Line-Symmetry-Aware k-Means for Skeleton-Based Human Posture Clustering with Reduced Redundancy

Hiroto Miyabe[1,2] and Masashi Nishiyama[1] [a]

[1] *Graduate School of Engineering, Tottori University, 101 Minami 4-chome, Koyama-cho, Tottori, 680-8550, Japan*

[2] *Corporate Research & Development Center, Toshiba Corporation, 1 Komukai Toshiba-cho, Saiwai-ku, Kawasaki, Kanagawa 212-8582, Japan*

hiroto.miyabe.t75@mail.toshiba, nishiyama@tottori-u.ac.jp

Keywords: Human Posture Clustering, Skeleton-Based Human Pose, Line Symmetry, Symmetry-Aware Clustering, K-Means Clustering.

Abstract: Investigating representative 2D postures from large collections of human images requires clustering with reduced redundancy when horizontal reflections are treated as equivalent. However, standard clustering methods can produce mirrored, near-duplicate cluster representatives; with a fixed number of clusters, these redundant representatives leave fewer centroids available for distinct posture patterns. We present Line-Symmetry-Aware k-means (LSAK), a coordinate-based specialization of transformation-invariant clustering. During assignment, LSAK selects between the original and reflected forms of each centered skeleton. It then updates centroids by averaging the aligned samples. Reflection reverses horizontal coordinates and permutes anatomical left-right keypoints. Our focus is redundancy among the resulting representative postures. Evaluation on DeepFashion-MultiModal shows that LSAK substantially reduces line-symmetrical centroid pairs compared with the evaluated baselines, and this trend is further examined under sample duplication, mirror augmentation, and multiple pair-detection thresholds. Qualitative results illustrate how the reduced redundancy changes the set of representative postures obtained for a fixed number of clusters.

1 INTRODUCTION

Advances in human pose estimation (Xu et al., 2022; Yang et al., 2023) facilitate reliable extraction of skeleton samples from human images, enabling large-scale analysis of diverse postures. Potential applications include studying how the postures of fashion models influence human impressions and product appeal. Such investigations require clustering similar postures and extracting a small number of representative postures for inspection without unnecessary comparisons. We focus on representative posture extraction from 2D skeletons, treating horizontal reflections as equivalent.

Clustering is commonly used to extract a small number of representative postures. Widely used methods include k-means (MacQueen, 1967), k-means++ (Arthur and Vassilvitskii, 2007), and Faster k-medoids (Schubert and Rousseeuw, 2019). Given N skeleton samples, these methods consolidate the samples into K clusters specified in advance $(K \ll N)$ based on a standard distance metric. The k-means variants represent each cluster by its mean, whereas k-medoids selects an observed sample as its representative. These K cluster representatives are treated as representative postures for inspection.

However, when existing methods (MacQueen, 1967; Arthur and Vassilvitskii, 2007; Schubert and Rousseeuw, 2019) are applied to clustering skeleton samples, the following issues arise.

I_1 : **Inter-cluster redundancy**. Postures that are highly similar after horizontal reflection in the normalized image-plane coordinate space appear multiple times among the representative postures. This occurs because centroids that are near-duplicates under horizontal reflection are generated redundantly in the clustering results.

I_2 : **Intra-cluster variation**. For a fixed number of clusters, redundant representatives can leave fewer centroids available for distinct posture patterns. Different patterns may then share a cluster, producing an intermediate centroid and large deviations from individual skeleton samples.

Among these, Issue I_1 is the primary target of this

[a] https://orcid.org/0000-0002-5964-3209

work. We treat Issue I_2 as a secondary property of the resulting clusters rather than as an equally weighted optimization objective.

Symmetry-aware clustering methods use point-symmetry-based distances (Su and Chou, 2001; Chung and Lin, 2007) or line symmetry in two-dimensional point sets (Chung and Lin, 2006). These methods model symmetry of point distributions rather than reflection of an entire skeleton. More broadly, transformation-invariant clustering jointly infers cluster identities and latent transformations (Frey and Jojic, 2003). That study does not evaluate skeleton-based posture clustering. Reflection-invariant pose distances have also been used for artwork retrieval (Jenicek and Chum, 2019). This retrieval method matches observed poses but does not address inter-cluster redundancy among representative postures for a fixed number of clusters.

