Ajou UNIVERSITY

Industrial Mathematics Project

FINAL REPORT

Optimal Solution of Rare Pose Detection

POSE!
201621146 Chae won Kang
201621136 Jae hyeob Lee
201521037 Chang hun Han
201621145 Hye mi Kim

Index

0. Abstract

This study aimed to find Rare pose in given dataset. There are several reasons of errors in Human pose estimation and we are focus on the case where Rare pose causes error.

First of all, we expressed each Human pose with 16 3D-vectors. To obtain a probability for vectors, 8-Parameter Fisher-Bingham Distribution(FB8) was used as a model. Also, parameters for each distribution was obtained by Maximum Likelihood Estimation(MLE). Then, we calculate rareness of pose by multiplying the Probability Density Function(PDF) of each vectors in single pose.

Finally, it was reached to identify the rareness pose by applying the previously calculated distributions.

We expect that intensive Rare Pose learning will help to increase the performance of 3D Human Pose estimation.

1. Introduction

3D pose estimation is a process of predicting the transformation of an object from a user-defined reference pose, given an image or a 3D scan. It arises in computer vision or robotics where the pose or transformation of an object can be used for alignment of a Computer-Aided Design models, identification, grasping, or manipulation of the object.

Recently, research on 3D Human Pose estimation has been actively conducted. Deep learning is mainly used for 3D Human Pose estimation. In particular, remarkable results are obtained by Convolutional Neural Networks(CNNs).

However, poor performance was observed on certain postures. It was usually in situations such as Rare pose, obscured body parts, and overlaps with other people. Among these, this study noted when Rare Pose occurs. If we knew the Rare Pose among Dataset, we can increase the accuracy of Pose estimation by adjusting the Rare Pose.

The purpose of this study is to identify the distribution of each joint vector, to find rare pose by calculating the rareness. By finding Rare pose and properly learning it will complement the problems of the existing dataset and improve its performance.

2. Related Woks

2.1 Fisher-Bingham Distribution

Kent is a probability distribution on the two-dimensional unit sphere S^2 in R^3. It is the analogue on the two-dimensional unit sphere of the bivariate normal distribution with an unconstrained covariance matrix.

The probability density function f(x) of the Kent distribution is given by

$$f(x) = \frac{1}{c(k,\beta)} \exp\{k\gamma_1^T \cdot x + \beta[(\gamma_2^T \cdot x)^2 - (\gamma_3^T \cdot x)^2]\} \quad \dots (1)$$

where x is a three-dimensional unit vector, $(.)^T$ denotes the transpose of(.), and the normalizing constant $c(k,\beta)$ is

$$c(k,\beta) = 2\pi \sum_{j=0}^{\infty} \frac{\Gamma(j+\frac{1}{2})}{\Gamma(j+1)} \beta^{2j} \left(\frac{1}{2}k\right)^{-2j-\frac{1}{2}} \mathrm{I}_{2j+\frac{1}{2}}(k) \quad \dots (2)$$

Where $\mathrm{I}_v(k)$ is the modified Bessel function and $\Gamma(.)$ Is the gamma function. Note that $c(0,0) = 4\pi(k^{-1})\sinh(k)$, the normalizing constant of the Von Mises – Fisher distribution.

The parameter k (with $k > 0$) determines the concentration or spread of the distribution, while β (with $0 \leq 2\beta < k$) determines the ellipticity of the contours of equal probability. The higher the k and β parameters, the more concentrated and elliptical the distribution will be, respectively. Vector γ_1 is the mean direction, and vectors $\gamma_2, \gamma_3$are the major and minor axes. The latter two vectors determine the orientation of the equal probability contours on the sphere, while the first vector determines the common center of the contours. The 3x3 matrix $(\gamma_1\gamma_2\gamma_3)$ must be orthogonal.

