

Towards a Fair Evaluation of 3D Human Pose Estimation Algorithms

Technical Report #001 - October 2009

C. Canton-Ferrer, J.R. Casas, M. Pardàs, E. Monte
Technical University of Catalonia, Barcelona, Spain

Abstract

Tracking of unrestricted human movement has received great attention by the computer vision community basically fostered by the number of applications that benefit from it. Despite this research focus, there are few established mechanisms for evaluating and comparing the performance of reported solutions. Existing metrics to quantify the performance of a given video-based 3D pose estimation algorithm assume that the committed errors follow a Gaussian distribution and this might yield to misleading and biased figures under certain circumstances. In order to conduct a fair evaluation of such algorithms, a set of new metrics are introduced towards providing a more accurate and realistic measure based on the study of the error distribution. Two markerless and one marker-based methods for 3D human motion capture are employed to analyze data contained in the standard HumanEva-I human motion analysis dataset. Ground truth information provided in this dataset allows computing already existing metrics and comparing them with the proposed ones proving their usefulness.

1. Introduction

Systematic evaluation of computer vision algorithms has raised a growing interest in recent times. Periodic evaluation campaigns allow fair comparison of different techniques, avoiding subjectivity through an agreed set of well-defined metrics for assessment and a reference corpus of pertinent data for testing. Along the same lines, noteworthy examples can be found in the field of face recognition [1], person tracking [2], articulated body motion [14] or gait recognition [13] among others.

In the field of articulated body motion, there is still no general agreement on a principled evaluation procedure using a common set of objective and intuitive metrics for measuring the performance of different articulated motion tracking algorithms. Due to this lack of metrics, some researchers present their tracking systems based on a qualita-

tive assessment [9]. On the other hand, a multitude of isolated measures were defined in individual contributions to validate their systems using various features and algorithms [4, 11, 15]. Recently, a significant contribution by Sigal and Black [14] released a large annotated dataset and proposed two metrics that have been adopted in several evaluation campaigns. Nevertheless, these metrics present some inconveniences and may produce biased scores under certain conditions.

In this paper, we analyze the existing human pose estimation evaluation metrics and propose an alternative set of measures that avoid the inaccuracies derived from the assumption of a Gaussian distribution of the error. Three state-of-the-art multi-camera systems are selected to be evaluated using the HumanEva-I dataset to prove the usefulness of the proposed evaluation methodology. All systems are based on the seminal annealed particle filtering (APF) principle [9] that has been found to produce a robust tracking outputs. Two of the evaluated systems are markerless and employ a 3D reconstruction of the scene together with a human body model (HBM) [6] or a scalable human body model (SHBM) [7]. The other system to be compared is marker-based [5], employing a set of color markers to retrieve the position of the joints of the performer. Finally, a comparison of all metrics is presented proving the adequateness of the proposed ones towards a fair evaluation of video-based 3D human pose algorithms.

2. Error measures

To quantitatively evaluate video-based human pose tracking, two elements are required: ground truth information and a set of metrics. The first have recently benefited from marker-based motion capture systems [4, 14] while sometimes hand-labeling labour has been also employed [11]. In order to define objective and informative performance evaluation metrics, two design criteria should be followed. First, they should allow to judge the tracker's precision in determining the exact location of the articulated structure landmarks. Second, they should reflect its ability

to consistently track the landmark locations through time, i.e., to correctly trace their trajectories. Finally, useful metrics should have as few adjustable thresholds as possible to help make evaluations straightforward and keep results comparable.

A number of metrics have been employed independently by several authors, most of them relying on the mean square distance, in either 2D or 3D, between the estimated positions of several landmarks on the human body and ground truth positions [11]. Other metrics rely on the angular root mean square error measured at each joint angle [15]. Quantitative evaluation of human pose estimation algorithms was addressed by Bălan *et al.* [4] presenting some performance scores for those based on particle filtering. HumanEva dataset [14] provided two generic metrics together with an annotated set of sequences involving several types of motion. This dataset has been largely adopted by the computer vision community as the reference tool for human pose performance assessment.

