

MỘT KỸ THUẬT ƯỚC LƯỢNG PHÁP TUYẾN BỀ MẶT ĐỐI TƯỢNG DỰA TRÊN TẬP ẢNH THU ĐƯỢC TỪ CAMERA

Trịnh Xuân Hùng¹, Đỗ Năng Toàn², Đỗ Văn Thiện¹

¹Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

²Viện Công nghệ thông tin, Đại học quốc gia Hà Nội

trixhung@gmail.com, donangtoan@gmail.com, thiendv.ioit@gmail.com

TÓM TẮT - Trong bài báo này, chúng tôi đề xuất một kỹ thuật ước lượng chính xác bản đồ pháp tuyến bề mặt (NormalMap) trên tập ảnh thu được từ đối tượng với vị trí camera cố định và hướng nguồn sáng thay đổi. Kỹ thuật đề xuất sử dụng phương pháp lặp tối ưu Gradient descent để giải ngược phương trình chiếu sáng Blinn-Phong với một số điều kiện giả định. Chúng tôi đã tiến hành cài đặt và thử nghiệm kỹ thuật đề xuất trên hai tập dữ liệu: một được sinh ra từ các mô hình 3D bằng các phương pháp chiếu sáng, hai là tập dữ liệu thu được từ camera với các điều kiện chiếu sáng biết trước được xây dựng bởi Ying Xiong và cộng sự. Kỹ thuật đề xuất cho kết quả tốt trên cả hai tập dữ liệu thử nghiệm với thời gian tính toán chấp nhận được.

Từ khóa - NormalMap, Surface Normal, NormalMapEstimate, Acquisition NormalMap.

I. GIỚI THIỆU

Các đối tượng phức tạp trong đồ họa ba chiều thường được biểu diễn thông qua một tập các đối tượng hình học cơ bản (điểm, đường và đa giác). Độ chi tiết của mỗi đối tượng phụ thuộc vào số lượng các đối tượng hình học cơ bản cấu thành. Để tăng độ chi tiết của đối tượng, cách đơn giản nhất là tăng số lượng đối tượng hình học cơ bản biểu diễn nó, việc này kéo theo nhu cầu bộ nhớ sử dụng và thời gian tính toán hiển thị đối tượng cũng tăng theo, đây là điều không mong muốn với các ứng dụng đồ họa ba chiều đặc biệt là các ứng dụng thực tại ảo vốn có số lượng lớn đối tượng cần biểu diễn theo thời gian thực. Ngoài hướng tiếp cận “tăng số lượng đối tượng hình học cơ bản cấu thành” có hai hướng tiếp cận khác thường được sử dụng: Một là sử dụng các kỹ thuật nội suy trong quá trình chiếu sáng nhằm tăng độ trơn, cong, liên tục cho bề mặt đối tượng qua đó làm tăng độ chi tiết của mô hình. Hai là sử dụng bản đồ pháp tuyến hoặc kết hợp bản đồ pháp tuyến với một số bản đồ kết cấu bề mặt để tăng độ chi tiết bề mặt đối tượng cần biểu diễn. Đi theo hướng tiếp cận thứ nhất có thể kể đến một số công trình tiêu biểu của các tác giả như Gouraud [6], Phong [21], Blinn[9] v.v.. Ưu điểm của các kỹ thuật này là có khả năng thể hiện tốt các tính chất cong, trơn, liên tục của bề mặt mặc dù nó được cấu thành từ các đa giác phẳng. Tuy nhiên, các kỹ thuật này không thể hiện được các bề mặt đối tượng có tính chất lồi, lõm phức tạp. Hướng tiếp cận thứ hai được tập trung nghiên cứu nhiều hơn trong thời gian gần đây, đi đầu cho hướng tiếp cận này là nhóm tác giả Wolfgang Heidrich với công trình[30] đã giới thiệu một phương pháp tăng cường chất lượng hiển thị bề mặt đối tượng bằng cách sử dụng bản đồ pháp tuyến. Sau đó, đã có nhiều công trình đã được đề xuất trong đó sử dụng bản đồ pháp tuyến kết hợp với các bản đồ khác như bản đồ chiều sâu (Depth-Map) của nhóm tác giả Kaneko [13], bản đồ dịch chuyển (Displacement-Map) của nhóm tác giả Jan Kautz [10], sử dụng kết hợp nhiều bản đồ pháp tuyến trên một bề mặt ba chiều của nhóm tác giả Sema Berkiten [25] v.v.. Hiện nay việc sử dụng bản đồ pháp tuyến là rất phổ biến và được hỗ trợ sẵn trong hầu hết các phần mềm thiết kế đồ họa ba chiều.

Bản đồ pháp tuyến bề mặt không chỉ được sử dụng trong việc tăng độ chi tiết của bề mặt đối tượng mà nó còn được sử dụng trong nhiều mục đích khác ví dụ: Trong xây dựng mô hình ba chiều từ ảnh tiêu biểu như các công trình của nhóm Aiyun Qing [1] khôi phục lại hình dạng khuôn mặt, Alexandros Panagopoulos [2] khôi phục lại hình dạng đối tượng dựa từ điểm tô bóng các đối tượng hình học cơ bản và nhóm Yusuke Yamaura [29] khôi phục lại bề mặt sử dụng nội suy B-Spline. Trong đồ họa hai chiều hướng pháp tuyến còn được sử dụng trong việc chiếu sáng lại cho các bức ảnh đã được chụp [17], hoặc tăng cường chất lượng hình ảnh, chuẩn hóa ánh sáng cho các bài toán phát hiện và nhận dạng đối tượng [3, 12, 23] v.v..

Do tính chất và tầm quan trọng, có khá nhiều kỹ thuật được đề xuất phục vụ việc xây dựng bản đồ pháp tuyến bề mặt [4, 5, 8, 7, 11, 16, 20, 22, 24, 26, 27] tuy nhiên hầu hết các phương pháp có độ chính xác chưa cao hoặc phức tạp trong việc tổ chức và thu thập dữ liệu, hoặc không không phù hợp với các dạng bề mặt có hệ số phản xạ gương cao. Có thể phân các kỹ thuật ước lượng pháp tuyến thành ba hướng chính:

Một là ước lượng pháp tuyến các dữ liệu ba chiều đã có như các loại mô hình ba chiều, ảnh chiều sâu, ảnh cắt lớp, hoặc tập đám mây điểm [7, 11] v.v., hướng tiếp cận này không thuộc phạm vi nghiên cứu của bài báo.