2.2 von Mises – Fisher Distribution

In directional statistics, the von Mises-Fisher distribution is a probability distribution on the (p-1)-sphere in R^p. If p=2 the distribution reduces to the von Mises distribution for the random p-dimensional unit vector X is given by

$$f_p(x;\mu,k) = C_p(k)\exp(k\mu^T x) \quad \cdots \ (3)$$

where $k \geq 0, \left||u|\right| = 1$ and the normalization constant $C_p(k)$ is equal to

$$C_p(k) = \frac{k^{\frac{p}{2}-1}}{(2\pi)^{\frac{p}{2}} I_{\frac{p}{2}-1}(k)} \quad \ldots (4)$$

where I_v denotes the modified *Bessel function* of the first kind at order v. If p=3 the normalization constant reduces to

$$C_3(k) = \frac{k}{4\pi \sinh k} = \frac{k}{2\pi(e^k - e^{-k})} \quad \ldots (5)$$

The parameters u and k are called the mean direction and concentration parameter, respectively. The greater the value of k, the higher the concentration of the distribution around the mean direction u. The distribution is unimodal for k>0, and is uniform on the sphere for k=0

The von Mises-Fisher distribution for p=3, also called Fisher distribution, was first used to model the interaction of electric dipoles in an electric field. Other applications are found in geology, bioinformatics, and text mining.

2.3 Maximum Likelihood Estimation

Maximum Likelihood Estimation (MLE) is one of the methods of estimate the parameter of Random Variable. It is a method of parameter estimation based on already given observation through several implementations. Let's find out by example.

Suppose we predict p by throwing a coin with a front coming out with a probability of p and a back coming out with a probability of 1-p. To calculate p by MLE, simply divide the number of front faces by the total number of times.

For a more detailed explanation, suppose you have an unknown probability density function f. And let's say X is the probability that it generates. Assuming that density function is a family of a distribution parameterized by θ as follows $\{f(\cdot\,|\theta)\}$, a value of k can be calculated as soon as the value of θ is immediately known, provided the observation $\{f(\cdot\,|\theta)\}$ is given. If f is Gaussian, θ will be mean μ and covariance Σ, and if it is Bernoulli, it will be $0 \leq p \leq 1$. By this definition, Likelihood can define as follows:

$$L(\theta;\, x_1, x_2, \cdots, x_n) \;=\; L(\theta; X) \;=\; f(X|\theta) \;=\; f(x_1, x_2, \cdots, x_n|\theta)$$

Maximum Likelihood Estimate (MLE) is one of the ways to estimate θ, and it is to choose a value that maximizes likelihood. If observation is i.i.d., it becomes $\hat{\theta} = arg\, max_\theta\, L(\theta; X) = arg\; max_\theta f(X|\theta)$, and if log is applied to it, it becomes in addition form. Because log is a monotonic increasing function, the point having the maximum value and the point having the original maximum value are the same when log is taken and the addition is easier than multiplication, so in many cases $log - likelihood$ is used to calculate parameter estimation.

2.4 3D Human Pose Estimation

Recently, 3D pose estimation has been approached through CNN. Estimate the human pose of a photo[1, 2, 3] or video[4] through CNN. In general, the estimation of a person's point by looking at the image was limited to the 2D image.[7, 8, 9] However, as the single-view human-pose reconstruction method [10,11] advances, the area of research is rapidly expanding to 3D human-pose. There are many different methods of estimating 3D human pose [2, 3, 4, 5].

Collecting data containing 3D human point information is difficult to collect unless there is adequate equipment, so general 3D human pose dataset is a data set filmed in a studio environment.[5, 6] However, studio environments can cause problems with lighting changes or backgrounds. These problems have recently been solved by creating synthetic datasets.[12, 13] Synthetic dataset has the advantage of knowing the exact location of the point and having many adjustable parts. In some studies, after the creation of synthetic data, it went through the domain adaptation process and learned to enhance the performance of the 3D-Pose Estimation model.

3. Research Process

Before using the general image aggregation method [14, 15] or the synthetic method [12, 13], we need to find rareness to find out what data we need to create additional learning data. In this section we use FB8 to get the probability of each joint vector, and multiply the probabilities to get the rareness of pose.

