Journal of Transfer and Generalization
Volume 1 · Number 1 · 2026

A Frontal Reference Gallery Transfers Unevenly Across Head Poses in CMU Captures

Blake King[1] · Noah Knight[2] · Benjamin Lewis[3]

[1] Slovenia · [2] Slovakia · [3] Malta

Original publication: November 1, 2026 | Article 663

Abstract

All thirty-two frontal queries in a complete sixteen-person CMU subset were identified by a frontal-only reference gallery. That perfect within-condition score concealed uneven transfer: left, right, and upward queries reached 65.63%, 78.13%, and 96.88% accuracy. The left-minus-frontal difference was -34.38 percentage points, with a working interval from -60.42 to -8.33; the right and upward intervals included zero. The comparison keeps query expression categories and identity membership common across poses. It shows why a favorable frontal result should not stand for uniform capture-condition performance, while remaining limited to raw whole-image matching on a small familiar roster.

Capture-condition comparison with a common known-identity roster. Head poses are source labels, not randomized treatments or inferred camera angles.

A Frontal Reference Gallery Transfers Unevenly Across Head Poses in CMU Captures

> Capture-condition comparison with a common known-identity roster. Head poses are source labels, not randomized treatments or inferred camera angles.

Abstract

All thirty-two frontal queries in a complete sixteen-person CMU subset were identified by a frontal-only reference gallery. That perfect within-condition score concealed uneven transfer: left, right, and upward queries reached 65.63%, 78.13%, and 96.88% accuracy. The left-minus-frontal difference was -34.38 percentage points, with a working interval from -60.42 to -8.33; the right and upward intervals included zero. The comparison keeps query expression categories and identity membership common across poses. It shows why a favorable frontal result should not stand for uniform capture-condition performance, while remaining limited to raw whole-image matching on a small familiar roster.

A reference view defines a domain of evidence

A reference gallery can be accurate under the conditions represented by its photographs while offering much less support for other views. If every reference faces forward, a successful frontal query establishes a relatively narrow result. It says little about whether a turned head will remain nearest to that person's references rather than another person's images. The practical question is not simply whether references are available, but which capture conditions they represent.

We investigate how a frontal-only gallery identifies the same people when query head pose changes. The analysis uses the source categories straight, left, right, and up. Query expression categories remain sad and angry, while references remain neutral and happy. We restrict all comparisons to the same people with complete query menus, so an apparent pose difference cannot arise merely because different participants entered each comparison.

Meng et al., 2026 propose a multi-view identity-mosaic mechanism for subject-preserving video generation. Its emphasis on multiple views motivates taking reference coverage seriously. The present experiment does not reproduce its architecture, evaluate generated video, or show how its multi-view mechanism behaves. It asks a preceding measurement question: how much a single reference-view condition can conceal uneven performance across observed query views.

The distinction matters because domain transfer has several meanings. Here the domain is a labeled capture condition within the same roster. We do not move to new people, another laboratory, another camera, a later year, or a different demographic population. Calling the comparison pose-domain transfer should preserve that specific meaning rather than suggest that every form of external generalization has been tested.

Why the same identity can become a distant image

A turn of the head changes which surfaces are visible, where facial structures appear, and how the head aligns with the background. Whole-image Euclidean distance does not represent those changes as a known geometric transformation. It compares pixels at the same array coordinates. Two images of one person can therefore separate in this metric even when a human observer would regard identity as unchanged.

Turk and Pentland, 1991 and Belhumeur et al., 1997 developed learned representations for image-based recognition. They provide a historical basis for distinguishing the representation from the identity label it is intended to recover. Our deliberately simple baseline does not learn either representation. It makes the dependence on the reference capture domain visible without attributing the result to an unexamined network embedding.

The sequence-modeling setting in Samaria and Harter, 1994 also differs from our independent still-image lookup. The Olivetti archive documented in AT&T Laboratories Cambridge, 2026 supplies another familiar small-roster face collection, but we do not combine its participants with CMU or treat results from the two archives as matched replications. Dataset identity and capture conditions should remain attached to every recognition claim.