2.1. Problem formulation

Given a HBM whose pose is represented by a state vector $\mathbf{y} \in \mathcal{X} \subset \mathbb{R}^D$, we may represent any adopted pose by a set of M virtual markers encoded as a vector $X = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_M\}$, where $\mathbf{p}_m \in \mathbb{R}^3$. A mapping from \mathbf{y} to X can be always derived, either if the state vector encodes landmark positions (using a linear mapping) or joint angles (applying forward kinematics). Let us define the error between an estimated pose \hat{X} with reference to the ground truth pose X as:

$$D(X, \hat{X}) = \frac{1}{M} \sum_{m=1}^M \|\mathbf{p}_m - \hat{\mathbf{p}}_m\|. \quad (1)$$

This error figure is usually assumed to have a Gaussian distribution and the first and second statistical moments can be derived. Hence, when a sequence of poses of length T is analyzed, the performance of the tracking algorithm may be assessed by averaging error along time and computing the standard deviation:

$$\mu = \frac{1}{T} \sum_{t=1}^T D(X_t, \hat{X}_t), \quad (2)$$

$$\sigma = \sqrt{\frac{1}{T} \sum_{t=1}^T \left(D(X_t, \hat{X}_t) - \mu \right)^2}. \quad (3)$$

These two metrics have been proposed in the HumanEva framework by [14] to assess the performance of a generic pose estimation algorithm. Good performance of a HBM tracking algorithm will yield low values of both μ and σ , whereas high values will denote a poor efficiency.

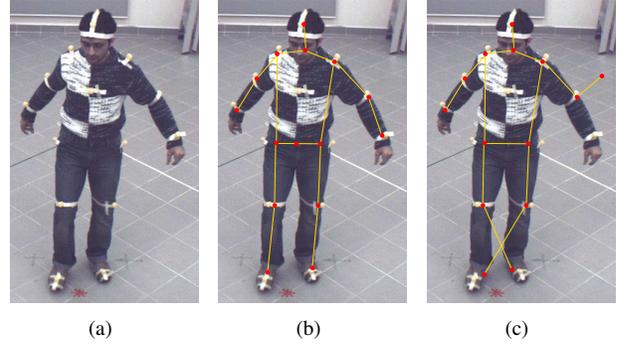


Figure 1. Point based metrics comparison example. In (a), the reference image with the visual yellow markers. In (b), a good body pose estimation produces $\mu = 48.91$ and $\sigma = 16.21$, and $MMTP = 48.91$ and $MMTA = 1.0$. In (c), a poor body pose estimation produces $\mu = 51.22$ and $\sigma = 24.37$, and $MMTP = 46.35$ and $MMTA = 0.77$ with $\delta = 50$ (all distance units in mm). $MMTA$ and $MMTP$ stand for the metrics presented in this paper.

These metrics produce meaningful results at a given time instant when the sets X and \hat{X} fulfill the condition:

$$\|\mathbf{p}_m - \hat{\mathbf{p}}_m\| \leq \delta, \quad \forall m, \quad (4)$$

being δ a fixed threshold. Parameter δ discriminates whether the position \mathbf{p}_m and the estimation $\hat{\mathbf{p}}_m$ can be considered as matched. This is the case of a pose configuration \hat{X} similar to the one depicted in Fig.1(b) where the estimation of the landmark positions are close to the ground truth positions. When this condition is not fulfilled for some values of m , then the algorithm outputs estimate poses \hat{X} resembling Fig.1(c). A limitation of the metrics proposed in Eq.2 and 3 is that the non matched case is not distinguished from a matched case. When the estimation of a certain landmark is clearly far away from the ground truth (that is when $\|\mathbf{p}_m - \hat{\mathbf{p}}_m\| > \delta$) the produced error is still accounted as a gross estimation inaccuracy thus severely penalizing both μ and σ scores. Hence, when a landmark subset is not tracked properly (typically, the end of the limbs), the figures produced by these metrics are not informative enough to describe the tracker's performance.

2.2. Statistics

In order to test the Gaussianity distribution assumption stated in Eqs.2 and 3, the error vector E is generated, $E = \{\mathbf{e}_k\} = \{\|\mathbf{p}_{m,t} - \hat{\mathbf{p}}_{m,t}\|\}, \forall m, t$, including the estimation error associated to every body marker along the whole analysis sequence. When analyzing the histogram of E for several human motion capture algorithms shown in Fig.2, it can be seen that there is a dominant peak with a Gaussian shape associated to error values fulfilling $\mathbf{e}_k \leq \delta$, while the long tail spreading to large error values is derived from those satisfying $\mathbf{e}_k > \delta$. Indeed, when analyzing the