3.1 FB8

The probability density function(PDF) of FB8 distribution is as follows.

$$f_8(\overrightarrow{x}) = c_8(k,\beta,\eta,\vec{v})^{-1} \exp\{k\vec{v}\cdot\vec{x} + \beta(x_2^2 - \eta x_3^2)\}$$

$$\vec{x} = (cos\theta, sin\theta, cos\emptyset, sin\theta sin\emptyset)$$

It is hard to calculate the parameter $c_8(k,\beta,\eta,\vec{v})$. Even though there are several ways such as Runge-Kutta method, numerical integration and saddle point, the cost is high.

$$f_8(\vec{x}) = \frac{\exp\{k(v_1 cos\theta + v_2 sin\theta cos\emptyset + v_3 sin\theta sin\emptyset) + \beta sin^2\theta(cos^2\emptyset - \eta sin^2\emptyset)\}}{c_8(k,\beta,\eta,\vec{v})}$$

$$c_8(k,\beta,\eta,\vec{v}) = \int_0^{\pi}\int_0^{2\pi} e^{k(v_1 cos\theta + v_2 sin\theta cos\phi + v_3 sin\theta sin\phi) + \beta sin^2\theta(cos^2\phi - \eta sin^2\phi)} d\phi sin\theta d\theta$$

$$\equiv \int_0^{\pi}\int_0^{2\pi} I d\phi d\theta$$

After expanding I by Taylor,

$$I = e^{kv_1 cos\theta} \sum_{l,k,j=0}^{\infty} \left\{ \frac{(kv_2 sin\theta cos\phi)^l}{l!} \frac{(kv_3 sin\theta cos\phi)^k}{k!} \frac{[\beta sin^2\theta(cos^2\phi - \eta sin^2\phi)]^j}{j!} \right\}$$

$$= e^{kv_1 cos\theta} \sum_{l,k,j=0}^{\infty} \sum_{i=0}^{j} \left\{ \frac{k^{l+k}\beta^j v_2^l v_3^k (-\eta)^i}{l!\,k!\,j!} \binom{j}{i} sin^{2j+l+k+1}\theta cos^{2(j-i)+l}\phi sin^{2i+k}\phi \right\}$$

Then, we can obtain

$$c_8(k,\beta,\eta,\vec{v}) = 2\int_0^{\pi}\sum_{l,k,j=0}^{\infty}\left\{\frac{k^{2(l+k)}\beta^j v_2^{2l}v_3^{2k}}{(2l)!\,(2k)!\,j!}e^{kv_1\cos\theta}\sin^{2(j+l+k)+1}\theta \times\sum_{i=0}^{j}(-\eta)^i\binom{j}{i}B\left(j-i+l+\frac{1}{2},i+k+\frac{1}{2}\right)\right\}d\theta$$

$$= 2\sqrt{\pi}\sum_{l,k,j=0}^{\infty}\left\{\frac{k^{2(l+k)}\beta^j v_2^{2l}v_3^{2k}}{(2l)!\,(2k)!\,j!}\left|\frac{kv_1}{2}\right|^{-j-l-k-\frac{1}{2}}I_{j+l+k+\frac{1}{2}}(|kv_1|) \times\sum_{i=0}^{j}(-\eta)^i\binom{j}{i}\Gamma\left(j-i+l+\frac{1}{2}\right)\Gamma\left(i+k+\frac{1}{2}\right)\right\}$$

$$= 2\sqrt{\pi}\sum_{l,k,j=0}^{\infty}\left\{\frac{k^{2(l+k)}\beta^j v_2^{2l}v_3^{2k}}{(2l)!\,(2k)!\,j!}\frac{\Gamma(k+\frac{1}{2})\Gamma(j+l+\frac{1}{2})}{\Gamma(j+l+k+3/2)} \times 0\_F_1\left(;j+l+k+\frac{3}{2};\frac{k^2v_1^2}{4}\right)\,2\_F_1\left(-j,k+\frac{1}{2};\frac{1}{2}-j-l;-\eta\right)\right\}$$

(γ : gamma function, β :beta function, I_v : modified Bessel function.) we can obtain approximation by setting j, k, l properly.

Each parameter was calculated by MLE.