Modern embedding methods seek stronger invariances. Schroff et al., 2015 learns an identity-oriented metric, while Wang et al., 2018 modifies the training margin. Kim et al., 2022 and Meng et al., 2021 incorporate quality-related considerations in representation learning. Those designs explain why our raw-pixel result cannot serve as a benchmark estimate for contemporary face recognition. Their representations, training data, and decision calibration are absent from this experiment.

A common roster with complete query conditions

The images come from Mitchell, 1997, with naming conventions and the original educational context available through Carnegie Mellon University, 2026. The archive contains 624 usable captures of twenty people. Sixteen planned captures carry a source bad-file marker and are excluded before any recognition analysis. Three supplied image sizes are derivatives of the same captures, so they do not increase the number of people or independently observed situations.

We read the quarter-size PGM files, thirty by thirty-two grayscale pixels, and parse identity, head pose, expression, and eye condition directly from their names. No demographic attribute is inferred. Source labels distinguish poses categorically; they do not provide measured yaw or pitch angles. Thus, left and right comparisons refer to the supplied capture categories rather than equal angular displacements in opposite directions.

References are the straight-pose, open-eye images with neutral or happy expressions. Each included identity must have both references. Queries use open eyes with sad and angry expressions at every one of the four poses. An identity enters the common cohort only if all eight such query captures are available. These rules depend on source availability and labels, not on whether the classifier recognizes the person well.

Sixteen people satisfy the complete-menu conditions, producing thirty-two gallery images and 128 queries. Each person contributes two queries per pose. The four omitted people remain documented in the source index; their absent or bad-marked query conditions are not imputed. A complete cohort improves within-roster comparability while narrowing the population to which the result directly applies.

This is a closed-roster identification problem. Every query identity is in the gallery, and the classifier must assign one of those sixteen identities. The design contains no unknown-person rejection threshold or claim about a verification false-accept rate. A perfect frontal score consequently has a different meaning from a safe identity-verification system operating in an open population.

Classification without adapting to the query pose

Pixels are converted to floating-point values in the interval from zero to one. Squared Euclidean distances compare each query with all thirty-two references. The identity of the nearest reference is returned, using lexicographic gallery order for an exact numerical tie. Neither the gallery nor the metric changes when the query pose changes.

There is no registration, facial cropping, brightness normalization, learned dimensionality reduction, or embedding network. Because we use the whole source frame, classification may exploit clothing, hair, backdrop, or capture-specific position as well as facial appearance. The outcome should therefore be called roster-identity lookup accuracy rather than an amount of preserved facial identity.

Collections such as Cao et al., 2017 and Huang et al., 2008 make substantially different demands on image variation. Kemelmacher-Shlizerman et al., 2015 introduces a much larger recognition setting, while Kärkkäinen and Joo, 2019 emphasizes balanced attribute coverage for a separate dataset purpose. Those comparisons establish boundaries rather than provide interchangeable evidence. Twenty source participants and sixteen complete-roster participants cannot support demographic fairness or unconstrained-recognition conclusions.

No source image from a query condition is used as its own reference. Expression membership is also separated: neutral and happy supply the gallery, while sad and angry supply queries. Nevertheless, the same people and capture setting appear on both sides. This is intentional for the within-roster question and should not be described as generalization to unseen identities.

What the paired pose contrasts estimate

We average each person's two correctness indicators within each pose. For left, right, and up separately, we subtract that person's frontal-query average. Each model therefore has sixteen paired identity rows and estimates the mean pose contrast. Equal numbers of expressions and common identity membership remove two simple composition differences from the comparison.

All three models are intercept-only OLS with HC3 standard errors and t15 intervals, following the heteroskedasticity-robust approach in White, 1980. There are no controls or fixed effects. R-squared is zero because the regressions summarize means. We retain the paired outcome construction instead of fitting 128 nominally independent binary observations from repeatedly photographed people.

