國立台灣科技大學
資訊工程研究所

碩 士 論 文

3D 點雲與 SIFT 特徵點室內影像定位之研究

3D Point Cloud and SIFT Descriptor Indoor Localization Research

研 究 生：陳致良
學 號：M9915057

指 導 教 授：項天瑞博士

中 華 民 國 一 百 零 一 年 七 月 十 日

3D 點雲與 SIFT 特徵點室內影像定位之研究

學生： 陳致良　　指導教授： 項天瑞博士

國立台灣科技大學資訊工程研究所

摘　　要

本篇論文探討如何利用單張彩色影像來重建出三維人臉模型。我們的方法是利用彩色影像來取代灰階值去建立張量模型 (tensor model)，在人臉資料庫中，是用典型相關分析 (Canonical correlation analysis) 是來建立彩色影像與深度資訊的對應關係，一旦建立好屬於各自的對應關係後，在重建人臉的過程中，就只需要單張彩色影像即可利用典型相關分析的對應關係來推算出正確的深度資訊。實驗中，我們的方法可以在不同的光線環境跟人臉角度得到不錯的效果。

3D Point Cloud and SIFT Descriptor Indoor Localization Research

Student: ZZhi-Liang Chen Advisor: Dr. Tien-Ruey Hsiang

Submitted to Department of Computer Science and Information Engineering
College of Electrical Engineering and Computer Science
National Taiwan University of Science and Technology

ABSTRACT

This paper develops a tensor-based 3D face reconstruction approach from a single color image. Instead of the grayscale image, we also consider additional color factors in constructing the tensor model. Canonical correlation analysis is applied to establish the relationship between the color image and the depth information in the face database. During the face reconstruction, given a single color face image, the depth estimation is computed from the CCA-based mapping between the tensor models. Experimental results show our approach is better suited under different lighting conditions and poses.

誌　　謝

本論文能夠完成，首先要感謝的是指導教授項天瑞老師。老師嚴謹的治學態度，讓我不但在學術研究上學習到更謹慎的思考，也在日常生活上獲益良多。

感謝實驗室的同學們，實驗室的生活有苦有樂，有你們才讓我能撐得下這三年漫長的時間。感謝建群、松翰學長給我的指導。感謝益偉、崇峰、承誌、誠儀、慶豪學長以及恩緯、盈樽、嘉駿、世寬學長給我的指導。感謝實驗室的學弟妹們，訓哲、立昂、致良、青緯、庭耀、宗博、雅筑，常常幫了我不少忙。更要感謝我的同梯們，貴彥、薇穎、冠佑，我們同甘共苦，一起奮鬥，特別是在弄計畫的過程中，冠佑幫助了我很多，有你們在我才能走到這一步。也感謝好朋友志強跟我一起修正文法與用詞。

最後我要把我最深的感謝留給我的家人。謝謝我的爸媽跟兩位姊姊，你們讓我沒有經濟壓力地讀完這這碩士學位，也常常給我很多鼓勵，今天我終於拿到這個學位，終於可以讓你們放下心上的一塊大石頭了。

Table of Contents

List of Tables

List of Figures

Chapter 1 Related Work

尺度不變特徵向量 (SIFT) 一開始由 lowe [1] 所提出，目的是尋找兩張圖片中的相似特徵向量來比對兩張圖片的相對關係，主要分成四個階段：

- (1) 區域空間極值分布
- (2) 特徵點定位與篩選
- (3) 特徵點方向分配
- (4) 特徵點描述向量建立

第一階段 (1) 區域空間極值篩選，先利用不同尺度間的高斯金字塔選擇區域中的最大極值，其高斯分布式子如下：

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} exp(-(x^2 + y^2)/2\sigma^2) \tag{1.1}$$

不同尺度的高斯分布利用摺積 (Convolution) 將影像模糊化。$I(x, y)$ 代表原始影像，$G(x, y, \sigma)$ 代表高斯函數：

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{1.2}$$

再利用每組影像相鄰的高絲模糊影像進行高斯差分 (Difference-Of-Gaussian)，目的用於在集合內 4 組高斯差分影像中找出極值，式子如 (1.3) 所示：

