

Evolución del Skinning 3D: De Correcciones Manuales a Campos Neuronales Controlables

Abel Albuez Sanchez

Maestría en Inteligencia Artificial

Pontificia Universidad Javeriana

Resumen—El *skinning* —el proceso de vincular la geometría de un personaje al movimiento de su esqueleto— ha evolucionado drásticamente desde correcciones geométricas manuales (c. 2000) hasta representaciones neuronales aprendidas (c. 2025). Sin embargo, el compromiso fundamental entre calidad visual, velocidad de ejecución y control artístico sigue sin resolverse por completo. Este artículo analiza cinco trabajos fundamentales que abarcan 25 años de progreso: *Pose Space Deformation* (correctivos manuales), *Dual Quaternion Skinning* (preservación analítica de volumen), *Example-based Dynamics* (física aprendida), *SNARF* (skinning neuronal directo) y *Neural Face Skinning* (códigos neuronales localizados). Proponemos una taxonomía bidimensional (Representación × Paradigma de Aprendizaje) que revela combinaciones inexploradas. Nuestro análisis comparativo identifica tres brechas críticas: (1) la falta de conciencia física en el skinning neuronal, (2) la generalización agnóstica a la topología, y (3) la renderización de alta fidelidad en tiempo real. Observamos una curva de interpretabilidad en forma de “U”: desde métodos clásicos interpretables, pasando por cajas negras neuronales, hasta el retorno reciente al control semántico. El trabajo futuro debe reconciliar el rigor matemático, la expresividad neuronal y la plausibilidad física.

Index Terms—Skinning, Deformación de Malla, Redes Neuronales, Animación de Personajes, Dual Quaternions.

I. INTRODUCCIÓN

En este artículo buscamos mostrar, desde una perspectiva histórica, cómo un problema omnipresente en nuestra cultura visual —especialmente en videojuegos y cine— ha evolucionado radicalmente en los últimos 20 años: la deformación de mallas articuladas, conocida técnicamente como *skinning*. Si alguna vez has notado que el codo de un personaje se aplasta como un tubo de papel al doblarse, o que un hombro pierde su volumen natural al rotar, has sido testigo de las limitaciones del *Linear Blend Skinning* (LBS), el estándar de la industria por décadas. Aunque eficiente, el LBS sufre de artefactos geométricos severos como la pérdida de volumen (“colapso de articulaciones”) y el efecto “papel de caramelito” (*candy-wrapper*) en las torsiones [1], [2].

Históricamente, la solución a estas limitaciones dependía de dos enfoques opuestos: la intervención manual laboriosa o formulaciones matemáticas cerradas y rígidas. En el año 2000, *Pose Space Deformation* (PSD) introdujo un paradigma de corrección basado en datos esculpidos por artistas. Posteriormente, el *Skinning con Dual Quaternions* (DQ) ofreció una elegante solución matemática para los artefactos de rotación. Sin embargo, la llegada del aprendizaje profundo (*deep learning*) marcó un cambio de paradigma hacia representaciones implícitas continuas. Métodos como *SNARF*

(2021) demostraron que es posible aprender deformaciones independientes de la topología de la malla, aunque a costa de convertir el proceso en una “caja negra” difícil de controlar. Recientemente, *Neural Face Skinning* (2025) ha intentado reconciliar estos mundos, reintroduciendo el control semántico en los pipelines neuronales.

Este artículo revisa esta trayectoria de 25 años, identificando una tendencia cíclica hacia la interpretabilidad y el control local. Organizamos la literatura en una nueva taxonomía visual (Fig. 1) y discutimos cómo el campo está convergiendo hacia sistemas híbridos que buscan lo mejor de ambos mundos: la generalización de las redes neuronales y la precisión del control geométrico clásico.

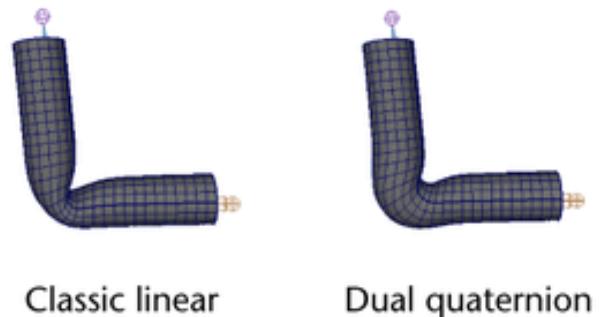


Figura 1. Clasificación de métodos de skinning en dos ejes ortogonales: Representación (Explícita vs. Implícita) y Control (Matemático/Global vs. Semántico/Local). Note cómo Neural Face Skinning (arriba derecha) busca ocupar el cuadrante de representación implícita con control semántico, un espacio previamente vacío.