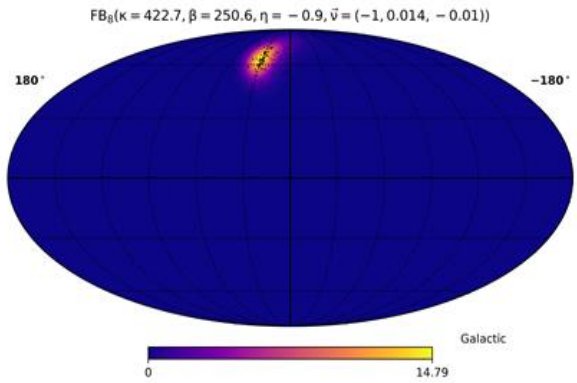

Figure 1

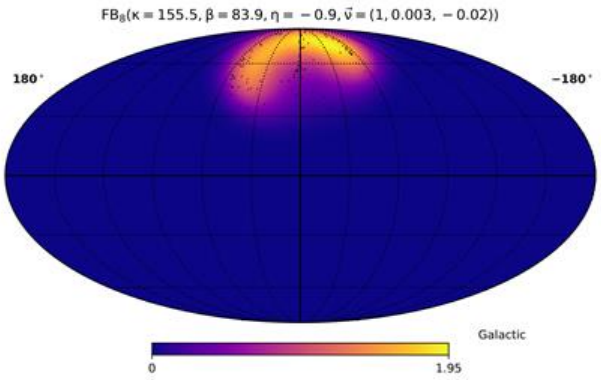

Figure 2

The wider operating range of the joints used in Figure 2 compared to Figure 1 indicates that the FB8 distribution of Figure 2 is more widespread.

3.2 Calculate Rareness

The human pose's joint(j_k) excluding the root node has a parent joint pair for one joint. By subtracting child joint($j_{k,des}$) from parent joint($j_{k,par}$), a single joint vector(v_k) can be created, and the distribution of joint vectors in all datasets can be expressed through FB8. The distribution of joint vector($f_{8,k}(v_k)$) is estimated through MLE. This distribution gives the probability of one joint occurring(p_k). the value of pdf of FB8 is extreme. So, log pdf is used.

$$v_k = j_{k,par} - j_{k,des}$$

$$p_k = log(f_{8,k}(v_k))$$

The probability of one pose is obtained by multiplying the probability of each joint. For log pdf values of low probability joints to not positively affect the overall probability values, this log pdf value must be between 0 to 1 or just 1. And in order for a highly probability joint to have a positive effect on the overall probability value, its probability value must be greater than 1.

The log pdf of each joint vector obtained through FB8 above starts with a negative range, and the minimum multiplied by -1 becomes the maximum value of pdf. So if one joint vector is less likely to come out because it is uncommon, the overall probability is multiplied by 1 to prevent it from affecting the probability, and if the common joint vector is multiplied by a number higher than 1, the probability of a pose have high value. To do this, add the maximum probability of log pdf of the results of FB8 (because the maximum probability is a negative number of the minimum probability), and add 1. When multiplying the probabilities of all joint vectors of a pose, the more common the pose, the higher the probability value, and the less common pose, the closer the probability to 1.

$$p'_k = p_k + max(p_k) + 1$$

Obtain adjusted probability values for all joints. The final rareness value is obtained by multiplying all of these values.

$$R = \prod_{k=1}^{n} p'_k$$

3.3 Visualization

Through the above method, We got the rareness values and visualized the results.