Roberts et al., 2017 motivates respecting the level at which dependence and the generalization claim arise. The source roster is not a probability sample of future users, so a robust covariance estimator does not turn these intervals into population coverage guarantees. Related capture conditions also remain dependent within each person; pairing handles their observed difference but does not reconstruct unobserved capture sessions.

The three displayed intervals are pointwise, not a jointly adjusted family. Benjamini and Hochberg, 1995 provides a formal multiplicity framework, but no such adjustment is claimed for our table. Cook, 1977 motivates retaining person-level results so that readers can inspect whether a contrast depends on a few participants. We do not remove poorly recognized identities or reselect the complete cohort after seeing accuracy.

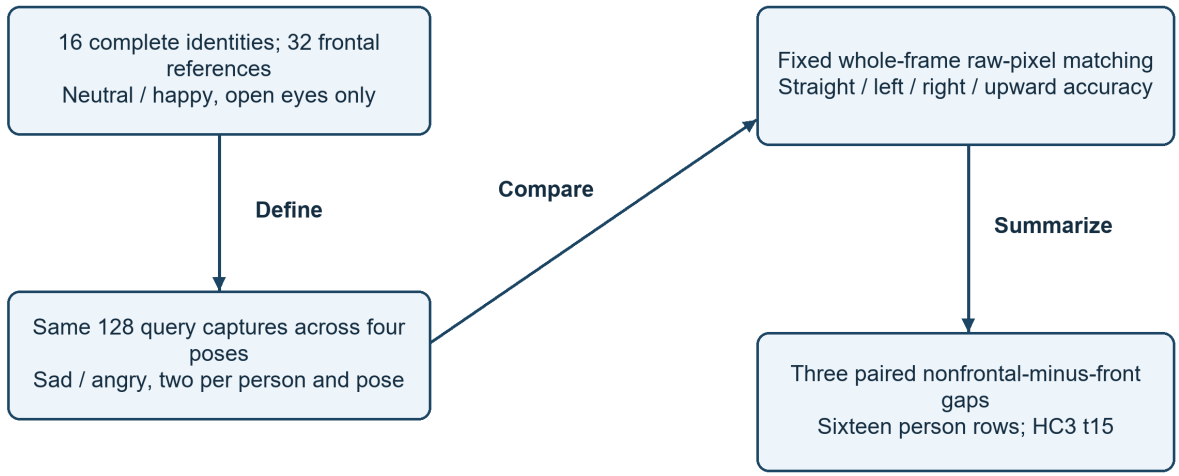


Figure 1. *The same known roster and query expressions are compared across source head-pose categories. Arrows describe measurement, not a randomized causal effect of turning a person's head.*

Perfect frontal accuracy did not extend uniformly

The frontal gallery identified all thirty-two frontal queries correctly. Mean frontal accuracy was therefore one in the observed cohort. Left queries produced an equal-person accuracy of 0.656250, right queries 0.781250, and upward queries 0.968750. These differences occurred with the same reference images, metric, identity roster, and expression categories.

The left-minus-frontal contrast was -0.343750 with standard error 0.122191 and working interval -0.604193 to -0.083307. The right contrast was -0.218750, with interval -0.442721 to 0.005221. The upward contrast was -0.031250, with interval -0.100042 to 0.037542. The left comparison showed the clearest recorded loss, while the other two intervals leave greater uncertainty about their mean contrasts.

Table 1. *Pose-specific minus frontal identity accuracy in the common sixteen-person cohort. N is sixteen paired person means in every model; intercept-only OLS, HC3 covariance, t15 intervals, no controls or fixed effects. Three pointwise comparisons are reported without a familywise claim.*

Model	Term	Estimate (SE)	95% CI	N
M1	Mean outcome/contrast	-0.343750 (0.122191)	[-0.604193, -0.083307]	16
M1: fit	R-squared = 0.0000	-	-	16
M2	Mean outcome/contrast	-0.218750 (0.105079)	[-0.442721, 0.005221]	16
M2: fit	R-squared = 0.0000	-	-	16
M3	Mean outcome/contrast	-0.031250 (0.032275)	[-0.100042, 0.037542]	16
M3: fit	R-squared = 0.0000	-	-	16

Model labels: M1 = left query minus straight query accuracy; M2 = right query minus straight query accuracy; M3 = up query minus straight query accuracy.