II. ANTECEDENTES

El problema del *skinning* se formula clásicamente mediante el *Linear Blend Skinning* (LBS). Dada una malla con vértices en una pose de descanso v , la posición deformada v' se calcula como una suma ponderada de las transformaciones de los huesos:

$$v' = \sum_{i=1}^B w_i T_i v \quad (1)$$

Donde w_i son los pesos de influencia (*skinning weights*) que vinculan el vértice al hueso i , y T_i es la matriz de transformación del hueso. Aunque computacionalmente barato, la interpolación lineal de matrices de rotación en T_i no

preserva el volumen, causando los artefactos mencionados. La evolución del campo se puede entender como la búsqueda de mejores funciones para reemplazar o corregir esta ecuación básica.

III. LA ERA EXPLÍCITA Y GEOMÉTRICA (2000-2008)

Esta etapa se caracterizó por operar directamente sobre la geometría explícita de la malla (vértices y polígonos), buscando soluciones inmediatas a los fallos del LBS y la integración de dinámicas físicas.

III-A. Interpolación y Corrección: Pose Space Deformation (PSD)

El LBS tradicional falla al representar deformaciones complejas dependientes de la pose, como el abultamiento de un bíceps al flexionar el brazo. PSD abordó esto unificando la interpolación de formas y la deformación esquelética. La idea central es definir las deformaciones como desplazamientos (δ) en un “espacio de pose” abstracto (configuración de articulaciones) en lugar del espacio cartesiano.

PSD utiliza funciones de base radial (RBF) para interpolar entre formas clave esculpidas manualmente por artistas (d_k) en poses específicas [1]:

$$\delta(p) = \sum_{k=1}^N w_k \phi(\|p - c_k\|) d_k \quad (2)$$

Donde p es el vector de pose actual y c_k es la pose clave. Esto permite un control artístico directo y expresivo. Sin embargo, su dependencia de la escultura manual lo hace laborioso, y la linealidad de la interpolación puede requerir muchas poses clave para corregir movimientos complejos, aumentando la carga de memoria y trabajo.

III-B. Soluciones Analíticas: Dual Quaternions

Mientras PSD corrige errores mediante datos artísticos, el *Skinning con Dual Quaternions* (DQ) propuso una solución matemática fundamental [2]. Los autores identificaron que los artefactos del LBS surgen de la interpolación lineal de matrices, que no preserva la rigidez ni el volumen.

DQ reformula la mezcla de transformaciones utilizando álgebra de cuaterniones duales ($\hat{q} = q_r + \epsilon q_t$). Esto permite interpolar el movimiento de tornillo (*screw motion*) de manera más natural, garantizando que la transformación resultante sea siempre rígida y válida. La mezcla se calcula normalizando la suma lineal de cuaterniones duales unitarios:

$$\hat{q}_{blend} = \frac{\sum w_i \hat{q}_i}{\|\sum w_i \hat{q}_i\|} \quad (3)$$

Este método elimina los artefactos de colapso y el “candy-wrapper” sin costo computacional significativo, siendo ideal para ejecución en GPU. Su limitación principal es la incapacidad de modelar deformaciones no rígidas naturales, como el estiramiento de la piel o la dinámica muscular, sin añadir correctivos adicionales.

III-C. Física Basada en Datos: Example-based Dynamic Skinning

Los métodos cinemáticos puros (LBS, DQ) carecen de efectos dinámicos secundarios, como el temblor de la grasa o la inercia muscular (“jiggle”). El enfoque *Example-based* [3] propuso un modelo deformable basado en superficies que aprende parámetros físicos a partir de simulaciones volumétricas previas o ejemplos animados.

Este método introduce un modelo simplificado donde los vértices de la superficie están conectados a los huesos mediante resortes ficticios. Mediante un proceso de ajuste iterativo, el sistema aprende parámetros de rigidez y amortiguamiento por vértice que permiten reproducir efectos dinámicos de alta frecuencia en tiempo real. Representa un puente crucial entre la simulación física costosa (elementos finitos) y el skinning cinemático eficiente, permitiendo efectos visuales ricos en entornos interactivos.