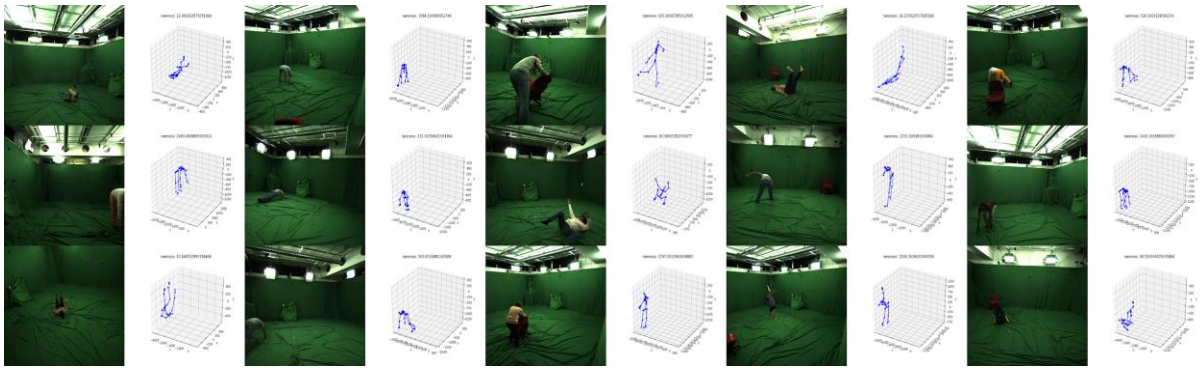

Figure 3. poses in which the head joint is rare

The results of the head joint, where the distribution of FB8 is heavily clustered, are shown in Figure 3. From the above results, it can be seen that data is collected that is not common in the direction of the neck. In the case of the neck, the area of activity is not wide, so the distribution of the total rareness value range is wide. It can be said that neck vectors do not significantly affect rareness.

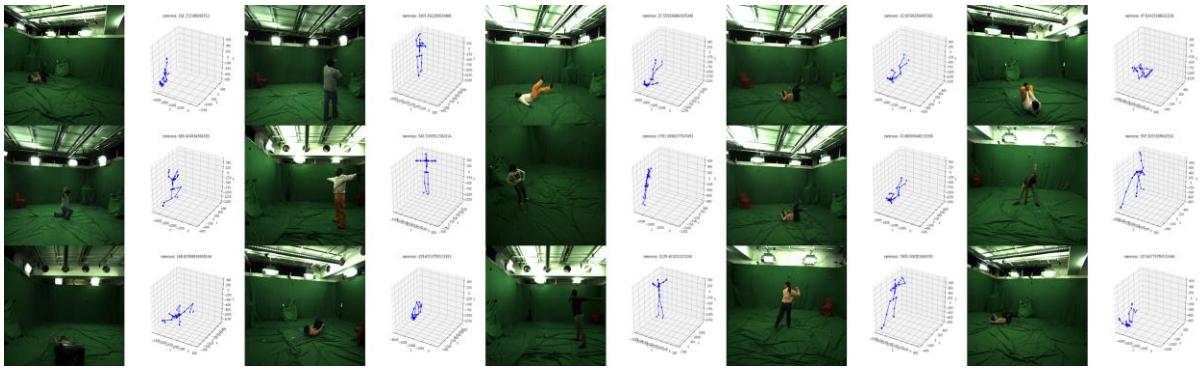

Figure 4. Pose in which the left elbow joint is rare

The results of the left elbow joint, where the distribution of FB8 is widespread, are shown in Figure 4. From the above results, you can see that the direction of the left elbow is not a common posture. There are also various rareness values in the results of rare left elbow vectors, but you can see that they have generally smaller values than the rare head vectors.

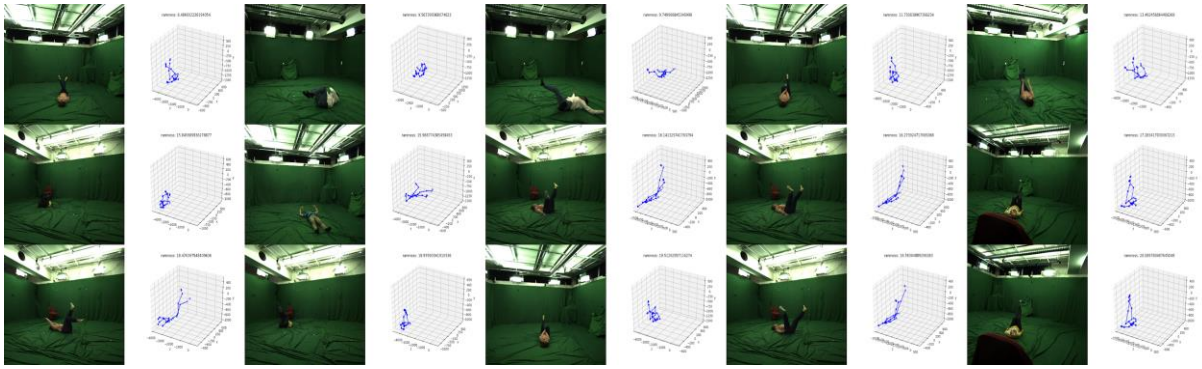

Figure 5. Poses with small rareness values.