The different significance labels that could be assigned to left and right do not establish a statistically demonstrated difference between those directions. Their comparison would require a direct

left-minus-right contrast, which is not one of the fitted models. Nor does the interval crossing zero for upward pose prove equivalence to frontal performance. The actual upward accuracy remained below the observed frontal score.

The perfect frontal reference also should not receive a zero-width population confidence statement. It is a finite count of successful queries under one gallery. Our models estimate paired pose differences, and their variation comes from the supplied person-level outcomes. The study makes no assertion that a newly photographed frontal query from these or other people must be recognized.

There is no measured angular dose behind the categorical ordering. Upward captures may retain more of the discriminative whole-frame appearance than lateral captures, but the archive does not isolate that explanation. Head position, visible contour, expression execution, and background alignment can change jointly. The comparatively small upward loss therefore cannot establish that pitch is inherently easier than yaw or that one direction has a universal advantage for identity recognition.

The common expression menu removes an obvious label-composition difference without holding every visual detail constant. A person's sad expression in a frontal capture is not an identical facial configuration rotated into another photograph. These are separate recorded images. The paired design controls which person contributes to each average; it does not manufacture a geometrically controlled counterfactual from observational captures. This distinction limits both causal language and the design of any proposed follow-up experiment.

Reference coverage and generative identity claims

Personalization approaches such as Li et al., 2023a use identity information to condition generation. Gal et al., 2022 and Kumari et al., 2022 investigate personalization and customization through different mechanisms. Their common relevance is that the information supplied as a reference can constrain what a model is expected to preserve. A frontal photograph is useful evidence, but it does not directly observe every view of a subject.

He et al., 2024 addresses identity-preserving human video generation, while Ren et al., 2024 studies visual consistency in image-to-video generation. Those objectives involve temporal outputs and generative choices not represented by our still-image classifier. A poor raw-pixel lookup result after a pose change does not prove identity drift in a generated video, just as a favorable frontal match would not prove temporal consistency.

The domain-shift framing in Ovadia et al., 2019 helps specify the transfer claim. This experiment changes labeled capture pose while keeping roster and expression categories common. Other dimensions remain fixed or unmeasured. We cannot infer whether an additional reference view would repair the loss, because the gallery is never expanded or rebalanced in the completed comparison.

Validation-selection concerns in Varma and Simon, 2006 explain why a new reference menu should be assessed on data separate from those used to choose it. A developer observing the left-view loss could reasonably design a multi-view follow-up, but choosing extra references after inspecting these same queries would answer a different, adaptively informed question. We retain the fixed frontal-only result rather than present such a repair as independent confirmation.

Interpreting the outcome as a construct

Gneiting and Raftery, 2007 emphasizes that evaluation should match its intended objective. If the objective is closed-roster whole-image lookup under these capture conditions, the table is directly relevant. If the objective is perceived identity in a novel generated video, additional measures and human assessment would be needed. The same correct-label percentage cannot silently serve both objectives.

The scale distinction in Stevens, 1946 also matters. A drop of 0.343750 is an accuracy change over the observed roster and conditions. It is not a percentage reduction in identity itself, a geometric pose-error magnitude, or the probability that every left-view output is unacceptable. Naming that unit prevents an otherwise vivid result from becoming an overgeneralized claim.

Source records follow the documentation perspective in Gebru et al., 2021 and reuse principles in Wilkinson et al., 2016. Capture identifiers, complete-cohort membership, reference definitions, query outcomes, and person summaries remain available locally. Other analyses in this batch reuse the CMU archive for masks and resolution; they investigate different operations and do not multiply the independent evidence for the pose contrast reported here.