Hướng tiếp cận thứ hai là ước lượng pháp tuyến bề mặt từ một ảnh duy nhất. Theo hướng tiếp cận này cách đơn giản nhất và cũng được sử dụng phổ biến nhất để tính bản đồ pháp tuyến là dựa trên các đạo hàm riêng của mỗi điểm ảnh trong ảnh theo các hướng x, và y theo đó hướng vector pháp tuyến của điểm đó sẽ được tính trực tiếp từ các giá trị đạo hàm riêng này. Việc tính này không đảm bảo sự chính xác về mặt giá trị nhưng khi ứng dụng bản đồ kết quả trong các kỹ thuật [10, 13, 25, 30] đều cho được hình ảnh chấp nhận được, đây chính là một trong những lý do mà kỹ thuật này vẫn được sử dụng phổ biến. Ngoài việc sử dụng trực tiếp các giá trị đạo hàm để tính ra hướng pháp tuyến có nhiều công trình nghiên cứu đã được đề xuất theo hướng này với mục đích tăng độ chính xác của bản đồ kết quả, tiêu biểu là

công trình của nhóm tác giả TAI-PANG WU [26] đã đề xuất một kỹ thuật ước lượng hướng pháp tuyến chính xác hơn bao gồm hai pha, pha thứ nhất tiến hành một cách tự động với kỹ thuật đề xuất SfS, pha thứ hai chính xác hóa lại các giá trị pháp tuyến bằng cách điều khiển các điểm đánh dấu. Tương tự như vậy tác giả M. Okabe [17] đã đưa ra một cách ước lượng hướng pháp tuyến bán tự động dựa trên các đường biên của ảnh v.v..

Hướng tiếp cận thứ ba là ước lượng vector pháp tuyến từ nhiều ảnh, đây là hướng tiếp cận được sự quan tâm và có nhiều kết quả đã được đăng tải [4, 5, 8, 16, 20, 22, 24, 26, 27]. Các phương pháp theo hướng tiếp cận này thường cho độ chính xác cao hơn hướng tiếp cận thứ hai tuy nhiên nó lại yêu cầu một số ràng buộc dữ liệu khi thu nhận. Không ngoài hướng tiếp cận này, chúng tôi đề xuất một kỹ thuật ước lượng chính xác hướng pháp tuyến bề mặt đối tượng trên tập ảnh thu được từ đối tượng với vị trí camera cố định và hướng nguồn sáng thay đổi. Kỹ thuật đề xuất sử dụng phương pháp lặp tối ưu Gradient descent để giải ngược phương trình chiếu sáng Blinn-Phong [9] với một vài điều kiện giả định.

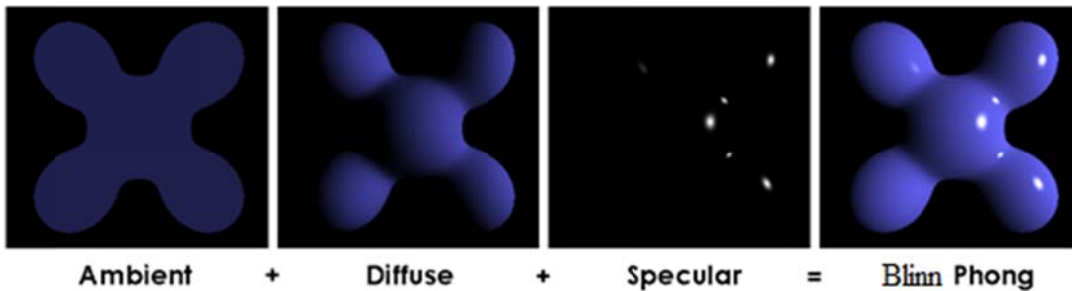
Phần còn lại của báo cáo được tổ chức như sau: Phần II trình bày các nghiên cứu liên quan. Phần III trình bày kỹ thuật ước lượng hướng pháp tuyến bề mặt đề xuất. Tiếp theo là cài đặt thử nghiệm và cuối cùng là kết luận về kỹ thuật đề xuất.

II. MỘT SỐ NGHIÊN CỨU LIÊN QUAN

Trong phạm vi nghiên cứu một cách hình thức, có thể định nghĩa bài toán ước lượng bản đồ pháp tuyến như sau: Cho một tập dữ liệu quan sát được $Z = \{z_1, z_2, \dots, z_M\}; M > 2$ trong đó $z_i = \{I_i, L_i\}$ với I_i là ảnh đối tượng thu được từ camera theo một góc nhìn cố định với thông số camera cố định trong điều kiện chiếu sáng L_i . Bài toán đặt ra là làm thế nào để tìm được bản đồ pháp tuyến của đối tượng trên góc quan sát của tập ảnh?. Để thực hiện được điều này ý tưởng cơ bản của chúng tôi là tìm một phương trình biểu diễn mối quan hệ giữa ảnh đối tượng thu được và các điều kiện sáng khi đã biết hướng pháp tuyến của từng điểm và các hệ số phản xạ của mô hình sau đó tiến hành giải ngược phương trình này dựa trên tập dữ liệu quan sát Z khi đó có thể tìm được bản đồ pháp tuyến và các hệ số phản xạ của bề mặt đối tượng. Trong nghiên cứu này dựa trên mô hình chiếu sáng Blinn-Phong [9] để xây dựng phương trình biểu diễn và giải ngược phương trình này bằng phương pháp Gradient Descent;

1. MÔ HÌNH CHIẾU SÁNG BLINN-PHONG

Blinn-Phong [9] là một mô hình chiếu sáng được cải tiến dựa trên mô hình chiếu sáng Phong [21] một mô hình chiếu sáng cục bộ được sử dụng phổ biến trong đồ họa ba chiều phản ánh tương đối đầy đủ và chính xác các hiệu ứng ánh sáng. Trong các mô hình này ánh sáng và các thuộc tính bề mặt đối tượng được phân thành các thành phần: nền (Ambient), phản xạ khuếch tán (Diffuse), phản xạ gương (Specular).