In this paper, we focus primarily on reducing inter-cluster redundancy in representative-posture extraction with a fixed number of clusters. Building on established transformation-invariant clustering principles, we formulate Line-Symmetry-Aware k-means (LSAK) in the centered 2D skeleton-coordinate space. We assess reflection-related redundancy in the resulting representative set using Pairs. Reducing redundant representatives can leave more centroids available for distinct posture patterns; we therefore examine within-cluster compactness as a complementary property. We summarize the contributions of our method as follows.

- **Reflection-aware estimation of representative postures.** LSAK selects the original or reflected form of each skeleton according to its distance to the current centroids. It aligns the selected forms within each cluster and averages their coordinates to update the representative posture. Reflection reverses horizontal coordinates and exchanges anatomical left-right keypoints. This allows reflection-related samples to share a representative rather than occupy separate clusters.
- **Reduction of reflection-based centroid redundancy.** Our primary objective is to suppress line-symmetrical near-duplicate centroids. We quantify this effect using Pairs, which counts mutually matched line-symmetrical centroid pairs. Experiments show that LSAK substantially reduces Pairs compared with the evaluated baselines. Additional experiments with sample duplication, mirror augmentation, and different pair-detection thresholds further examine the robustness of this effect.
- **Compactness and qualitative analysis of representative postures.** We evaluate within-cluster compactness and qualitatively inspect the extracted representative postures. These analyses characterize the working representations and illustrate the posture patterns covered by a fixed number of representatives.

2 PROBLEM SETTING AND DESIGN CONCEPT

We consider N standardized 2D skeleton samples and seek K representative postures, where K is fixed in advance. For representative-posture inspection, horizontally reflected skeletons are treated as equivalent. Under this setting, the central problem is whether reflection-related postures unnecessarily occupy separate representatives, thereby reducing the number of centroids available for genuinely different posture patterns. We first use Figure 1 to make the two resulting issues concrete, and then explain the design concept used to address the primary redundancy problem.

Figure 1(a) schematically illustrates the inter-cluster redundancy introduced in Section 1, together with the associated intra-cluster variation. We first examine Issue I_1, inter-cluster redundancy. Although the centroid of each cluster initially appears to represent a different posture, horizontally reflecting one in the normalized image-plane coordinate space makes the postures nearly coincide. This high similarity under horizontal reflection indicates that ignoring symmetry can produce redundant centroids. Next, we describe Issue I_2, intra-cluster variation. Each cluster in Figure 1(a) contains both postures with one leg raised and postures without a raised leg. The centroid therefore represents an intermediate posture, differing from the individual skeleton samples and leading to large intra-cluster variation.

Here, we consider how to avoid allocating multiple centroids to postures that are highly similar under horizontal reflection. We focus on horizontal reflection relationships among normalized skeleton samples and primarily target the inter-cluster redundancy in Issue I_1. Figure 1(b) schematically illustrates the intended effect of our method. For each skeleton sample, the method considers its original and reflected representations during clustering so that reflection-related postures need not consume separate representatives. With a fixed number of clusters, the resulting centroids can therefore cover a broader set of posture patterns. We additionally examine within-cluster compactness to characterize the resulting clusters.

The design estimates representative postures directly in the skeleton-coordinate space. Each centroid is a coordinate-wise mean of the selected, orientation-

