Supplementary Material:
Can facial pose and expression be separated with weak perspective camera?

Appendix A

Here we prove Theorem 3.1 of the main paper, which for convenience we copy below.

Theorem 3.1. *Suppose we have 3D facial points with a neutral expression,* $\{\bar{\mathbf{X}}_i\}_{i=1}^N$, *and 2D image points* $\{\mathbf{x}_i\}_{i=1}^N$ *corresponding to those 3D points but with a rotation. Let* $\mathbf{R}^*$ *and* $\boldsymbol{\sigma}^*$ *minimize* $J_0(\mathbf{R}, \boldsymbol{\sigma})$, $\mathbf{r}$ *be defined as in (9), and* $\mathbf{B}\in\mathbb{R}^{3N\times M}$ *be a matrix such that rank* $\left(\dot{\mathbf{W}}(\boldsymbol{\sigma}^*)\dot{\mathbf{R}}^*\mathbf{B}\right) = M < 2N$. *Then,*

$$\min_{\mathbf{R},\mathbf{e},\boldsymbol{\sigma}} J_{\mathbf{B}}(\mathbf{R}, \mathbf{e}, \boldsymbol{\sigma}) \le \min_{\mathbf{R},\boldsymbol{\sigma}} J_0(\mathbf{R}, \boldsymbol{\sigma}). \tag{S.1}$$

*Moreover, this inequality holds strictly (*i.e.*, without equality) if* $\mathbf{r} \notin$ *Null* $\left[\left(\dot{\mathbf{W}}(\boldsymbol{\sigma}^*)\dot{\mathbf{R}}^*\mathbf{B}\right)^T\right]$, *in which case* $||\mathbf{e}^*||>0$ *where* $\mathbf{e}^*$ *is the minimizer of* $J_{\mathbf{B}}(\mathbf{R}, \mathbf{e}, \boldsymbol{\sigma})$ *w.r.t. variable* $\mathbf{e}$.

Proof. Consider the function $J_{\mathbf{B}}(\mathbf{R}^*, \mathbf{e}, \boldsymbol{\sigma}^*)$, which can be treated as a single-variable function (of $\mathbf{e}$) as $\mathbf{R}^*$ and $\boldsymbol{\sigma}^*$ are fixed values. It is trivial to show that $J_{\mathbf{B}}(\mathbf{R}^*, \mathbf{e}, \boldsymbol{\sigma}^*)$ can be written as:

$$J_{\mathbf{B}}(\mathbf{R}^*, \mathbf{e}, \boldsymbol{\sigma}^*) = ||\mathbf{x} - \dot{\mathbf{W}}(\boldsymbol{\sigma}^*)\dot{\mathbf{R}}^*(\bar{\mathbf{X}} + \mathbf{B}\mathbf{e})|| \tag{S.2}$$
$$= ||\mathbf{r} - \dot{\mathbf{W}}(\boldsymbol{\sigma}^*)\dot{\mathbf{R}}^*\mathbf{B}\mathbf{e}||, \tag{S.3}$$

where $\mathbf{r}, \tilde{\mathbf{x}}, \bar{\mathbf{X}}, \dot{\mathbf{R}}^*$ and $\dot{\mathbf{W}}(\boldsymbol{\sigma}^*)$ are defined in Eq. (9) of the main paper and the text that follows it. Let us define the matrix $\mathbf{C}$ as the $2N \times M$ matrix

$$\mathbf{C} := \dot{\mathbf{W}}(\boldsymbol{\sigma}^*)\dot{\mathbf{R}}^*\mathbf{B}. \tag{S.4}$$

Then, (S.3) can be more compactly rewritten as

$$J_{\mathbf{B}}(\mathbf{R}^*, \mathbf{e}, \boldsymbol{\sigma}^*) = ||\mathbf{r} - \mathbf{C}\mathbf{e}||. \tag{S.5}$$