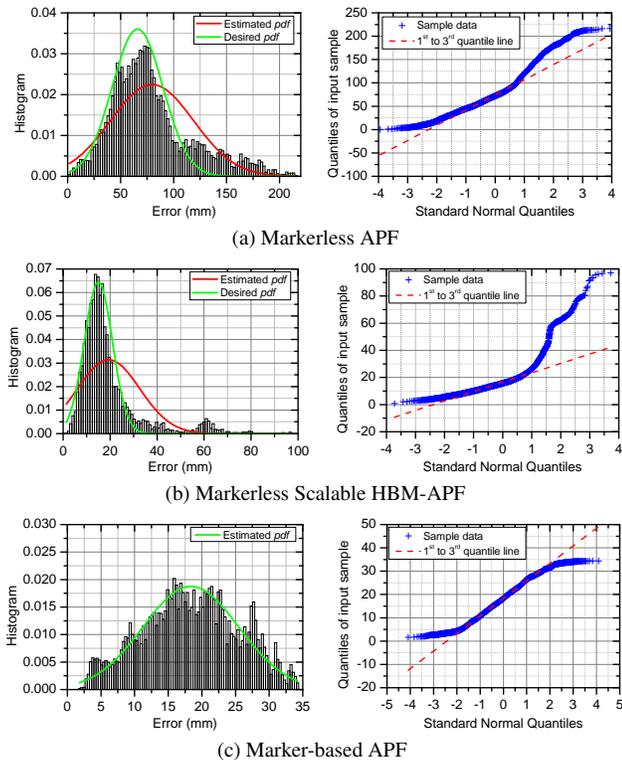


Figure 2. Histograms associated to the estimation error and the quantile-quantile plot between the error vector E and a reference normal distribution.

estimated pdf associated to E , it can be seen how it does not properly fit to a Gaussian distribution (red line). In an ideal case, as in Fig.2c, the estimated pdf matches a Gaussian function while, in the other cases, the desired (green line) and computed pdf differ substantially.

The quantile-quantile plot is an efficient way to assess the Gaussianity of a distribution [10] in the sense that the empirical quantiles of the data are plotted versus the quantiles of a Gaussian. If the data belongs to a Gaussian distribution, the points are spread roughly following a line. If the data is skewed or has longer/shorter tails than a Gaussian, instead of having a line, the scatter plot shows either flat or vertical parts. As it can be seen in Fig.2, both markerless algorithms error data do not properly align with the regression line while the marker-based one does, as expected from the associated histograms.

3. Metrics

Assessing the efficiency of a human pose estimation and tracking algorithm assuming a Gaussian distribution of the involved errors E may lead to biased performance figures (mean and variance). A set of metrics is proposed to circumvent this problem towards a fair evaluation procedure.

3.1. Point-based metrics

A new point based metric is proposed in order to better express the performance of a human motion capture algorithm. It takes into account that there might be situations where a subset of the landmarks in \hat{X} is not estimated properly while the rest is done accurately. A similar problem is found in the field of multiple object tracking where Bernardin *et al.* [3] proposed a set of metrics that were validated and largely accepted as performance and comparison scores in international evaluation campaigns [1]. The underlying concept of this performance metrics may be extended to the field of pose estimation evaluation to produce two intuitive and more informative metrics.

Let us define the set Ω as the set of pairs estimation-ground truth locations whose distance is below the threshold δ , $\Omega = \{(\mathbf{p}_m \in X, \hat{\mathbf{p}}_m \in \hat{X}) \mid \|\mathbf{p}_m - \hat{\mathbf{p}}_m\| \leq \delta\}$. The two metrics can be defined:

1. The Multiple Marker Tracking Precision ($MMTP$),

$$MMTP = \frac{\sum_{t=1}^T \sum_{m \in \Omega_t} \|\mathbf{p}_{t,m} - \hat{\mathbf{p}}_{t,m}\|}{\sum_{t=1}^T |\Omega_t|}, \quad (5)$$

where $|\Omega|$ denoted the cardinality of the set Ω . This metric shows the total position error for the matched estimation-ground truth pairs, averaged by the total number of matches made along time. It reflects the ability of the tracker to estimate precise landmark positions, independent of the performance of the algorithm to correctly match all the landmarks in the HBM.

2. The Multiple Marker Tracking Accuracy ($MMTA$),

$$MMTA = 1 - \frac{\sum_{t=1}^T |\Omega_t|}{M \cdot T}, \quad (6)$$

where M is the total number of landmarks in the HBM. This score accounts for the ability of the of tracker at producing matched estimation-ground truth pairs.