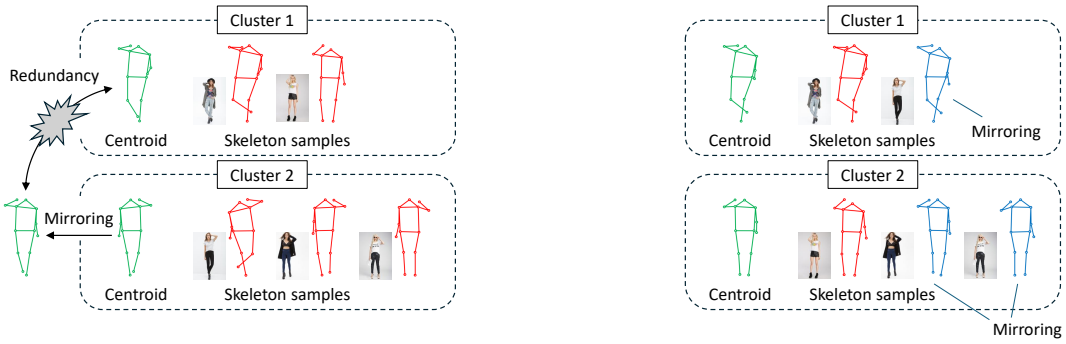


Figure 1: Schematic comparison of conventional clustering and LSAK. (a) Ignoring horizontal reflection can produce redundant centroids and clusters containing distinct posture patterns. (b) LSAK is designed primarily to suppress mirror-related redundant centroids; its reflection-aware representation can also affect within-cluster compactness.

aligned samples; restoring the dataset mean gives keypoint coordinates that can be displayed directly, without reconstructing a pose from a separate descriptor. Only the predefined reflection is treated as equivalent in the centered space; arbitrary in-plane rotations are not identified.

A possible alternative is to pre-canonicalize each skeleton to a fixed orientation before applying conventional clustering. Such a rule selects orientation independently of the current clusters, whereas LSAK selects between the two representations relative to the current centroids. This distinguishes the procedures but does not establish superiority over pre-canonicalization without a direct comparison.

3 LSAK: LINE-SYMMETRY-AWARE K-MEANS

3.1 Line-Symmetry-Aware Distance Function

We use a reflection-aware distance on centered skeleton-coordinate vectors. Let the per-sample standardized skeleton vector be denoted by $\boldsymbol{x}_n \in \mathbb{R}^{2z}$ ($n \in \{1,\ldots,N\}$), where

$$\boldsymbol{x}_n = (x_{n,1}, y_{n,1}, \ldots, x_{n,z}, y_{n,z})^{\mathrm{T}}. \tag{1}$$

Each vector concatenates the 2D coordinates of z keypoints. For each skeleton, we standardize the horizontal and vertical coordinate components separately across the z keypoints so that each component has zero mean and unit variance. This per-sample normalization suppresses translation and scale differences in the image plane before reflection relationships are considered.

As an implementation detail, we further center the standardized vectors by their feature-wise dataset mean before iterative clustering. This common translation leaves ordinary Euclidean k-means assignments unchanged. We define

$$\boldsymbol{\mu} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n, \qquad \bar{\boldsymbol{x}}_n = \boldsymbol{x}_n - \boldsymbol{\mu}. \tag{2}$$

Hereafter, $\boldsymbol{x}_n$ denotes the standardized skeleton itself, whereas $\bar{\boldsymbol{x}}_n$ denotes the centered working representation used during iterative clustering. Keeping these two quantities distinct is useful because the dataset mean is added back only when the final centroids are returned for evaluation and visualization.

We define horizontal reflection by

$$R = P \otimes H, \qquad H = \mathrm{diag}(-1, 1), \tag{3}$$

where P is the $z \times z$ permutation matrix that exchanges anatomical left-right keypoint pairs while leaving the central keypoint unchanged, and $\otimes$ denotes the Kronecker product. The permutation follows the keypoint order of the input skeletons. The line-symmetrical counterpart of the centered working representation is

$$\hat{\boldsymbol{x}}_n = l(\bar{\boldsymbol{x}}_n) = R\bar{\boldsymbol{x}}_n. \tag{4}$$

In the implementation, the reflected standardized samples $R\boldsymbol{x}_n$ are centered by their own feature-wise mean. Because their mean is $R\boldsymbol{\mu}$, this gives $R\boldsymbol{x}_n - R\boldsymbol{\mu} = R(\boldsymbol{x}_n - \boldsymbol{\mu}) = R\bar{\boldsymbol{x}}_n$. Thus, centering preserves the reflection relation exactly in the working space even though the original and reflected arrays are centered separately in the implementation. When mapped back to the standardized coordinates, this working-space reflection corresponds to $R(\boldsymbol{x}_n - \boldsymbol{\mu}) + \boldsymbol{\mu}$; it coincides with $R\boldsymbol{x}_n$ when $R\boldsymbol{\mu} = \boldsymbol{\mu}$. The reflection reverses horizontal coordinates and exchanges corresponding anatomical left-right keypoints; it models 2D reflection equivalence and does not imply general 3D viewpoint invariance.