Finally, the results of the poses with small rareness values are shown in Figure 5. Overall, it is unusual to lie down and raise your legs up and look up. This pose is generally not done, and appears less in the data set.

4. Conclusion

By observing 16 FB8 distributions for each joint vectors, we could see that each distribution was affected by characteristics of joint. For example, in the case of head joints and neck joints, the distribution was concentrated because of its low diversity of pose. However, distributions about arms are widely distributed due to wide range of movements of arms. In addition, we could sort out Rare pose by using Rareness which is already calculated.

After augmentation rare pose, Domain Adaptation is expected to enable 3D Pose Optimization with improved performance. Some examples below can be found below.

(1) Health Care Application - In the aftermath of the COVID-19, more activities are being done at home than outdoors. In particular, LG U+ and Kakao VX recently collaborated to introduce an app called 'Smart Home Training'. If the user follows the motion with the screen in front of him, the AI recognizes the reflection on the camera and adjusts the motion. At this time, 3D Human Pose Estimation plays an important role.

(2) Rehabilitation treatment - Applications for 'treatment purposes' such as rehabilitation can be developed. It will be useful to develop an application that allows senior citizens to comfortably rehabilitate at home without going to a hospital for rehabilitation treatment.

If you use the 3D Pose Estimation Model with improved performance, it will be useful when creating applications like these. It is also expected that the methods used in this study can be expanded not only in 3D Human Pose but also in 2D Human Pose, Human Finger, etc.

5. Reference

[1] Li, Sijin, Weichen Zhang, and Antoni B. Chan. "Maximum-margin structured learning with deep networks for 3d human pose estimation." Proceedings of the IEEE international conference on computer vision. 2015.

[2] Mehta, Dushyant, et al. "Vnect: Real-time 3d human pose estimation with a single rgb camera." ACM Transactions on Graphics (TOG) 36.4 (2017): 1-14.

[3] Nibali, Aiden, et al. "3d human pose estimation with 2d marginal heatmaps." 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2019.

[4] Zhou, Xiaowei, et al. "Sparseness meets deepness: 3d human pose estimation from monocular video." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[5] Mehta, Dushyant, et al. "Monocular 3d human pose estimation in the wild using improved cnn supervision." 2017 international conference on 3D vision (3DV). IEEE, 2017.

[6] Ionescu, Catalin, et al. "Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments." IEEE transactions on pattern analysis and machine intelligence 36.7 (2013): 1325-1339.

[7] Chen, Xianjie, and Alan L. Yuille. "Articulated pose estimation by a graphical model with image dependent pairwise relations." Advances in neural information processing systems. 2014.

[8] Newell, Alejandro, Kaiyu Yang, and Jia Deng. "Stacked hourglass networks for human pose estimation." European conference on computer vision. Springer, Cham, 2016.

[9] Toshev, Alexander, and Christian Szegedy. "Deeppose: Human pose estimation via deep neural networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.

[10] Mori, Greg, and Jitendra Malik. "Recovering 3d human body configurations using shape contexts." IEEE Transactions on Pattern Analysis and Machine Intelligence 28.7 (2006): 1052-1062.

[11] Ramakrishna, Varun, Takeo Kanade, and Yaser Sheikh. "Reconstructing 3d human pose from 2d image landmarks." European conference on computer vision. Springer, Berlin, Heidelberg, 2012.

[12] Chen, Wenzheng, et al. "Synthesizing training images for boosting human 3d pose estimation." 2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016.

[13] Varol, Gul, et al. "Learning from synthetic humans." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.

[14] Frid-Adar, Maayan, et al. "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification." Neurocomputing 321 (2018): 321-331.

[15] Perez, Luis, and Jason Wang. "The effectiveness of data augmentation in image classification using deep learning." arXiv preprint arXiv:1712.04621 (2017).

[16] Tianlu Yuan. “The 8-Parameter Fisher-Bingham Distribution on the sphere.”(2019)