreal, Quebec, Canada (June 2009)