For two centered skeleton samples $\bar{\boldsymbol{x}}_1$ and $\bar{\boldsymbol{x}}_2$, we compute

$$d_{1,2} = \|\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2\|_2, \tag{5}$$

$$\hat{d}_{1,2} = \|R\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2\|_2 = \|\bar{\boldsymbol{x}}_1 - R\bar{\boldsymbol{x}}_2\|_2. \tag{6}$$

Here, $\|\cdot\|_2$ denotes the Euclidean norm. We then define the line-symmetry-aware distance as

$$\tilde{d}_{1,2} = \begin{cases} d_{1,2} & \text{if } d_{1,2} \leq \hat{d}_{1,2}, \\ \hat{d}_{1,2} & \text{otherwise.} \end{cases} \tag{7}$$

Because R is an involutive orthogonal transformation, $\tilde{d}_{1,2}$ is a pseudometric on the centered sample space: distinct reflection-related samples can have zero distance. This distance captures the equivalence relation central to our primary goal: two postures should not require separate representatives merely because one is the horizontal reflection of the other.

3.2 Incorporating Our Distance Function into K-Means

LSAK specializes transformation-invariant clustering (Frey and Jojic, 2003) to two transformations, the identity and the predefined reflection R, in the centered skeleton-coordinate space. It combines reflection-aware assignment with coordinate-wise centroid updates to reduce redundant representatives. Unlike the observed-pose matching used for retrieval (Jenicek and Chum, 2019), these updates estimate K representatives from the assigned samples.

For each centered skeleton sample, LSAK considers two candidate representations, $\bar{\boldsymbol{x}}_n$ and its line-symmetrical counterpart $\hat{\boldsymbol{x}}_n = l(\bar{\boldsymbol{x}}_n)$. It computes the distances from both representations to the working centroids and uses the closer representation for cluster assignment. Before each centroid update, a cluster-wise orientation alignment is applied so that reflection-equivalent representations within a cluster are averaged in a consistent orientation. This alignment does not merge clusters after the fact; rather, it determines the common orientation in which the members of each current cluster contribute to its next centroid. The method then repeats assignment and centroid update until its stopping criterion is satisfied. Centroids are initialized by random sampling, as in k-means. We describe sample assignment in Section 3.3 and centroid update in Section 3.4.

Here, the hat denotes reflection of the centered representation, and the overbar is omitted beneath the hat for compact notation. Let $K \in \mathbb{N}$ be the number of clusters, $\bar{\boldsymbol{c}}_k^{(t)}$ the centered working centroid of cluster $k \in \{1,\ldots,K\}$ at iteration t, and $\bar{C}^{(t)} = \{\bar{\boldsymbol{c}}_1^{(t)},\ldots,\bar{\boldsymbol{c}}_K^{(t)}\}$ the centered centroid set. The overbar is used only for quantities in the dataset-centered working space; the final centroids after adding back $\boldsymbol{\mu}$ are denoted without an overbar.