Conclusion

A perfect frontal score concealed substantial variation across query-pose conditions in this small raw-pixel experiment. The clearest loss appeared for left-labeled captures, while right and upward comparisons remained less precise. The result supports reporting capture-condition coverage alongside an aggregate identity score. Its proper scope is a fixed frontal gallery and a common familiar roster, with no claim to have tested a modern identity embedding or a subject-preserving generation system.

References

1. He, Xuanhua; Liu, Quande; Qian, Shengju; Wang, Xin; Hu, Tao; Cao, Ke; Yan, Keyu; Zhang, Jie. (2024). ID-Animator: Zero-Shot Identity-Preserving Human Video Generation. arXiv:2404.15275. https://arxiv.org/abs/2404.15275.
2. Cao, Qiong, et al. "VGGFace2: A dataset for recognising faces across pose and age" *arXiv preprint arXiv:1710.08092*, 2017. https://arxiv.org/abs/1710.08092.
3. Cook, R. Dennis. (1977). Detection of Influential Observation in Linear Regression. Technometrics. https://doi.org/10.1080/00401706.1977.10489493.
4. Benjamini, Yoav and Yosef Hochberg. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 57, no. 1, 1995, pp. 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
5. Turk, M.A.; Pentland, A.P.. (1991). Face recognition using eigenfaces. Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/cvpr.1991.139758.
6. Ovadia, Yaniv, et al. “Can You Trust Your Model’s Uncertainty? Evaluating Predictive Uncertainty Under Dataset Shift.” *Advances in Neural Information Processing Systems*, vol. 32, 2019. https://arxiv.org/abs/1906.02530.
7. AT&T Laboratories Cambridge. (2026). Modified Olivetti faces: Sam Roweis MATLAB version [Data archive; mirror accessed 7 November 2026]. File identity and row-label layout verified against the scikit-learn maintainer loader. https://ndownloader.figshare.com/files/5976027.
8. Meng, Zijie; Liu, Jiwen; Liu, Yufei; Tong, Chengzhuo; Liu, Xiaoqiang; Zhang, Yuanxing; Xu, Yulong; Wan, Pengfei. (2026). ARGUS: Stacked Multi-View Identity Mosaic Injection for Subject-Preserving Video Generation. arXiv:2606.11670. https://arxiv.org/abs/2606.11670.
9. Gal, Rinon, et al. "An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion." *arXiv preprint arXiv:2208.01618*, 2022. https://arxiv.org/abs/2208.01618.
10. White, Halbert. “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.” *Econometrica*, vol. 48, no. 4, 1980, pp. 817-838. https://doi.org/10.2307/1912934.

11. Schroff, Florian, et al. "FaceNet: A Unified Embedding for Face Recognition and Clustering" *arXiv preprint arXiv:1503.03832*, 2015. https://arxiv.org/abs/1503.03832.

12. Ren, Weiming; Yang, Huan; Zhang, Ge; Wei, Cong; Du, Xinrun; Huang, Wenhao; Chen, Wenhu. (2024). ConsistI2V: Enhancing Visual Consistency for Image-to-Video Generation. arXiv:2402.04324. https://arxiv.org/abs/2402.04324.

13. Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan; Appleton, Gabrielle; Axton, Myles; Baak, Arie; Blomberg, Niklas; Boiten, Jan-Willem; da Silva Santos, Luiz Bonino; Bourne, Philip E.; Bouwman, Jildau; Brookes, Anthony J.; Clark, Tim; Crosas, Mercè; Dillo, Ingrid; Dumon, Olivier; Edmunds, Scott; Evelo, Chris T.; Finkers, Richard; Gonzalez-Beltran, Alejandra; Gray, Alasdair J.G.; Groth, Paul; Goble, Carole; Grethe, Jeffrey S.; Heringa, Jaap; 't Hoen, Peter A.C; Hooft, Rob; Kuhn, Tobias; Kok, Ruben; Kok, Joost; Lusher, Scott J.; Martone, Maryann E.; Mons, Albert; Packer, Abel L.; Persson, Bengt; Rocca-Serra, Philippe; Roos, Marco; van Schaik, Rene; Sansone, Susanna-Assunta; Schultes, Erik; Sengstag, Thierry; Slater, Ted; Strawn, George; Swertz, Morris A.; Thompson, Mark; van der Lei, Johan; van Mulligen, Erik; Velterop, Jan; Waagmeester, Andra; Wittenburg, Peter; Wolstencroft, Katherine; Zhao, Jun; Mons, Barend. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. https://doi.org/10.1038/sdata.2016.18.