Since (S.5) involves the ℓ_2 norm and according to the assumption of Theorem 3.1 $\mathbf{C}$ is a skinny matrix with full column rank, the minimizer of (S.5) is [1]

$$\mathbf{e}^* := \arg_{\mathbf{e}} \min ||\mathbf{r} - \mathbf{C}\mathbf{e}|| = \mathbf{C}^\dagger\mathbf{r} \tag{S.6}$$

where $\mathbf{C}^\dagger$ is the Moore-Penrose inverse of $\mathbf{C}$. Thus, the minimal value that (S.5) can take is $\left|\left|\mathbf{r} - \mathbf{C}\mathbf{C}^\dagger\mathbf{r}\right|\right|$, which can be written as $\left|\left|(\mathbf{I} - \mathbf{C}\mathbf{C}^\dagger)\mathbf{r}\right|\right|$. Since $\mathbf{C}\mathbf{C}^\dagger$ is an orthogonal projection matrix, the matrix norm $\left|\left|\mathbf{I} - \mathbf{C}\mathbf{C}^\dagger\right|\right|$ is 1 [3]. Therefore, by definition of matrix norm [1], it holds that

$$\left|\left|\mathbf{r} - \mathbf{C}\mathbf{C}^\dagger\mathbf{r}\right|\right| = \left|\left|\left(\mathbf{I} - \mathbf{C}\mathbf{C}^\dagger\right)\mathbf{r}\right|\right|$$
$$\le \left|\left|\mathbf{I} - \mathbf{C}\mathbf{C}^\dagger\right|\right| \, ||\mathbf{r}|| = ||\mathbf{r}||. \tag{S.7}$$

Therefore, the following clearly holds

$$\begin{aligned}\min_{\mathbf{R},\mathbf{e},\boldsymbol{\sigma}} J_{\mathbf{B}}(\mathbf{R}, \mathbf{e}, \boldsymbol{\sigma}) &\le \min_{\mathbf{e}} J_{\mathbf{B}}(\mathbf{R}^*, \mathbf{e}, \boldsymbol{\sigma}^*) \\ &= \min_{\mathbf{e}} ||\mathbf{r} - \mathbf{C}\mathbf{e}|| \\ &= \left|\left|\mathbf{r} - \mathbf{C}\mathbf{C}^\dagger\mathbf{r}\right|\right| \\ &\le ||\mathbf{r}|| \\ &= J_0(\mathbf{R}^*, \boldsymbol{\sigma}^*) \\ &= \min_{\mathbf{R},\boldsymbol{\sigma}} J_0(\mathbf{R}, \boldsymbol{\sigma}). \end{aligned} \tag{S.8}$$

Thus, we proved the main statement (S.1) in Theorem 3.1. To prove the rest of the theorem, let us suppose, as specified in the theorem, that rank$(\mathbf{C})$ = rank$(\dot{\mathbf{W}}(\boldsymbol{\sigma}^*)\dot{\mathbf{R}}^*\mathbf{B}) = M$ and that $\mathbf{r} \notin$ Null$(\mathbf{C}^T)$. Note that $\mathbf{C}\mathbf{C}^\dagger$ is an orthogonal matrix that projects a vector on the column space of $\mathbf{C}$; therefore $\left|\left|\mathbf{r} - \mathbf{C}\mathbf{C}^\dagger\mathbf{r}\right|\right|$, which is the projection error, is always smaller than the norm of the projected vector $||\mathbf{r}||$, unless $\mathbf{r}$ is orthogonal to the column space of $\mathbf{C}$. The latter condition holds if and only if $\mathbf{r}$ is orthogonal to every column of $\mathbf{C}$, *i.e.* if $\mathbf{r}^T\mathbf{C} = \mathbf{0}$, or, equivalently, if $\mathbf{C}^T\mathbf{r} = \mathbf{0}$ (*i.e.*, if $\mathbf{r} \in$ Null$(\mathbf{C}^T)$). Thus, if $\mathbf{r} \notin$ Null$(\mathbf{C}^T)$, (S.7) holds strictly, and, as a result, inequality (S.8) also holds strictly:

$$\min_{\mathbf{R},\mathbf{e},\boldsymbol{\sigma}} J_{\mathbf{B}}(\mathbf{R}, \mathbf{e}, \boldsymbol{\sigma}) < \min_{\mathbf{R},\boldsymbol{\sigma}} J_0(\mathbf{R}, \boldsymbol{\sigma}).$$