3.3 Sample Assignment

At iteration t, we compute the minimum distances from the centered original sample $\bar{\boldsymbol{x}}_n$ and its reflected counterpart $\hat{\boldsymbol{x}}_n$ to the K centered working centroids, and select the smaller distance:

$$\begin{aligned} d_n^{(t)} &= \min_{1\leq k\leq K} \|\bar{\boldsymbol{x}}_n - \bar{\boldsymbol{c}}_k^{(t)}\|_2, \\ \hat{d}_n^{(t)} &= \min_{1\leq k\leq K} \|\hat{\boldsymbol{x}}_n - \bar{\boldsymbol{c}}_k^{(t)}\|_2, \\ \tilde{d}_n^{(t)} &= \min\{d_n^{(t)}, \hat{d}_n^{(t)}\}. \end{aligned} \tag{8}$$

The selected representation $\tilde{\boldsymbol{x}}_n$ and cluster label $\tilde{k}_n$ are recomputed at each iteration; their iteration superscripts are omitted for brevity. The tilde therefore denotes the representation selected for the current assignment step, not an additional normalization of the input. We determine whether the sample is represented by its centered original or reflected form as

$$\tilde{\boldsymbol{x}}_n = \begin{cases} \bar{\boldsymbol{x}}_n & \text{if } d_n^{(t)} \leq \hat{d}_n^{(t)}, \\ \hat{\boldsymbol{x}}_n & \text{otherwise.} \end{cases} \tag{9}$$

We then assign the selected representation to the nearest centered centroid:

$$\tilde{k}_n = \operatorname*{argmin}_{k\in\{1,\ldots,K\}} \|\tilde{\boldsymbol{x}}_n - \bar{\boldsymbol{c}}_k^{(t)}\|_2. \tag{10}$$

We repeat Eqs. (8)–(10) until all N skeleton samples are processed. The use of two reflection-equivalent candidates at this stage is the mechanism that directly targets the redundancy problem: mirrored postures compete for the same cluster representatives instead of being forced to occupy separate cluster representatives.

3.4 Centroid Update

At iteration t, let $\mathcal{A}_k^{(t)}$ be the set of sample indices assigned to cluster k. Among these samples, let $n_{k,\text{orig}}^{(t)}$ and $n_{k,\text{ref}}^{(t)}$ denote the numbers for which Eq. (9) selects the centered original and line-symmetrical representations, respectively. To keep reflection-equivalent samples in a consistent orientation before averaging, the implementation jointly switches the selected representations when the reflected representation is selected by a strict majority in the cluster. For $n \in \mathcal{A}_k^{(t)}$,

we therefore define the representation used for centroid update as

$$\boldsymbol{s}_n^{(t)} = \begin{cases} \tilde{\boldsymbol{x}}_n, & \text{if } n_{k,\text{orig}}^{(t)} \geq n_{k,\text{ref}}^{(t)}, \\ l(\tilde{\boldsymbol{x}}_n), & \text{otherwise.} \end{cases} \tag{11}$$

This operation changes only the common orientation used to represent a cluster; it does not change its sample membership. The next centered working centroid is then

$$\bar{\boldsymbol{c}}_k^{(t+1)} = \frac{1}{|\mathcal{A}_k^{(t)}|} \sum_{n \in \mathcal{A}_k^{(t)}} \boldsymbol{s}_n^{(t)}. \tag{12}$$

Each input entry is counted separately, including repeated vectors in the duplication experiments. This update makes the role of the orientation alignment explicit: it aligns reflection-equivalent members within a cluster before taking their Euclidean mean, while the cluster assignment itself remains determined by Eqs. (8)–(10).

The iterations stop when the cluster labels remain unchanged between two consecutive iterations or when the maximum of $T_{\max} = 300$ iterations is reached. If an initialization produces an empty cluster, that run is discarded and clustering is restarted from a new initialization. The assignment/update steps can be interpreted through the Lloyd-type objective

$$J(\bar{C}) = \sum_{n=1}^{N} \min_{1 \leq k \leq K} \min \left\{ \|\bar{\boldsymbol{x}}_n - \bar{\boldsymbol{c}}_k\|_2^2, \|\hat{\boldsymbol{x}}_n - \bar{\boldsymbol{c}}_k\|_2^2 \right\}. \tag{13}$$