14. Huang, Gary B., Marwan Mattar, Tamara Berg, and Erik Learned-Miller. "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments." *ECCV Workshop on Faces in Real-Life Images*, 2008. https://cs.umass.edu/~elm/papers/Huang_eccv2008-lfw.pdf.

15. Gebru, Timnit, et al. "Datasheets for datasets" *Communications of the ACM*, 2021. https://doi.org/10.1145/3458723.

16. Kemelmacher-Shlizerman, Ira, et al. "The MegaFace Benchmark: 1 Million Faces for Recognition at Scale" *arXiv preprint arXiv:1512.00596*, 2015. https://arxiv.org/abs/1512.00596.

17. Stevens, S. S.. (1946). On the Theory of Scales of Measurement. Science. https://doi.org/10.1126/science.103.2684.677.

18. Varma, Sudhir and Richard Simon. "Bias in error estimation when using cross-validation for model selection." *BMC Bioinformatics*, vol. 7, no. 1, 2006. https://doi.org/10.1186/1471-2105-7-91.

19. Carnegie Mellon University. (2026). Neural Networks for Face Recognition: companion data and file documentation [Original educational archive; accessed8November2026]. https://www.cs.cmu.edu/~tom/faces.html.

20. Belhumeur, P.N.; Hespanha, J.P.; Kriegman, D.J.. (1997). Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence. https://doi.org/10.1109/34.598228.

21. Wang, Hao, et al. "CosFace: Large Margin Cosine Loss for Deep Face Recognition" *arXiv preprint arXiv:1801.09414*, 2018. https://arxiv.org/abs/1801.09414.

22. Meng, Qiang, et al. "MagFace: A Universal Representation for Face Recognition and Quality Assessment." *arXiv preprint arXiv:2103.06627*, 2021. https://arxiv.org/abs/2103.06627.

23. Kumari, Nupur; Zhang, Bingliang; Zhang, Richard; Shechtman, Eli; Zhu, Jun-Yan. (2022). Multi-Concept Customization of Text-to-Image Diffusion. arXiv:2212.04488. https://arxiv.org/abs/2212.04488.

24. Mitchell, Tom. (1997). CMU Face Images [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5JC79.

25. Gneiting, Tilmann; Raftery, Adrian E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association. https://doi.org/10.1198/016214506000001437.

26. Kärkkäinen, Kimmo and Jungseock Joo. "FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age" *arXiv preprint arXiv:1908.04913*, 2019. https://arxiv.org/abs/1908.04913.

27. Kim, Minchul, et al. "AdaFace: Quality Adaptive Margin for Face Recognition." *arXiv preprint arXiv:2204.00964*, 2022. https://arxiv.org/abs/2204.00964.

28. Li, Zhen, et al. "PhotoMaker: Customizing Realistic Human Photos via Stacked ID Embedding." *arXiv preprint arXiv:2312.04461*, 2023. https://arxiv.org/abs/2312.04461.

29. Roberts, David R., et al. "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure." *Ecography*, vol. 40, no. 8, 2017, pp. 913-929. https://doi.org/10.1111/ecog.02881.

30. Samaria, F.S.; Harter, A.C.. (1994). Parameterisation of a stochastic model for human face identification. Proceedings of 1994 IEEE Workshop on Applications of Computer Vision. https://doi.org/10.1109/acv.1994.341300.