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \tag{1.3}$$

在此 k 為高斯模糊的尺度比值，設為 $\sqrt{2}$，若某個像素的極值為 26 個相鄰的像素中最大或最小的話，則此像素的位址即為區域極值的所在。

第二階段 (2) 特徵點定位篩選，其主要的目的在於找出真正有用的特徵點，在此特徵點的精準度必須要達到次像素的精度。有的特徵點其極值為低對比

度的點，這時候這些低對比度的特徵點就會不予採用，剩下的特徵點即可為下一階段所使用。作法首先將 (1.3) 利用泰勒展開得到 (1.4):

$$D(x) = D + \frac{\delta D^T}{\delta X}X + \frac{1}{2}X^T\frac{\delta^2 D}{\delta X^2}X \tag{1.4}$$

式中 X 為極值 $(x, y, \delta)^T$，D 為高斯差分後的結果，再將 (1.4) 對 X 作偏微分可得 $\vec{X}$ 算出 X 為極值點的的偏移量。

$$\vec{X} = -\frac{\delta^2 D}{\delta X^2}^{-1}\frac{\delta D}{\delta X} \tag{1.5}$$

若是 $\vec{X} >= 0.5$，或是 $\sigma > k/2$，表示此區域極值點較靠近相鄰的點位，則需要再將此點移至相鄰的極值再經 (1.5) 計算後得到最佳的位置。若將 $\vec{X}$ 帶入 (1.4) 中，可得 (1.6) 我們所用來篩選的式子:

$$D(\vec{X}) = D + \frac{1}{2}\frac{\sigma D^T}{\sigma X}\vec{X} \tag{1.6}$$

利用 (1.6) 將求出的絕對值與其他絕對值相比，可將對比度小的特徵點刪除以達到過濾的效果。

第三階段 (3) 特徵點方向分配，目的在於當對比的圖片有旋轉或者是尺度上的變化，相同的特徵點為了保有相同方向的特性，必須賦予每個特徵點一組特定的方向。其做法則是利用統計的方式，將所有的梯度值以角度每 10 個單位做值方圖記錄，並且記錄每個直方圖的強度，以 (1.7) 表示：

$$\mathrm{m(x,y)} = \sqrt{(\frac{\delta L}{\delta x})^2 + (\frac{\delta L}{\delta y})^2}$$

Face reconstruction and recognition in digital photographs are still challenging problems. Principal component analysis (PCA) [2] and independent component analysis (ICA) [3] are popular methods used in facial image process and other image analysis topics. In fact, other factors such as lighting, viewpoint, and expression may also be useful in particular applications. However, these methods often address only single factor variations in images. For example, extra information may strengthen the representation of a person's identity, thus improve the performance in recognition.

In PCA-based face reconstruction algorithms, a face is represented with a shape vector $S = (X_1, Y_1, Z_1, X_2, ..., Y_n, Z_n)^T \in \mathbb{R}^{3n}$, containing the X, Y, Z coordinates of its n vertices. Then, let $\bar{S}$ be the average representation of a face, $P \in \mathbb{R}^{3n \times m}$ be the matrix of the first m eigenvectors. A new face S' can be expressed as

$$S' = \bar{S} + P\vec{\alpha} \tag{1.7}$$

where $\vec{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_m)^T \in \mathbb{R}^m$ is the coefficients of the shape eigenvectors. Romdhani [4] extends Inverse Composition Image Alignment (ICIA) [5] algorithm which fits 2D models. It first defines the representation of the 3D shape. In order to fit the shape, it needs to solve the inverse shape projection problem. The texture model is also fitted in the same way. Then, the iterative ICIA algorithm is used to improve the fitting precision efficiently. However, this alignment algorithm requires manual initialization and is usually slow in fitting a face. Hu [6] proposed an automatic linear algorithm to speed up the reconstruction process according to sparsely corresponded 2D facial feature points. The whole reconstruction process starts from locating key facial points without manual labeling in a frontal view input image. Next, iteratively computing the coefficients α and using PCA coefficients of eigenvectors to reconstruct 3D shape. The texture is extracted from the input image.

Aligning faces and finding the facial landmark points on frontal face image are two important issues. In this paper, we apply the algorithm introduced in [7] that locates facial features with an extended active shape model [8] to align faces. [8] proposed a method to find a set of points, known as landmarks, to describe a

shape of image object. By the ASM refinement algorithm, shapes are represented as one-dimensional vectors. To obtain more reliable profile matches, [7] uses a two-dimensional landmark templates instead of a one-dimensional one. A 2D profile area captures more information around the landmark and this information can be used to fit a better result. To get accurate position of start shape, [7] uses two ASM searches in series, using the results of the first search as the start shape for the second search. During the training process, they adds noise to the training shapes and that helps the trained model can generalize a wider variety of face. Obtained by [7], we can get 76 facial points automatically in the front image without manual labeling.