For fixed centroids, the assignment step selects the minimum term for each sample; for fixed assignments and selected orientations, Eq. (12) minimizes the within-cluster squared error. The cluster-wise orientation alignment reflects all selected representations of a cluster jointly, so orthogonality of R preserves the minimized within-cluster squared distances. Accordingly, each completed assignment/update cycle is non-increasing in J. The implementation nevertheless uses label stability (or $T_{\max}$) as its practical stopping rule. Let L denote the final iteration returned by this procedure. The dataset mean is then added back to each centered working centroid,

$$\boldsymbol{c}_k^{(L)} = \bar{\boldsymbol{c}}_k^{(L)} + \boldsymbol{\mu}, \tag{14}$$

and $\boldsymbol{c}_k^{(L)}$ is the final centroid used in the centroid-level evaluation and visualization. Separating $\bar{\boldsymbol{c}}_k^{(t)}$ from $\boldsymbol{c}_k^{(L)}$ avoids conflating the numerically centered optimization space with the coordinates in which the final representative postures are reported.

4 EXPERIMENTS

4.1 Dataset

All experiments used DeepFashion-MultiModal (Liu et al., 2016; Jiang et al., 2022). We selected 1,215 images of female fashion models with annotated skeletons, excluding samples with partially missing skeletons or sitting postures (i.e., seated on a chair). Each skeleton contains $z = 13$ keypoints: the nose and the bilateral shoulders, elbows, wrists, hips, knees, and ankles, giving a 26-dimensional vector.

4.2 Quantitative Evaluation

4.2.1 Quantitative Settings

We evaluated inter-cluster redundancy (Issue I_1) using Pairs and reported SSE to describe compactness. The baselines were k-means (MacQueen, 1967), k-means++ (Arthur and Vassilvitskii, 2007), Faster k-medoids (Schubert and Rousseeuw, 2019), SBKM (Su and Chou, 2001), and MPSK (Chung and Lin, 2007). We used $K \in \{8, 16\}$ and report means over 100 trials per setting. For k-means, k-means++, and LSAK, iterations stopped at unchanged cluster labels or 300 iterations. SBKM and MPSK used three k-means initialization iterations followed by at most 300 fine-tuning iterations, stopping at stable labels or a repeated label pattern. Faster k-medoids used its standard swap-based stopping rule. The additional evaluations in Sections 4.4 and 4.5 use the same settings unless stated otherwise. Mirror augmentation uses the reflection in Section 3.1, including left-right keypoint permutation.

4.2.2 Primary Metric: Number of Line-Symmetrical Centroid Pairs (Pairs)

Pairs counts disjoint line-symmetrical centroid pairs to quantify inter-cluster redundancy. For each returned centroid $\boldsymbol{c}_k^{(L)}$, we find the nearest other centroid to its reflection $l(\boldsymbol{c}_k^{(L)})$:

$$j_k = \operatorname*{argmin}_{i \in \{1,\ldots,K\}, i \neq k} \left\| l(\boldsymbol{c}_k^{(L)}) - \boldsymbol{c}_i^{(L)} \right\|_2, \tag{15}$$

and define

$$\delta_k = \left\| l(\boldsymbol{c}_k^{(L)}) - \boldsymbol{c}_{j_k}^{(L)} \right\|_2. \tag{16}$$

A pair is counted once when $j_k = i$, $j_i = k$, and both distances are below ε. All centroids remain candidates throughout the search; none is removed after matching. Since each centroid selects one partner, mutual matching produces disjoint pairs, giving

Table 1: Mean Pairs and SSE over 100 trials. Pairs measures centroid redundancy; SSE describes compactness in the working representation used by each method. Lower values are indicated by ↓.

K	Method	Pairs ↓	SSE/10^3 ↓
8	k-means	2.94	1.67
	k-means++	2.94	1.68
	Faster k-medoids	2.00	1.85
	SBKM	2.88	1.69
	MPSK	2.45	1.77
	LSAK (ours)	0.00	1.35
16	k-means	4.75	1.31
	k-means++	4.52	1.30
	Faster k-medoids	2.00	1.59
	SBKM	3.82	1.50
	MPSK	3.28	1.58
	LSAK (ours)	0.69	1.04

Pairs $\leq K/2$. We set $\varepsilon = 0.9$ and examine threshold sensitivity in Section 4.5. Fewer detected pairs indicate less reflection-related redundancy. LSAK does not directly minimize this centroid-level count: its assignment and update steps use sample-to-centroid distances.