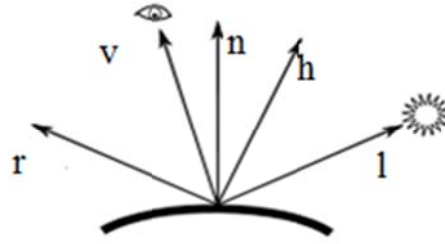
Hình 1. Các thành phần trong mô hình phản xạ Phong và Blinn-Phong

Theo đó với một tập nguồn sáng $\{L_1, L_2, \dots, L_n\}; L_j = (i_{jd}, i_{js}, \vec{l}_j)$ trong cảnh phương trình biểu diễn chiếu sáng cho một điểm trên bề mặt đối tượng được biểu diễn như sau:

$$L = k_a i_a + \sum_{m=1..n} k_d (\vec{l}_m \cdot \vec{n}) i_{md} + k_s (\vec{n} \cdot \vec{h}_m)^\alpha i_{ms} \quad (1)$$

$$\vec{h}_m = \frac{\vec{v} + \vec{l}_m}{|\vec{v} + \vec{l}_m|} \quad (2)$$

với i_a : là cường độ ánh sáng nền hay còn gọi là ánh sáng môi trường (Ambient); i_{md}, i_{ms} : tương ứng với các cường độ ánh sáng phản xạ khuếch tán và phản xạ gương của nguồn sáng thứ m ; k_a, k_d, k_s và α tương ứng với các hệ số phản xạ nền, phản xạ khuếch tán, phản xạ gương và độ bóng của bề mặt đối tượng. Thông thường thì mỗi giá trị i_a, i_{md}, i_{ms} , và k_a, k_d, k_s được biểu diễn bởi 3 thành phần tương ứng với ba màu cơ bản RGB; \vec{l}_m là vector hướng của nguồn sáng thứ m ; \vec{n} là vector pháp tuyến của điểm đang xét; \vec{v} : là vector hướng quan sát; \vec{h} là vector trung bình của hướng quan sát và hướng nguồn sáng được tính theo công thức 2. Hình 2 biểu diễn mối các vector và mối quan hệ của chúng.



Hình 2. Các vector sử dụng trong mô hình Phong và Blinn-Phong

Trong công thức 1 ở mô hình chiếu sáng nguyên gốc, tác giả sử dụng vector phản xạ \vec{r} , Blinn đã cải tiến bằng cách thay \vec{r} bằng \vec{h} tạo thành mô hình chiếu sáng Blinn-phong. Với cải tiến này, mô hình chiếu sáng Blinn-phong cho hình ảnh giống thực tế hơn vì vậy, các bộ thư viện đồ họa phổ biến như OpenGL hoặc DirectX đã chuyển từ việc hỗ trợ mặc định mô hình Phong sang mô hình Blinn-Phong.

Theo các mô hình trên giả sử ta có một tập ảnh thu được từ đối tượng, tương ứng với mỗi ảnh là một tập tham số đặc trưng cho các nguồn sáng và camera. Khi đó nếu có thể giải ngược được phương trình 1 ta có thể ước lượng được chính xác bản đồ pháp tuyến và các hệ số phản xạ bề mặt của mô hình. Đây chính là mục tiêu chính của nghiên cứu.

2. GRADIEN DESCENT

Gradient Descent còn được gọi “xuống đồi theo hướng vector đạo hàm” là một phương pháp được sử dụng phổ biến nhất để giải các bài toán tối ưu cực tiểu hàm mục tiêu, nó đặc biệt được dùng nhiều trong bài toán tìm tham số cho các mô hình học máy [14, 16, 29].

Trong hầu hết trường hợp có thể định hiệu phương pháp tối ưu Gradient Descent như sau: Giả sử chúng ta cần đi tìm lời giải cho một hàm f_θ , với θ là các tham số cần tìm dựa trên một tập các giá trị quan sát được $Z = \{z_1, z_2, \dots, z_M\}$, khi đó ta đi xây dựng hàm mục tiêu:

$$C(\theta) = \frac{1}{M} \sum_{i=1}^M D(f_\theta, z_i) \quad (3)$$

trong đó $D(f_\theta, z_i)$ là hàm sai số của 1 mẫu quan sát, z_i là một mẫu quan sát có thể biểu diễn $z_i = (x_i, y_i)$ với $x_i \in R^n$ là đặc trưng đầu vào và $y_i \in R$ là kết quả của hàm $f_\theta(x_i)$.

Ý tưởng để giải bài toán tối ưu này là từ giá trị hiện tại của tham số θ , ta di chuyển tham số 1 khoảng trong không gian của nó để làm giảm giá trị hàm mục tiêu. Người ta đã chứng minh giá trị hàm mục tiêu giảm khi ta dịch chuyển θ ngược chiều đạo hàm của hàm mục tiêu tại θ tức là tại bước thứ $k+1$: $\theta_{k+1} = \theta_k - \alpha_k C'(\theta_k)$ thì luôn có $C(\theta_{k+1}) \leq C(\theta_k)$ với α_k đủ nhỏ. Thực hiện việc lặp cho đến khi giá trị hàm mục tiêu không giảm nữa có nghĩa ta đã tìm được bộ tham số θ sao cho hàm mục tiêu đạt cực tiểu, hay nói cách khác ta đã tìm được lời giải cho hàm f_θ (hàm lời).