Finally, a supplementary metric might be defined: the standard deviation of the $MMTP$ score, σ_{MMTP} , as a measure of the quality of the estimation of the correctly matched estimation-ground truth pairs.

In the example depicted in Fig.1, these two sets of metrics are compared. When the condition expressed in Eq.4 is fulfilled as in Fig.1b, metrics $\mu = 48.91$ and $\sigma = 16.21$ properly evaluate the estimated pose. In Fig.1c, a typical situation of landmark estimation swapping is found in the ankles while the left hand track is lost. In this case, the implicit assumption that these landmark estimation inaccuracies follow a Gaussian distribution clearly biases the scores, $\mu = 51.22$ and $\sigma = 24.37$. $MMTP$ and $MMTA$ can nicely handle both situations: in the first case $MMTA = 1.0$

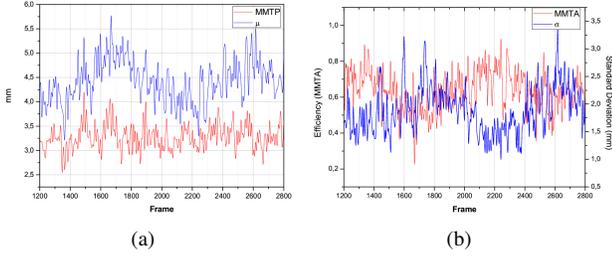


Figure 3. Quantitative performance of point based metrics. In (a), landmark estimation precision scores μ and $MMTP$ along time and, in (b), the plot of the evolution of scores σ and $MMTA$.

indicates that the tracker has correctly produced a valid estimation for all markers and the average precision is $MMTP = 48.91$. In the second case, $MMTA = 0.77$ indicates that the tracker could only track the 77% of the landmarks during the analysis period of time and the average precision of the correctly tracked landmarks was $MMTP = 46.35$ which is not biased by the non matched pairs.

A quantitative comparison of the temporal evolution of the presented point-based metrics is depicted in Fig.3. It is shown that the score μ is more sensitive than $MMTP$ since it agglutinates both information from precision and lost tracks. An instantaneous version of σ and $MMTA$ computed every frame is depicted to show the noticeable correlation between both scores: when there are less matched pairs estimation-ground truth, $MMTA$ figure decreases while the deviation of the error increases and vice versa. However, the value of σ has little physical interpretation when some landmarks are not tracked properly while $MMTA$ presents the amount of correctly tracked landmarks.

It must be noted that these results have been presented for $\mathbf{p}_m \in \mathbb{R}^3$ but these metrics can be adapted to the case where the landmark locations are measured directly on images, that is $\mathbf{p}_m \in \mathbb{R}^2$, or when we deal with 2D HBMs.

Parameter-free evaluation

Selecting an adequate value of δ is crucial to obtain meaningful $MMTP$ and $MMTA$ scores. When selecting small values of δ , the proposed metrics will be very restrictive thus yielding to a low $MMTA$ and high $MMTP$ values. On the other hand, large values of δ will report a tendency to $MMTA \rightarrow 1$ and $MMTP \rightarrow \mu$. Although δ may be set up manually allowing a maximum allowed error, a parameter-free evaluation procedure would be desirable.

The optimal value of δ , δ_{opt} , should be one that partitions the histogram of E in such a way that values fulfilling $e_k \leq \delta_{opt}$ tend to have a Gaussian distribution, as shown in Fig.4a. Therefore, $MMTP$ and σ_{MMTP} will stand for the mean and variance of the green bins approximated by the Gaussian function plotted in blue. In this way, $MMTA$ will

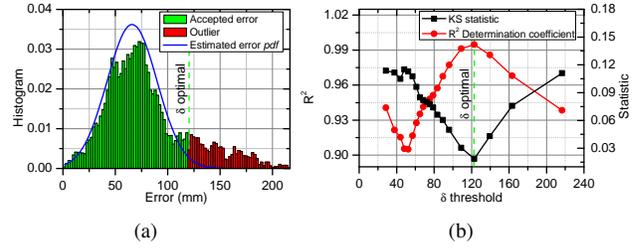


Figure 4. δ selection. In (a), parameter δ_{opt} partitions the error histogram between the Gaussian and outlier parts.

account for the fraction of the error that can not be considered as belonging to this Gaussian distribution.