In recent years, [9,10] used tensor techniques in face recognition. A tensor is a multidimensional array (or N-way array). A matrix is considered as a second order tensor. High-order tensors have applications in computer vision, numerical analysis, signal processing. In matrix decomposition, a matrix $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2}$ is a two mode mathematical object that contains a row vector and a column vector. SVD orthogonalizes these two spaces and decomposes the matrix as $D = U_1 \sum U_2^T$. We follow the notations used in [9,10], the SVD decomposition can be expressed as $D = \sum \times_1 U_1 \times_2 U_2$, where $\times_i$ is the mode-i product. N-mode SVD is an extension of SVD that orthogonalizes these N spaces and expresses the tensor as the mode-n products of N-orthogonal spaces

$$D = Z \times_1 U_1 \times_2 U_2 ... \times_n U_n ... \times_N U_N \tag{1.8}$$

as illustrated in Fig 1.1, where tensor Z in the core tensor. The core tensor governs the interaction between the mode matrices U_n, for $n = 1, ..., N$. And mode matrix U_n contains the orthogonal vectors spanning the column space of the mode-n flattening of D.

Lei et al. [11] proposed a new method for recovering a face from a single image. They use a single near infrared (NIR) image as input, and construct a mapping from the NIR tensor space to 3D tensor space. Statistical learning techniques are applied to remove noises and redundant information among parameter vectors form tensor spaces. The NIR images are captured from active NIR imaging system [12],

which can provide strong lights to produce a clear frontal-lighted face image without causing disturbance to human eyes and minimize environmental lighting. Images produced by NIR imaging systems are much less sensitive to environmental lighting. The approach proposed by [11] is compared against two other methods, named CCA [13] and CSM [14]. Their experimental results show that by using NIR images, the reconstructed faces are better than using visible light images as the input.

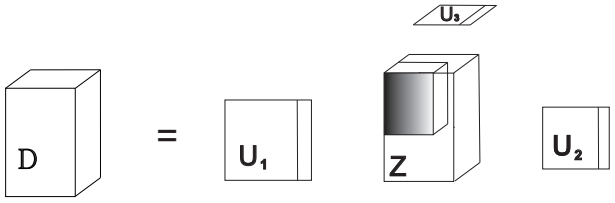

Figure 1.1: An N-mode SVD orthogonalizes the N vector spaces associated with an order-N order (the case $N = 3$).

Bibliography

[1] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[2] I. Jolliffe and MyiLibrary, *Principal component analysis*. Wiley Online Library, 2002, vol. 2.

[3] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*. Wiley-interscience, 2001, vol. 26.

[4] S. Romdhani and T. Vetter, "Efficient, robust and accurate fitting of a 3d morphable model," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. IEEE, 2003, pp. 59–66.

[5] S. Baker and I. Matthews, "Equivalence and efficiency of image alignment algorithms," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. I–1090.

[6] Y. Hu, D. Jiang, S. Yan, L. Zhang *et al.*, "Automatic 3d reconstruction for face recognition," in *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*. IEEE, 2004, pp. 843–848.

[7] S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," *Computer Vision–ECCV 2008*, pp. 504–513, 2008.

[8] T. Cootes, C. Taylor, D. Cooper, J. Graham *et al.*, "Active shape models-their training and application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[9] M. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," *Computer Vision—ECCV 2002*, pp. 447–460, 2002.

[10] T. Kolda and B. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, 2009.

[11] Z. Lei, Q. Bai, R. He, and S. Li, "Face shape recovery from a single image using cca mapping between tensor spaces," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–7.

[12] S. Li, R. Chu, S. Liao, and L. Zhang, "Illumination invariant face recognition using near-infrared images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 4, pp. 627–639, 2007.

[13] M. Reiter, R. Dormer, G. Langs, and H. Bischof, "3d and infrared face reconstruction from rgb data using canonical correlation analysis," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 1. IEEE, 2006, pp. 425–428.

[14] M. Castelan and E. Hancock, "A simple coupled statistical model for 3d face shape recovery," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 1. IEEE, 2006, pp. 231–234.