III. ƯỚC LƯỢNG PHÁP TUYẾN BỀ MẶT

A. ƯỚC LƯỢNG PHÁP TUYẾN DỰA TRÊN BLINN-PHONG VÀ GRADIEN DESCENT

1. PHƯƠNG TRÌNH CHIẾU SÁNG

Để giải ngược được phương trình (1) chúng tôi đưa ra một số giả định để đơn giản hóa bài toán, các giả định này có thể thực hiện một cách dễ dàng khi thu thập ảnh dữ liệu:

- Chỉ có duy nhất một nguồn sáng đẳng hướng ảnh hưởng trực tiếp đến đối tượng, các giá trị đặc trưng cho nguồn sáng là các tham số đã biết.
- Góc và thông số camera là cố định, khoảng cách từ camera đến đối tượng lớn hơn nhiều so với kích thước của đối tượng do đó có thể coi phép chiếu của camera là phép chiếu song song khi đó hướng quan sát từ camera đến mọi điểm thuộc đối tượng luôn là $(0, 0, -1)$.
- Đối tượng được cấu thành từ một chất liệu có các hệ số phản xạ bề mặt như nhau trên mọi điểm của đối tượng. Giá trị màu sắc thu được khi quan sát đối tượng là tích hợp của ánh sáng và kết cấu bề mặt đối tượng.

Khi đó theo mô hình Blinn-Phong cường độ sáng tại mỗi điểm trên một ảnh trên mỗi kênh màu cơ bản với nguồn sáng $l_s = \{i_a, i_d, i_s, \vec{l} = (l_x, l_y, l_z)\}$ có thể được tính bằng công thức sau:

$$L(p, l_s) = k_a i_a + k_d i_d \text{Max}(\vec{l} \cdot \vec{n}_p, 0) + k_s i_s (\text{Max}(\vec{h} \cdot \vec{n}_p, 0))^\alpha \tag{5}$$

Giá trị quan sát được là giá trị chiếu sáng kết hợp với kết cấu bề mặt tại P như dưới đây:

$$I(p, l_s) = L(p, l_s) t_p \tag{6}$$

trong đó: i_a, i_d, i_s : là các giá trị đặc trưng cho các cường độ sáng (Ambient, Deffuse, Specular) $\in [0, 1]$; l_x, l_y, l_z : là các thành phần của vector chuẩn hóa đặc trưng cho hướng nguồn sáng; k_a, k_d, k_s : là các giá trị đặc trưng cho các hệ số phản xạ bề mặt của đối tượng (Ambient, Deffuse, Specular) $\in [0, 1]$; α là giá trị đặc trưng cho độ bóng của đối tượng $\alpha \in [0, \infty]$; n_{px}, n_{py}, n_{pz} : là các thành phần của vector chuẩn đặc trưng cho hướng pháp tuyến bề mặt tại điểm P; t^p : là giá trị $\in [0, 1]$ đặc trưng cho kết cấu của bề mặt tại P, có thể ước lượng trước được t^p bằng cách lấy trung bình của tập ảnh mẫu hoặc dựa vào ảnh của đối tượng trong điều kiện chỉ có sự chiếu sáng nền.

Trong các tham số của phương trình (5), các giá trị $k_a, k_d, k_s, \alpha, n_{px}, n_{py}, n_{pz}$ là các giá trị cần tìm các giá trị còn lại là các giá trị quan sát hoặc có thể ước lượng trước được. Hàm Max trong (5) là hàm trả về giá trị lớn hơn việc sử dụng hàm Max để đảm bảo tác động của các thành phần ánh sáng đến bề mặt đối tượng là không âm, điều này là hoàn toàn hợp lý và dễ hiểu.

2. XÂY DỰNG HÀM TỐI ƯU

Từ (5), (6) với một tập dữ liệu đầu vào $Z = \{z_1, z_2, z_M\}$; $M > 0$ trong đó $z_j = (I_{s_j}, l_{s_j})$, với I_{s_j} là ảnh thu được từ camera tương ứng với nguồn sáng l_{s_j} , khi đó ta có sai số của một mẫu quan sát j trên một điểm ảnh P được biểu diễn bằng công thức sau:

$$D_j(p) = (I_{s_{jp}} - I(p, l_{s_j}))^2 = (I_{s_{jp}} - L(p, l_{s_j}) t_p)^2 = \left(I_{s_{jp}} - \left(k_a i_{ja} + k_d i_{dj} \text{Max}(\vec{l}_j \cdot \vec{n}_p, 0) + k_s i_{js} (\text{Max}(\vec{h}_j \cdot \vec{n}_p, 0))^\alpha \right) t_p \right)^2 \tag{7}$$

Như vậy sai số trên toàn tập dữ liệu là:

$$C(\theta) = \frac{1}{M} \sum_{j=1, \dots, M} \frac{1}{|I_s|} \sum_{p \in I_s} D_j(p) \tag{8}$$

với $|I_s|$ là số điểm trên ảnh mẫu; $\theta = (k_a, k_d, k_s, \alpha, \vec{n}_{p0}, \vec{n}_{p1}, \dots, \vec{n}_{p|I_s|})$ là các tham số cần tìm. Công việc của chúng ta là đi tìm các θ sao cho giá trị hàm sai số trên toàn tập (hàm tối ưu) là nhỏ nhất. Để thực hiện việc này theo phương pháp Gradient Descent đầu tiên chúng ta tiến hành khởi tạo các tham số ban đầu cho θ sau đó thực hiện việc cập nhật θ lặp đi lặp lại với giá trị θ tại bước thứ $\eta+1$ là: $\theta_{\eta+1} = \theta_\eta - \delta_\eta C'(\theta_\eta)$ trong đó δ_η là tốc độ hội tụ δ_η càng nhỏ thì độ chính xác càng cao nhưng số vòng lặp càng lớn có nhiều phương pháp chọn δ_η tuy nhiên trong nghiên cứu này chúng tôi chọn δ_η theo phương pháp [19]. Như vậy công việc tiếp theo của chúng ta là đi tính các đạo hàm riêng theo từng thành phần của θ , tuy nhiên có một vấn đề trong công thức (7) đó là hàm Max là một hàm rời rạc không có đạo hàm do đó chúng ta phải tìm một liên tục hàm xấp xỉ cho hàm Max, hàm chúng tôi chọn là: $\frac{a}{1 + e^{-\xi a}}$, dễ thấy với ξ đủ lớn thì giá trị của hàm này là xấp xỉ bằng hàm Max. Như vậy công thức 7 có thể viết lại như sau:

$$D_j(p) = (I_{s_{jp}} - I(p, l_{s_j}))^2 = (I_{s_{jp}} - L(p, l_{s_j}) t_p)^2 = \left(I_{s_{jp}} - \left(k_a i_{ja} + k_d i_{dj} \frac{\vec{l}_j \cdot \vec{n}_p}{1 + e^{-\xi(\vec{l}_j \cdot \vec{n}_p)}} + k_s i_{js} \left(\frac{\vec{h}_j \cdot \vec{n}_p}{1 + e^{-\xi(\vec{h}_j \cdot \vec{n}_p)}} \right)^\alpha \right) t_p \right)^2 \tag{9}$$