In order to select the adequate value of δ_{opt} , we formulated the following optimization problem:

$$W(\delta) = \{e_k \in E \mid e_k \leq \delta\}, \quad (7)$$

$$\delta_{opt} = \min_{\delta} f(W(\delta)), \quad (8)$$

where $f(\cdot)$ stands for a normality test function over the values of $W(\delta)$. Two options have been considered for $f(\cdot)$. First, we employed the Kolmogorov-Smirnov statistic [8] that measures the maximum difference between the empirical cumulative distribution function (CDF) of the input data $W(\delta)$ and the theoretical CDF of a Gaussian. This statistic measures a local feature of both CDF's, which is the worst discrepancy. In our optimization problem, we search the value of δ that minimizes the maximum absolute difference between the empirical CDF up to δ and the theoretical CDF of a Gaussian. Second, a linear regression [10] is applied over the quantiles of the input data with reference to the quantiles of a Gaussian. Then, we compute the coefficient of determination R^2 which is related to the explained variance and measures a global feature, i.e. the dispersion around the regression line. Values of R^2 near 1 mean that the data is aligned with the regression line, while low values hint at a lack of linear dependence. When employing the R^2 figure, Eq.8 minimum is replaced by a maximum.

Although the two scores measure different aspects of the problem, we have found that they usually agree on the value of δ_{opt} as depicted in Fig.4b.

3.2. Angle-based metrics

A natural choice in evaluating the performance of an articulated motion capture system would be to produce a score directly related to the defining parameters of the HBM, that is angles. The advantage over point-based metrics lies in the fact that measured angles are relative to the two vertices of the articulation ending at the joint and, therefore, tracking errors do not accumulate towards the end of the limbs, as happens with the spatial position measured by point-based metrics. Encoding the pose of a given HBM \mathcal{H} by a set of

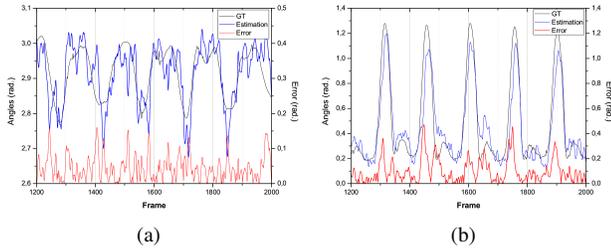


Figure 5. Angular re-parameterization example for the elbow, (a), and the knee (b) articulations executing the action *walking*. Once the angles of the ground truth (black) and the estimation (blue) are expressed with the same HBM parameters, the error (red) between them can be computed and analyzed by the angular metrics.

N angles¹, $\Theta_{\mathcal{H}} = \{\theta_1, \theta_2, \dots, \theta_N\}$, $\theta_n \in \mathbb{R}$, is a common approach in articulated motion tracking, since these magnitudes are directly related with the kinematic structure of the human body [7, 6].

Defining metrics based on an angle representation of the HBM presents some issues to be taken into account. For example, every pose encoding based on angles assumes a parameterization of the human body that can not be the same among algorithms enforcing different degrees of freedom in every joint. Furthermore, joint angle representations are not unique (quaternions, Euler angles, exponential maps, etc.) thus making comparisons among algorithms difficult. Some researchers have already proposed metrics measuring the error in terms of degrees at every joint [15] but due to the aforementioned issues, no angular metric has yet been widely adopted by the community. The authors propose a general method for evaluating the performance of an articulated motion capture system in terms of angles, regardless of the parameterization employed during the analysis.

In order to define an angle based metric, a reference HBM $\tilde{\mathcal{H}}$ representation should be adopted. An obvious choice would be to define a transformation between the HBM \mathcal{H} used by the tracking algorithm and $\tilde{\mathcal{H}}$ but this mapping cannot be always computed due to the differences among HBM parameterizations. Instead, we propose the following re-parameterization technique. First, a given pose $\Theta_{\mathcal{H}}$ is transformed into a set of 3D coordinates, X , by applying forward kinematics [14]. This set X of 3D coordinates is implicitly labeled because it is known which body landmark is described by every 3D location. Finally, the inverse kinematic problem has to be solved by extracting the angles $\Theta_{\tilde{\mathcal{H}}}$ of $\tilde{\mathcal{H}}$ from the set X . We propose using an enough detailed HBM as $\tilde{\mathcal{H}}$; in our case, we chose the one described in [12]. Moreover, this particular choice of $\tilde{\mathcal{H}}$ allows an algebraic expression relating X with all of its joint angles. An example of this process is depicted in Fig.5

¹The body root position and orientation are omitted for the sake of notation simplicity.

where $\mathcal{H} \neq \tilde{\mathcal{H}}$ to prove the described re-parameterization approach.