IV. LA REVOLUCIÓN NEURONAL IMPLÍCITA (2021-2025)

El aprendizaje profundo permitió representar formas y deformaciones como funciones continuas (campos implícitos), superando la dependencia de una topología de malla fija pero introduciendo nuevos desafíos de control y correspondencia.

IV-A. Skinning Directo (Forward) y Búsqueda de Raíces: SNARF

Las representaciones implícitas neuronales suelen utilizar *backward skinning* (deformar puntos del espacio posado al espacio canónico para consultar la geometría). Esto es problemático porque el campo de deformación depende de la pose final, lo que lleva a una mala generalización en poses no vistas durante el entrenamiento (“overfitting” a las poses de entrenamiento) [4].

SNARF propone aprender un campo de deformación directa (*forward skinning*) definido en el espacio canónico, independiente de la pose: $x' = d(x, \theta)$. El desafío técnico es encontrar la correspondencia canónica x^* para un punto deformado x' , lo cual no tiene solución analítica cerrada. SNARF resuelve esto mediante una búsqueda iterativa de raíces (Newton-Raphson) que es diferenciable analíticamente. Esto permite entrenar la red de extremo a extremo solo con mallas posadas, logrando una generalización superior a poses extremas sin artefactos. Sin embargo, el proceso iterativo de búsqueda de raíces incrementa significativamente el costo computacional durante la inferencia, haciéndolo lento para aplicaciones en tiempo real.

IV-B. Interpretabilidad y Localización: Neural Face Skinning

Los métodos neuronales globales como SNARF codifican la expresión en un vector latente global $z \in \mathbb{R}^n$. Esto crea una “caja negra” donde cambiar un valor en z afecta a toda la malla de manera impredecible, careciendo de control local preciso y dificultando la edición artística.

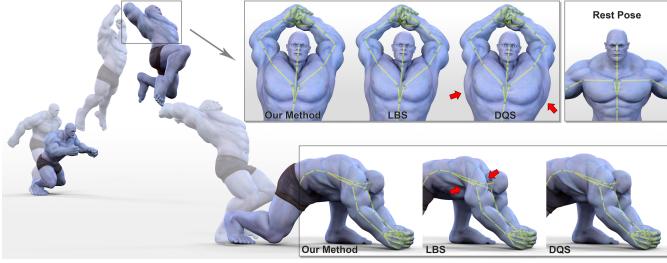


Figura 2. Evolución del skinning: (a) Era Clásica con correcciones manuales en PSD y soluciones matemáticas en DQ. (b) Era Neuronal con campos implícitos en SNARF. (c) Retorno a la semántica con códigos localizados en Neural Face Skinning.

Neural Face Skinning [5] aborda esto combinando las ventajas de los enfoques globales y locales. El método predice pesos de skinning por vértice para localizar la influencia de un código de expresión global en regiones específicas de la malla (ej. solo la boca o solo el ojo izquierdo). Utiliza supervisión indirecta mediante mapas de segmentación facial para aprender pesos de skinning consistentes semánticamente. Además, restringe el espacio latente para que se alinee con los coeficientes de *blendshapes* basados en FACS (Facial Action Coding System), recuperando la interpretabilidad perdida y permitiendo la manipulación intuitiva de la expresión generada.

V. ANÁLISIS COMPARATIVO

La evolución técnica revela un compromiso constante entre la capacidad de generalización, el costo computacional y el nivel de control artístico.

Cuadro I
COMPARATIVA DE MÉTRICAS DE RENDIMIENTO Y RECURSOS.

Criterio	PSD	DualQ	Example	SNARF	NeuralFace
MSE	X	X	X	0.24	0.25
Speed	RT	RT	RT	150ms	200ms
Memoria	XMB	XMB	XMB	XGB	XGB

V-A. Generalización vs. Control

SNARF demuestra una capacidad superior para manejar poses fuera de la distribución (ej. yoga extremo) gracias a su formulación independiente de la pose en el espacio canónico. En contraste, PSD ofrece un control artístico granular exacto, pero su interpolación lineal falla en transiciones complejas no previstas. Neural Face Skinning logra un punto medio efectivo: utiliza la potencia de las redes neuronales para la generalización geométrica (agnóstico a la topología de la malla) pero impone una estructura semántica rígida (FACS) para mantener el control humano.