3. CÔNG THỨC CẬP NHẬT THAM SỐ

Để cập nhật tham số θ của bước thứ $\eta+1$ đầu tiên ta phải tính các đạo hàm riêng của hàm tối ưu $C(\theta)$ theo từng thành phần của θ . Từ các công thức 7, 8, 9 có thể thấy các đạo hàm riêng theo từng thành phần của θ là:

- Đạo hàm riêng theo hệ số phản xạ nền (Ambient) k_a :

$$C'(\theta)_{k_a} = \frac{1}{M} \sum_{j=1, \dots, M} \frac{1}{|I_s|} \sum_{p \in I_s} 2i_{ja} t_p (L(p, l_{s_j}) t_p - I_{s_{jp}}) \tag{10}$$

- Đạo hàm riêng theo hệ số khếch tán (Diffuse) k_d :

$$C'(\theta)_{k_d} = \frac{1}{M} \sum_{j=1, \dots, M} \frac{1}{|I_s|} \sum_{\forall p \in I_s} 2i_{jd} t_p \frac{\vec{l}_j \cdot \vec{n}_p}{1 + e^{-\xi(\vec{l}_j, \vec{n}_p)}} (L(p, I_s)_j t_p - I_s_{jp}) \quad (11)$$

- Đạo hàm riêng theo hệ số phản xạ gương (Specular) k_s :

$$C'(\theta)_{k_s} = \frac{1}{M} \sum_{j=1, \dots, M} \frac{1}{|I_s|} \sum_{\forall p \in I_s} 2i_{js} \left(\frac{\vec{h}_j \cdot \vec{n}_p}{1 + e^{-\xi(\vec{h}_j, \vec{n}_p)}} \right)^\alpha t_p (L(p, I_s)_j t_p - I_s_{jp}) \quad (12)$$

- Đạo hàm riêng theo độ bóng bề mặt α :

$$C'(\theta)_\alpha = \frac{1}{M} \sum_{j=1, \dots, M} \frac{1}{|I_s|} \sum_{\forall p \in I_s} 2k_s i_{js} \left(\frac{\vec{h}_j \cdot \vec{n}_p}{1 + e^{-\xi(\vec{h}_j, \vec{n}_p)}} \right)^\alpha \ln \left(\frac{\vec{h}_j \cdot \vec{n}_p}{1 + e^{-\xi(\vec{h}_j, \vec{n}_p)}} \right) t_p (L(p, I_s)_j t_p - I_s_{jp}) \quad (13)$$

- Đạo hàm riêng của pháp tuyến tại P là tổng hợp của ba đạo hàm riêng thành phần $C'(\theta)_{\vec{n}_p} = (C'(\theta)_{n_{px}}, C'(\theta)_{n_{py}}, C'(\theta)_{n_{pz}})$, mỗi đạo hàm riêng thành phần có thể tính bằng công thức sau:

$$C'(\theta)_{\vec{n}_p} = \frac{2t_p}{M} \sum_{j=1, \dots, M} \left(\begin{array}{l} k_d i_{jd} \left(\frac{l_{jq} n_{pq}}{1 + e^{-\xi(\vec{l}_j, \vec{n}_p)}} + \frac{(\vec{l}_j \cdot \vec{n}_p) \xi e^{-\xi(\vec{l}_j, \vec{n}_p)}}{(1 + e^{-\xi(\vec{l}_j, \vec{n}_p)})^2} \right) + k_s i_{js} \alpha h_{jq} n_{pq} \\ \left(\frac{\vec{h}_j \cdot \vec{n}_p}{1 + e^{-\xi(\vec{h}_j, \vec{n}_p)}} \right)^{\alpha-1} \left(\frac{1}{1 + e^{-\xi(\vec{h}_j, \vec{n}_p)}} + \frac{(\vec{l}_j \cdot \vec{n}_p) \xi e^{-\xi(\vec{h}_j, \vec{n}_p)}}{(1 + e^{-\xi(\vec{h}_j, \vec{n}_p)})^2} \right) \end{array} \right) (L(p, I_s)_j t_p - I_s_{jp}) \quad (14)$$

trong đó q tương ứng với các thành phần của các vector pháp tuyến, hướng nguồn sáng và vector trung bình h có thể là x hoặc y hoặc z.

Như vậy giá trị của θ tại bước thứ $\eta+1$ là:

$$\theta_{\eta+1} = \left(\begin{array}{l} k_a + \delta_\eta C'(\theta)_{k_a}, k_d + \delta_\eta C'(\theta)_{k_d}, k_s + \delta_\eta C'(\theta)_{k_s}, \alpha + \delta_\eta C'(\theta)_\alpha, \\ \vec{n}_{p0} + \delta_\eta C'(\theta)_{\vec{n}_{p0}}, \vec{n}_{p1} + \delta_\eta C'(\theta)_{\vec{n}_{p1}}, \dots, \vec{n}_{p|I_s|} + \delta_\eta C'(\theta)_{\vec{n}_{p|I_s|}} \end{array} \right) \quad (15)$$

B. THUẬT TOÁN

Tương tự như thuật toán giải các bài toán tối ưu khác bằng Gradient Descent thuật toán gồm hai pha chính pha thứ nhất là khởi tạo tham số ban đầu, pha thứ hai là pha lặp trong đó thực hiện việc cập nhật các giá trị của θ cho đến khi các giá trị của hàm tối ưu là nhỏ nhất. Có thể mô tả sơ bộ thuật toán bằng giải mã như sau:

Đầu vào: Tập dữ liệu $Z = \{z_1, z_2, \dots, z_M\}$ với $z_j = \{I_s_j, I_s_j\}$ như đã mô tả phần trên

Đầu ra: các giá trị của $\theta = (k_a, k_d, k_s, \alpha, \vec{n}_{p0}, \vec{n}_{p1}, \dots, \vec{n}_{p|I_s|})$

Các bước chính của thuật toán:

Bước 1: Ước lượng ảnh kết cấu bề mặt t.