We will now use proof by contradiction to prove that the minimizer of $J_{\mathbf{B}}(\mathbf{R}, \mathbf{e}, \boldsymbol{\sigma})$ w.r.t. $\mathbf{e}$, namely $\mathbf{e}^*$, satisfies the inequality $||\mathbf{e}^*|| > 0$ (or equivalently $\mathbf{e}^* \neq \mathbf{0}$) when rank$(\mathbf{C}) = M$ and $\mathbf{r} \notin$ Null$(\mathbf{C}^T)$. To establish a contradiction, let us suppose that

$$\min_{\mathbf{R},\boldsymbol{\sigma}} J_{\mathbf{B}}(\mathbf{R}, \mathbf{0}, \boldsymbol{\sigma}) \le \min_{\mathbf{e}} J_{\mathbf{B}}(\mathbf{R}^*, \mathbf{e}, \boldsymbol{\sigma}^*) \tag{S.9}$$

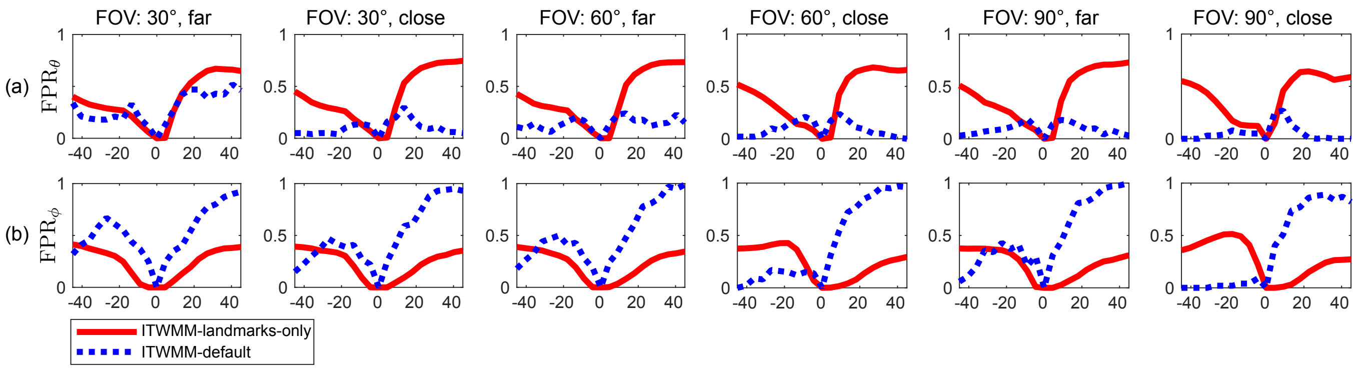

Figure S.1. The AU false positive rates (FPRs) against the rotation amount, shown separately for lower- and upper-face AUs (see Section **??** for various FOV and subject-to-camera distances. (a) Yaw rotation; (b) pitch rotation.

for $\mathbf{r} \notin \text{Null}(\mathbf{C}^T)$. But $J_\mathbf{B}(\mathbf{R}, \mathbf{0}, \boldsymbol{\sigma}) = J_\mathbf{0}(\mathbf{R}, \boldsymbol{\sigma})$, therefore (S.9) can be rewritten as

$$\min_{\mathbf{R},\boldsymbol{\sigma}} J_\mathbf{0}(\mathbf{R}, \boldsymbol{\sigma}) \leq \min_{\mathbf{e}} J_\mathbf{B}(\mathbf{R}^*, \mathbf{e}, \boldsymbol{\sigma}^*). \tag{S.10}$$

Also according to (S.8) and the argument that follows it, when $\mathbf{r} \notin \text{Null}(\mathbf{C}^T)$, we have that

$$\min_{\mathbf{e}} J_\mathbf{B}(\mathbf{R}^*, \mathbf{e}, \boldsymbol{\sigma}^*) < \min_{\mathbf{R},\boldsymbol{\sigma}} J_\mathbf{0}(\mathbf{R}, \boldsymbol{\sigma}). \tag{S.11}$$

But (S.10) contradicts (S.11), therefore our original supposition (S.9) cannot hold. In other words, the $\mathbf{e}$ that minimizes $J_\mathbf{B}(\mathbf{R}, \mathbf{e}, \boldsymbol{\sigma})$ cannot be zero when $\mathbf{r} \notin \text{Null}(\mathbf{C}^T)$, therefore $||\mathbf{e}^*|| > 0$.