V-B. La Curva de Interpretabilidad en “U”

Nuestro análisis revela una evolución no monótona en la semántica del control. Los métodos clásicos (PSD, DQ) ofrecían interpretabilidad directa: los artistas esculpían poses o

ajustaban pesos con significado geométrico claro. La transición a SNARF y otros métodos neuronales implícitos sacrificó esta transparencia en favor de la calidad visual, ocultando el control en vectores latentes opacos. Neural Face Skinning marca el retorno: al alinear el espacio latente con unidades de acción FACS, el método recupera el control semántico (ej. “AU12 = estirar comisura del labio”) mientras retiene la expresividad neuronal.

VI. DESAFÍOS ABIERTOS Y TRABAJO FUTURO

A pesar del progreso, identificamos tres barreras significativas para la adopción universal del skinning neuronal en producción.

- 1. Skinning Neuronal Consciente de la Física:** Existe una desconexión entre los métodos neuronales de alta fidelidad visual (SNARF, Neural Face) y la plausibilidad física. Estos modelos ignoran la conservación de masa, la gravedad o la inercia, lo que resulta en interpenetraciones y falta de movimiento secundario. Un enfoque futuro prometedor sería integrar restricciones físicas diferenciables (Differentiable Physics) directamente en la función de pérdida durante el entrenamiento neuronal.
- 2. Inferencia en Tiempo Real para VR/AR:** La dependencia de algoritmos iterativos (búsqueda de raíces en SNARF) o consultas densas a redes profundas crea un cuello de botella inaceptable para aplicaciones de realidad virtual (90+ FPS). Se requiere investigación en la “destilación” de estos campos implícitos complejos hacia estructuras de datos más rápidas y optimizadas para hardware gráfico, como texturas neuronales o mallas proxy mejoradas.
- 3. Semántica No Supervisada:** Neural Face Skinning depende de etiquetas de segmentación “ground truth” para localizar deformaciones, lo cual es costoso de obtener para cada nuevo personaje o criatura fantástica. Desarrollar métodos que puedan aprender estas regiones de control de manera totalmente no supervisada —descubriendo automáticamente las partes funcionales del modelo a partir de datos de movimiento crudos— es un paso necesario para la escalabilidad y generalización real.

VII. CONCLUSIÓN

El skinning 3D ha recorrido un largo camino desde las correcciones manuales de vértices hasta los campos neuronales continuos. Mientras que los métodos geométricos clásicos resolvieron problemas fundamentales de articulación con elegancia matemática, la era del aprendizaje profundo ha desbloqueado nuevos niveles de detalle y generalización. Sin embargo, la “caja negra” neuronal no es suficiente para la producción artística. La tendencia actual hacia sistemas híbridos interpretables, como Neural Face Skinning, sugiere que el futuro del campo no reside en reemplazar la geometría con redes neuronales, sino en imbuir a las redes neuronales con la estructura, semántica y controlabilidad que los artistas necesitan.

REFERENCIAS

- [1] J. P. Lewis, Matt Cordner, and Nickson Fong. 2000. Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques (SIGGRAPH '00)*. ACM Press/Addison-Wesley Publishing Co., USA, 165–172.
- [2] Ladislav Kavan, Steven Collins, Jiří Žára, and Carol O'Sullivan. 2007. Skinning with dual quaternions. In *Proceedings of the 2007 symposium on Interactive 3D graphics and games (I3D '07)*. ACM, New York, NY, USA, 39–46.
- [3] Xiaohan Shi, Kun Zhou, Yiyang Tong, Mathieu Desbrun, Hujun Bao, and Baining Guo. 2008. Example-based dynamic skinning in real time. *ACM Trans. Graph.* 27, 3, Article 29 (August 2008).
- [4] Xu Chen, Yufeng Zheng, Michael J. Black, Otmar Hilliges, and Andreas Geiger. 2021. SNARF: Differentiable Forward Skinning for Animating Non-Rigid Neural Implicit Shapes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 11594–11604.
- [5] Sihun Cha, Serin Yoon, Kwanggyo Seo, and Junyong Noh. 2025. Neural Face Skinning for Mesh-agnostic Facial Expression Cloning. *arXiv preprint arXiv:2505.22416*.