This process is applied to both the ground truth 3D positions X to obtain the ground truth angles $\Theta_{\tilde{\mathcal{H}}}$ and to the estimated pose $\hat{\Theta}_{\mathcal{H}}$ to derive \hat{X} and, then, to obtain the estimated angles $\hat{\Theta}_{\tilde{\mathcal{H}}}$. The error between an estimated pose $\hat{\Theta}_{\tilde{\mathcal{H}}}$ to the ground truth pose $\Theta_{\tilde{\mathcal{H}}}$ is defined as:

$$D(\Theta_{\tilde{\mathcal{H}}}, \hat{\Theta}_{\tilde{\mathcal{H}}}) = \frac{1}{M} \sum_{n=1}^N |(\theta_n - \hat{\theta}_n) \bmod \pm \pi|. \quad (9)$$

Two angular metrics are proposed: the angular mean estimation error, μ_{θ} , and its associated standard deviation, σ_{θ} . However, computing these scores directly over all angles over a period of length T would generate biased results due to the already discussed Gaussianity assumptions. Therefore, it is proposed to compute these metrics over the angles associated to two vertices fulfilling the matching criterion described in Eq.4 and averaged along all time instants, mimicking Eq.5.

The proposed angular metrics complement the information provided by the point based metrics and can not be presented alone. While the efficiency of the tracking system is assessed by the score *MMTA*, both *MMTP* and σ_{MMTP} and the pair μ_{θ} and σ_{θ} provide information about the precision of the system in the spatial and angular domains respectively.

4. Evaluation and Results

In order to assess the adequateness of the proposed metrics for human pose estimation algorithms, three multi-camera algorithms [7, 6, 5] have been employed to process the data contained in the HumanEva-I dataset. The obtained tracking results have been evaluated using the standard metrics used by the human motion capture community (mean and variance) and the proposed metrics in this paper (for $\delta = 100$ mm and δ_{opt}) as reported in Tab.1.

When comparing the obtained results using μ and σ metrics, it is not trivial to decide which system has a better performance, for instance between the marker-based APF and the markerless SHBM-APF. The marker-based one has high μ and low σ scores while the markerless SHBM-APF shows the opposite. This issue is straightforwardly solved when comparing scores proposed in this paper. Although both algorithms exhibit a similar behavior in the Gaussian region of the error (reported by figures *MMTP* and σ_{MMTP}), the *MMTA* score computed with $\delta = 100$ mm indicates that in the 95% of the cases the estimation is within the acceptance region, showing its superiority in comparison with the 76% of the markerless SHBM-APF algorithm. Values of μ_{θ} and σ_{θ} are usually correlated with the behavior of *MMTP* and σ_{MMTP} but provide a more natural way to express the performance of a given algorithm.

Method	μ	σ	$\delta = 100 \text{ mm}$					δ_{opt}					
			<i>MMTP</i>	σ_{MMTP}	<i>MMTA</i>	μ_{θ}	σ_{θ}	<i>MMTP</i>	σ_{MMTP}	<i>MMTA</i>	μ_{θ}	σ_{θ}	δ_{opt}
Marker-based APF	59.88	17.35	45.85	13.70	95.32	7.09	4.21	41.37	11.63	92.18	6.82	3.91	82
Markerless APF	121.18	45.92	90.17	35.18	71.36	10.12	3.30	93.05	36.41	72.83	10.76	3.71	105
Markerless SHBM-APF	51.34	28.51	45.31	24.52	76.42	6.73	4.97	42.22	22.71	75.02	5.69	4.52	89

Table 1. Performance results comparison.

Although a beforehand agreed δ parameter should be employed in an evaluation campaign to fix the maximum allowed error in the estimation of a pose, its value should be carefully selected not to produce biased values of *MMTP* and σ_{MMTP} due to a wrong Gaussian distribution assumption of set $e_k \leq \delta$. In a more thorough evaluation process, the value of δ_{opt} may give a useful clue to determine the range of correct operation of an algorithm, understood as the error range where limbs can be considered correctly tracked.