Bước 2: Khởi tạo các tham số ban đầu gồm (θ , bước nhảy step, và các biến môi trường khác). Tính giá trị hàm tối ưu với tham số θ .

Bước 3: Cập nhật giá trị cho tham số θ theo các công thức (10-15), cập nhật lại bước nhảy. Tính giá trị hàm tối ưu tương ứng với θ mới.

Bước 4: Lặp lại bước 2 cho đến khi sai khác giữa giá trị hàm tối ưu của bước sau với bước trước nhỏ hơn một ngưỡng cho trước.

Ảnh kết cấu bề mặt “t” đặc trưng cho phân bố màu sắc cơ bản trên bề mặt của đối tượng. Có thể ước lượng t bằng một số cách thức ví dụ: lấy ảnh với điều kiện chiếu sáng môi trường không có nguồn sáng nào ảnh hưởng trực tiếp đến đối tượng, hoặc có thể ước lượng t bằng cách tính trung bình của toàn bộ tập ảnh thu được với nhiều điều kiện chiếu sáng khác nhau phân phối đều trong không gian v.v...

Việc khởi tạo tham số ban đầu có vai trò quan trọng ảnh hưởng quyết định đến kết quả của thuật toán. Nếu tham số ban đầu càng gần với giá trị thật thì tốc độ hội tụ và độ chính xác của thuật toán càng cao. Trong θ có thể phân thành

hai loại tham số: Loại thứ nhất đó là thông số bề mặt của đối tượng quan tâm thứ hai là tập các vector pháp tuyến trên bề mặt đối tượng. Các thông số bề mặt đối tượng chúng ta có thể lựa chọn dựa trên kinh nghiệm, đặc biệt là kinh nghiệm của những họa sỹ vẽ mô hình ba chiều khi lựa chọn các thông số bề mặt cho đối tượng cần thể hiện. Giá trị của các thông số bề mặt đưa vào thuật toán có mối liên hệ tương đồng so với giá trị thiết lập khi sử dụng các phần mềm ba chiều để tái hiện lại mô hình. Với tập vector pháp tuyến chúng tôi khởi tạo giá trị ban đầu cho nó là giá trị ngược chiều với hướng quan sát $(0, 0, 1)$. Sở dĩ lựa chọn giá trị này là vì khi quan sát một đối tượng, chúng ta chỉ có thể nhìn thấy những điểm có vector pháp tuyến ngược chiều với hướng quan sát đồng thời khu vực có hướng pháp tuyến gần giá trị $(0, 0, 1)$ thường là lớn nhất do đó giá trị này sẽ là giá trị gần đúng nhất trên toàn ảnh.

Tham số θ được cập nhật theo công thức (15), việc cập nhật θ có thể đưa các giá trị thành phần ra khỏi tập xác định giới hạn của chúng vì vậy sau bước này chúng ta phải hiệu chỉnh lại các giá trị thành phần về khoảng giới hạn giá trị sau khi hiệu chỉnh mới là giá trị được sử dụng cho bước tiếp theo.

IV. THỬ NGHIỆM

Chúng tôi đã tiến hành cài đặt thuật toán đề xuất trên ngôn ngữ lập trình C++ với bộ công cụ phát triển Eclipse trên nền tảng hệ điều hành Linux. Sau khi cài đặt chúng tôi tiến hành thu thập dữ liệu và thử nghiệm trên hai tập dữ liệu: một được sinh ra từ mô hình ba chiều bằng một công cụ do chúng tôi xây dựng, tập khác được lấy từ nguồn từ <http://vision.seas.harvard.edu/qsfs/Data.http>, đây là tập dữ liệu ảnh các đối tượng thu được từ camera với góc quan sát và thông số camera cố định, được chiếu sáng bởi một nguồn sáng đẳng hướng. Tập dữ liệu này được xây dựng bởi Ying Xiong và cộng sự phục vụ cho bài toán xây dựng lại các thông số bề mặt từ ảnh.

A. THỬ NGHIỆM TRÊN TẬP DỮ LIỆU SINH TỪ MÔ HÌNH

Để thực hiện các thực nghiệm này chúng tôi sử dụng một số hiện vật ba chiều có được trong quá trình xây dựng Bảo tàng Nhân học - Đại học Quốc gia Hà Nội. Cụ thể chúng tôi đã sử dụng 10 hiện vật, mỗi hiện vật tiến hành render để được 10 ảnh kết quả, các thông số camera trong các lần render là cố định, hướng nguồn sáng thay đổi đều theo các hướng khác nhau. Hình dưới đây là một số ảnh hiện vật thử nghiệm.



Hình 3. Một số ảnh hiện vật trong ca thử nghiệm

Chúng tôi cũng tiến hành chạy thử thuật toán trên các bộ dữ liệu có được sau đó, đối sánh kết quả bản đồ pháp tuyến thu được với hướng pháp tuyến trên mô hình ba chiều mẫu. Việc đối sánh dựa trên hai tiêu chí đó là khoảng cách trung bình và khoảng cách lớn nhất giữa các vector pháp tuyến trên bản đồ pháp tuyến kết quả và hướng pháp tuyến trên mô hình mẫu. Dưới đây là bảng kết quả thu được sau khi tiến hành thử nghiệm.