□

Appendix B

We now discuss why the vector $\mathbf{r}$ defined in Eq. (9) is not likely to be in the nullspace of the transpose of the $2N \times M$ matrix $\mathbf{C} := \dot{\mathbf{W}}(\boldsymbol{\sigma}^*)\dot{\mathbf{R}}^*\mathbf{B}$ with rank M (see Appendix A). Recall that $\mathbf{r}$ can be interpreted as the minimal 3D-to-2D mapping error for the WP camera (Section 3); that is, the difference between the true 2D projection of a set of 3D points and their 2D projection according to the WP camera. First of all, note that $\mathbf{r}$ cannot be exactly $\mathbf{0}$ unless the object that is being projected into 2D is planar because the WP camera does not have the perspective effect [2]. Since faces are not planar objects, we can ignore the possibility of $\mathbf{r}$ being $\mathbf{0}$ and since $\text{rank}(\mathbf{C}) = M > 0$ (see Theorem 3.1), $\mathbf{C} \neq \mathbf{0}$, therefore we can ignore the possibility that $\mathbf{r} \in \text{Null}(\mathbf{C}^T)$ is satisfied trivially.

Moreover, if we are allowed to treat $\mathbf{r}$ as a random vector, we can show that the probability that it lies in $\text{Null}(\mathbf{C}^T)$ is 0. To this end, let $\mathbf{r}$ be a continuous random vector and define $\mathbf{y} = (y_1, \ldots, y_M)$ as the M-dimensional random vector $\mathbf{y} := \mathbf{C}^T\mathbf{r}$. Clearly, $\mathbf{y} = \mathbf{0}$ if and only $||\mathbf{y}||$ is 0. The probability of event $\mathbf{r} \in \text{Null}(\mathbf{C}^T)$ can be written as:

$$P(\mathbf{r} \in \text{Null}(\mathbf{C}^T)) \\ = P(\mathbf{C}^T\mathbf{r}=\mathbf{0}) = P(\mathbf{y}=\mathbf{0}) = P(\mathbf{z}=0), \tag{S.12}$$

where $\mathbf{z}$ is a (one-dimensional) random variable defined through the transformation of the random vector $\mathbf{r}$ as $\mathbf{z} := \sum_{i=1}^M y_i^2 = \sum_{i=1}^M (\mathbf{c}_i^T\mathbf{r})^2 = 0$, where $\mathbf{c}_i$ is the ith column of $\mathbf{C}$. Thus, we have reduced the probability of the event $\mathbf{r} \in \text{Null}(\mathbf{C}^T)$ to the probability of a random variable, namely $\mathbf{z}$, taking the value 0[1]. This probability is 0 since the probability of a continuous random variable taking any given value is 0 [4].

Appendix C

In this section we report, similarly to Section 4.5 of the main text, the false positives of Action Units (AUs) in the presence of pure head movements, but using the original source code of the ITWMM method[2]. Fig. S.1 reports the false positive rate (FPR) in AU detection w.r.t. amount of rotation for yaw and pitch rotation. Results are reported for two settings: (i) using the default parameters of the code and (ii) using only landmarks by setting the weight of the texture component to zero. In both cases, the weight of the smoothing component in the code has been set to 0. Results suggest that, similarly to results with our own implementation (Fig. 8 of the main text), spurious expressions (*i.e.*, false positives) are generated in the presence of pure rotations.

References

[1] Stephen Boyd and Lieven Vandenberghe. *Introduction to applied linear algebra: vectors, matrices, and least squares*. Cambridge university press, 2018. 1

[1] As mentioned earlier in this appendix, $\mathbf{C}$ cannot be $\mathbf{0}$, therefore $\mathbf{z}$ is indeed a random variable and not a deterministic one.

[2] https://github.com/menpo/itwmm/

[2] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 2

[3] David A Harville. Matrix algebra from a statistician's perspective, 1998. 1

[4] Athanasios Papoulis and S Unnikrishna Pillai. *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education, 2002. 2