The presented results prove how the information contained in μ and σ is not adequate to quantize the efficiency of a human motion capture algorithm and it does not allow a proper comparison among techniques. Instead, given a fixed δ , the *MMTA* score should be the first figure to assess the performance of a set of algorithms, followed by the values of *MMTP* and σ_{MMTP} (or μ_{θ} and σ_{θ}). Finally, δ_{opt} would provide a supplementary criteria for ranking. In our case, despite the diversity of employed techniques, we would first rank the marker-based algorithm since it is the one yielding to the highest value of *MMTA*. Then, we would compare the *MMTP* values of both markerless systems ranking the markerless SHBM-APF in second place and, finally, the markerless APF.

5. Conclusions

Defining a set of metrics to evaluate video-based human motion capture algorithms is an important issue towards allowing a fair comparison among techniques. It has been proved that the usually employed scores based on the assumption that the committed errors follow a Gaussian distribution may produce biased figures. The analysis of the histogram of the errors reveals that these values follows a Gaussian distribution up to a cutoff value δ_{opt} , and an unpredictable distribution shape afterward. Two point-based metrics are introduced to assess the performance of a given human motion capture algorithm as the fraction of the estimated tracked landmarks (i.e. joints) that are wrongly estimated, *MMTA*, and the mean and variance of the estimations that are in the Gaussian region of the error, *MMTP* and σ_{MMTP} . Mean and variance are also computed for the angles of correctly tracked vertices, namely μ_{θ} and σ_{θ} , through a re-parameterization approach to harmonize measures obtained when using different HBMs. As a future work, we expect to disseminate the conclusions of this study to be included in future evaluation campaigns.

References

- [1] CLEAR - Classification of Events, Activities and Relationships Evaluation and Workshop. <http://www.clear-evaluation.org>, 2007. 1, 3
- [2] PETS - Performance Evaluation of Tracking and Surveillance. <http://pets2007.net>, 2007. 1
- [3] K. Bernardin, A. Elbs, and R. Stiefelhagen. Multiple object tracking performance metrics and evaluation in a smart room environment. In *Proc. IEEE Intl. Workshop on Visual Surveillance*, 2006. 3
- [4] A. Bălan, L. Sigal, and M. Black. A quantitative evaluation of video-based 3D person tracking. In *Proc. 2nd Joint IEEE Intl. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pages 349–356, 2005. 1, 2
- [5] M. P. C. Canton-Ferrer, J.R. Casas. Towards a low cost multi-camera marker based human motion capture system. In *Proc. IEEE Int. Conf. on Image Processing*, 2009. 1, 5
- [6] M. P. C. Canton-Ferrer, J.R. Casas. Voxel based annealed particle filtering for markerless 3D articulated motion capture. In *Proc. IEEE Conf. on 3DTV*, 2009. 1, 5
- [7] C. Canton-Ferrer, J. Casas, and M. Pardàs. Exploiting structural hierarchy in articulated objects towards robust motion capture. In *Proc. Conf. on Articulated Motion and Deformable Objects*, volume 5098 of *Lecture Notes on Computer Science*, pages 82–91, 2008. 1, 5
- [8] W. Conover. *Practical Nonparametric Statistics*. 1980. 4
- [9] J. Deutscher and I. Reid. Articulated body motion capture by stochastic search. *Intl. Journal of Computer Vision*, 61(2):185–205, 2005. 1
- [10] R. Johnson and D. Wichern. *Applied multivariate statistical analysis*. 2007. 3, 4
- [11] M. Lee and R. Nevatia. Human pose tracking using multi-level structured models. In *Proc. European Conf. on Computer Vision*, volume 3, pages 368–381, 2006. 1, 2
- [12] I. Mikič. *Human body model acquisition and tracking using multi-camera voxel data*. PhD thesis, University of California, San Diego, 2003. 5
- [13] S. Sarkar, P. Phillips, Z. Liu, I. Vega, P. Grother, and K. Bowyer. The HumanID gait challenge problem: data sets, performance, and analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(2):162–177, 2005. 1
- [14] L. Sigal and M. Black. HumanEva: Synchronized video and motion capture dataset for evaluation of articulated human motion. Technical Report CS-06-08, Department of Computer Science, Brown University, 2006. 1, 2, 5
- [15] C. Sminchisescu, K. Kanaujia, Z. Li, and D. Metaxas. Discriminative density propagation for 3D human motion estimation. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 390–397, 2005. 1, 2, 5