Bảng 1. Kết quả thực nghiệm trên tập ảnh thu được từ các mô hình 3D mẫu

STT	Tên hiện vật	Số ảnh	Sai số trung bình	Sai số cực đại
1	Giáo đồng	10	0.0138	0.0556
2	Bình gốm	10	0.0371	0.1878
3	Gạch 1	10	0.0092	0.0622
4	Gạch 2	10	0.0085	0.0798
5	Mũi tên đồng	10	0.0096	0.0643
6	Trống đồng	10	0.0103	0.0687
7	Đĩa men ngọc	10	0.0587	0.2453
8	Bình gốm men rạn	10	0.0196	0.1748
9	Ngói hoa sen	10	0.0098	0.0843
10	Ngói vẩy cá	10	0.0089	0.0794








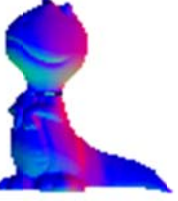

Các kết quả của thuật toán trên tập dữ liệu thử nghiệm là tương đối tốt với các sai số trung bình và sai số cực đại là khá nhỏ. Với kết quả này bản đồ pháp tuyến thu được ngoài việc sử dụng cho việc tăng cường chất lượng trong việc

render còn có thể sử dụng trong các mục đích khác yêu cầu độ chính xác cao như nhận dạng đối tượng, khôi phục hình dạng đối tượng v.v.. Bảng kết quả (1) cũng chỉ ra với các đối tượng có độ bóng cao như đĩa men ngọc hoặc bình gốm sai số vẫn cao hơn đáng kể so với những đối tượng khác nhưng vẫn ở mức chấp nhận được. Với kết quả này có thể khẳng định thuật toán phần nào khắc phục được nhược điểm của một số thuật toán khác là khó áp dụng cho những hiện vật có độ bóng bề mặt cao.

B. THỬ NGHIỆM TRÊN DỮ LIỆU THU ĐƯỢC TỪ CAMERA

Việc thiết kế một hệ thống thu nhận dữ liệu từ camera là không khó tuy nhiên để đảm bảo tính khách quan chúng tôi sử dụng tập dữ liệu được xây dựng bởi Ying Xiong và cộng sự. Chúng tôi tiến hành chạy thuật toán đề xuất và thuật toán của Qingxiong Yanga và cộng sự để được một ảnh pháp tuyến bề mặt. Tiếp theo chúng tôi sử dụng chính ảnh pháp tuyến bề mặt kết quả làm đầu vào cho việc render lại ảnh chiếu sáng với nguồn sáng là dữ liệu đầu vào của thuật toán xây dựng pháp tuyến, đối sánh ảnh kết quả thu được từ quá trình render với ảnh đầu vào của thuật toán trong ứng thể hiện ở bảng dưới đây.

Bảng 2. Một số kết quả thử nghiệm trên tập cơ sở dữ liệu Ying Xiong và cộng sự

Tập ảnh ban đầu	Số lượng	Ảnh bản đồ pháp tuyến (kỹ thuật đề xuất)	Ảnh sau render lại (kỹ thuật đề xuất)	Sai số kỹ thuật đề xuất	Sai số Qingxiong
	20			0.080	0.103
	20			0.075	0.0962
	20			0.093	0.0984

Nhìn chung trên tập dữ liệu của Ying Xiong kết quả thực hiện của thuật toán đều tốt hơn so với kết quả của Qingxiong [22]. Tuy nhiên kỹ thuật đề xuất cũng có một nhược điểm tương tự như nhược điểm của các thuật toán lập tối ưu khác đó là tốc độ thực hiện tương đối chậm trung bình khoảng 5.376 phút trên tập dữ liệu mẫu nhưng vấn đề này hoàn toàn có thể chấp nhận được vì việc tính bản đồ pháp tuyến không yêu cầu thời gian thực. Cũng có thể tăng tốc đáng kể cho thuật toán này nếu triển khai trên nền tảng GPU đây cũng là hướng tiếp theo của nhóm tác giả.

V. KẾT LUẬN

Ước lượng pháp tuyến bề mặt là một bài toán quan trọng, có nhiều ứng dụng trong thực tại ảo, đồ họa máy tính và thị giác máy. Trong báo cáo này chúng tôi đã đề xuất một kỹ thuật ước lượng pháp tuyến bề mặt dựa trên mô hình chiếu sáng Blinn-Phong và phương pháp lập tối ưu Gradient-Descent. Kỹ thuật đề xuất đã được cài đặt thử nghiệm trên hai tập dữ liệu: một được sinh ra từ mô hình ba chiều, một được thu thập từ camera trong các điều kiện giới hạn, các kết quả trên cả hai tập dữ liệu thử nghiệm có khả năng ứng dụng tốt cho bài toán tăng chất lượng hiển thị mô hình trong đồ họa ba chiều, xa hơn nữa có khả năng ứng dụng trực tiếp trong bài toán nhận dạng mặt người vì đặc điểm pháp tuyến là bất biến với ánh sáng, đây cũng là hướng tiếp theo của nghiên cứu này.

VI. TÀI LIỆU THAM KHẢO

- [1] Aiyun Qing, et al. "Face recognition under generic illumination based on harmonic lighting" International Journal of Pattern Recognition and Artificial Intelligence Vol.19, No.4513–531(2005).
- [2] Alexandros Panagopoulos, Sunil Hadap, Dimitris Samaras "Reconstructing Shape from Dictionaries of Shading Primitives" Lecture Notes in Computer Science Volume 7727, pp 80-94 (2013).
- [3] David C. Wyld et al. (Eds) "3D Modelling simulation and prediction of facial wrinkles" CCSIT, SIPP, AISC, PDCTA, NLP pp. 317–326, (2014).
- [4] Giuseppe Claudio Guarnera, Pieter Peers, Paul Debevec, Abhijeet Ghosh "Estimating Surface Normals from Spherical Stokes Reflectance Fields" Lecture Notes in Computer Science Volume 7584, pp 340-349 (2012)

- [5] Giljoo Nam et al. "Multispectral Photometric Stereo for Acquiring High-Fidelity Surface Normals" *Computer Graphics and Applications*, IEEE (Volume:34 , Issue: 6) p.57 - 68 (2014).
- [6] Gouraud, Henri "Continuous shading of curved surfaces". *IEEE Transactions on Computers* C-20 (6): 623–629. doi:10.1109/T-C.1971.223313.(1971).
- [7] Holzer, et al "Adaptive neighborhood selection for real-time surface normal estimation from organized point cloud data using integral images" *Intelligent Robots and Systems (IROS)*, IEEE/RSJ International Conf(2012)
- [8] Jacquet, et al "Real-World Normal Map Capture for Nearly Flat Reflective Surfaces" *Computer Vision (ICCV)*, p.713 – 720, (2013).
- [9] James F. Blinn. "Models of light reflection for computer synthesized pictures". *Proc. 4th annual conference on computer graphics and interactive techniques*: 192–198. doi:10.1145/563858.563893. (1977).
- [10] Jan Kautz, Wolfgang Heidrich, Hans-Peter Seide "Real-time bump map synthesis" *HWWS '01 Proceedings of the ACM SIGGRAPH/EUROGRAPHICS workshop on Graphics hardware* Pages 109-114 ACM New York, NY, USA © (2001).
- [11] Jian Ren, Xiquan Cui, Lap Man Lee, and Changhui Yang "Quantitative surface normal measurement by a wavefront camera" *Optics Letters*, Vol. 37, Issue 2, pp. 199-201 (2012).
- [12] J. Filip & M. Haindl "BTF Modelling using BRDF Texels" *Advances in Machine Vision, Image Processing, and Pattern Analysis Lecture Notes in Computer Science* Volume 4153, pp 475-484 (2006).
- [13] Kaneko, T., et al. "Detailed Shape Representation with Parallax Mapping" *Proceedings of ICAT* (2001).
- [14] Kiwiel, Krzysztof "Methods of Descent for Nondifferentiable Optimization". Berlin: Springer Verlag. p. 362. ISBN 978-3540156420. MR 0797754 (1985).
- [15] Marco Tarini, et al. "Real Time, Accurate, Multi-Featured Rendering of Bump Mapped Surfaces" *Computer Graphics Forum (Eurographics Conference Issue)*, Blackwell Publishers, vol. 19(3), pp 119-130 (2000).
- [16] Ma et al. "Rapid Acquisition of Specular and Diffuse Normal maps from Polarized Spherical Gradient Illumination" *Proceeding of Render techniques*, Jun. 2007, pp. 183-194 (2007).
- [17] M. Okabe et al. "Single-View Relighting with Normal Map Painting" *Proc. Pacific Graphics*, pp. 27-34, (2006).
- [18] Morten Mikkelsen "Simulation of Wrinkled Surfaces Revisited" *ACM Transactions on Graphics* 5(1), pp 30–50 (2008).
- [19] Bertsekas, Dimitri P. "Convex Optimization Algorithms" (Second ed.). Belmont, MA.: AthenaScientific. ISBN 978-1-886529-28-1 (2015).
- [20] L'ubor Ladický, Bernhard Zeisl, Marc Pollefeys "Discriminatively Trained Dense Surface Normal Estimation" *Computer Vision – ECCV 2014 Lecture Notes in Computer Science* Volume 8693, pp 468-484 (2014).
- [21] Phong Bùi Tường "Illumination for computer generated pictures", *Communications of ACM* 18, no. 6, pp 311–317 (1975).
- [22] Qingxiong Yanga, Narendra Ahujab "Surface reflectance and normal estimation from photometric stereo" *Computer Vision and Image Understanding* Volume 116, Issue 7, July 2012, Pages 793–802 (2012)
- [23] Robert Anderson, Björn Stenger, Vincent Wan, Roberto Cipolla "An expressive text-driven 3D talking head" *SIGGRAPH '13 ACM SIGGRAPH 2013 Posters* (2013).
- [24] Rongjiang Pan et al. "Normal Map Acquisition of Nearly Flat Objects Using a Flatbed Scanner" *Virtual Reality and Visualization (ICVRV)*, 2013 International Conference on 68 - 73 14-15 Sept. 2013.
- [25] Sema Berkiten, Xinyi Fan, Szymon Rusinkiewicz "Merge2-3D: Combining Multiple Normal Maps with 3D Surfaces" *3DV 2014, International Conference on 3D Vision*, December 2014.
- [26] TAI-PANG WU et al "Interactive normal reconstruction from a single image" *ACM SIGGRAPH 2008 Pa No. 119* ACM New York, NY, USA ©2008 ISBN: 978-1-4503-1831 (2008).
- [27] Xiaolong Wang, David F. Fouhey, Abhinav Gupta "Designing Deep Networks for Surface Normal Estimation" *Computer Science-Computer Vision and Pattern Recognition* (2014).
- [28] Yannick Francken et al. "Fast Normal Map Acquisition Using an LCD Screen Emitting Gradient Patterns" *Computer and Robot Vision, 2008. CRV '08. Canadian Conference* (2008).
- [29] Yuan, Ya-xiang "Step-sizes for the gradient method" (PDF). *AMS/IP Studies in Advanced Mathematics* (Providence, RI: American Mathematical Society) 42 (2): 785.(1999)

- [30] Yusuke Yamaura et al. "Shape reconstruction from a normal map in terms of uniform bi-quadratic B-spline surfaces" *Computer-Aided Design* Volume 63, June 2015, Pages 129–140 (2015).
- [31] Wolfgang Heidrich et al "Realistic, hardware-accelerated shading and lighting" *SIGGRAPH '99 Proceedings of the 26th annual conference on Computer graphics and interactive pp171-178 ACM (1999).*

AN ESTIMATED NORMAL MAP TECHNIQUE BASED ON IMAGES

Trinh Xuan Hung, Do Nang Toan, Do Van Thien

ABSTRACT- In this paper, we propose a technique that helpful in building Normal Map from object images which are collected by static camera with the change of light direction. The proposal technique uses iterative optimization Gradient descent to inverse solve the Blinn-Phong function with some assumptions. We run the proposal technique with two datasets: one-we collect from 3d model with different types of lighting, the other is collection images from camera with definite light built by Ying Xiong et al. We get good result with 2 data sets in acceptable time.