4.2.3 Sum of Squared Errors (SSE)

SSE describes compactness in the working representation used by each method:

$$\mathrm{SSE} = \sum_{n=1}^{N} \left\| \boldsymbol{s}_n - \bar{\boldsymbol{c}}_{a_n}^{(L)} \right\|_2^2, \tag{17}$$

where a_n is the assigned cluster. For the baselines, $\boldsymbol{s}_n = \bar{\boldsymbol{x}}_n$; for LSAK, $\boldsymbol{s}_n$ is the final reflection-aware representation used for centroid estimation. The centered centroid $\bar{\boldsymbol{c}}_{a_n}^{(L)}$ is used because SSE is computed before restoring the dataset mean. The distance is unweighted Euclidean. As the working representations differ, SSE is descriptive rather than a common reflection-invariant comparison across methods.

4.2.4 Quantitative Results

Table 1 reports Pairs and SSE. For $K = 8$ and $K = 16$, LSAK reduced mean Pairs from the five-baseline averages of 2.6 and 3.7 to 0.00 and 0.69, respectively. These results support the primary objective of reducing reflection-based redundancy among representative postures. LSAK also yielded the lowest reported SSE for both values of K, describing compact clusters in its reflection-aware working representation. Sections 4.4 and 4.5 further examine Pairs under alternative input constructions and detection thresholds.

4.3 Qualitative Evaluation of Representative Postures

4.3.1 Qualitative Settings

We compared k-means (MacQueen, 1967) and LSAK at $K = 8$ by showing each centroid with its five nearest skeleton samples. For LSAK visualization only, one final centroid was chosen as a reference; each remaining centroid was displayed as $\boldsymbol{c}_k^{(L)}$ or $l(\boldsymbol{c}_k^{(L)})$, whichever was closer to that reference. This *display-only centroid reflection* was applied after computing Pairs and SSE and did not affect either metric. It is distinct from the cluster-wise orientation alignment in Eq. (11), which is part of iterative clustering. The other settings followed Section 4.2.1.

4.3.2 Qualitative Results

Figure 2(a) highlights three visually identified line-symmetrical centroid pairs. The pairs $(\boldsymbol{c}_1^{(L)}, \boldsymbol{c}_2^{(L)})$, $(\boldsymbol{c}_3^{(L)}, \boldsymbol{c}_4^{(L)})$, and $(\boldsymbol{c}_5^{(L)}, \boldsymbol{c}_6^{(L)})$ in the k-means result correspond visually, up to reflection, to $\boldsymbol{c}_1^{(L)}$, $\boldsymbol{c}_2^{(L)}$, and $\boldsymbol{c}_3^{(L)}$ in the LSAK result, respectively. The LSAK centroids $\boldsymbol{c}_6^{(L)}$, $\boldsymbol{c}_7^{(L)}$, and $\boldsymbol{c}_8^{(L)}$ show additional posture patterns that are less apparent among the k-means centroids. These are visual correspondences between independently obtained results, not a post-processing step that merges baseline clusters.

Original (red) and reflected (blue) samples appear within the same clusters in Figure 2(b), illustrating how reflection-related samples can contribute to a common representative posture. The five nearest samples illustrate this behavior rather than the variability of entire clusters; centroid redundancy is evaluated separately by Pairs. Overall, the figure illustrates how reduced redundancy can leave more representatives available for distinct posture patterns when the number of clusters is fixed. The resulting potential to reduce unnecessary comparisons during posture inspection remains to be evaluated with users.

4.4 Effect of Sample Duplication and Mirror Augmentation

We compared three input conditions: the original samples (Original), an exact duplication of the sample list (Duplication), and the original samples concatenated with their reflected counterparts (Mirror augmentation). Original contains $N = 1{,}215$ entries; the other conditions contain $2N = 2{,}430$ entries, including repeated vectors rather than additional independently observed postures. Dataset-level centering

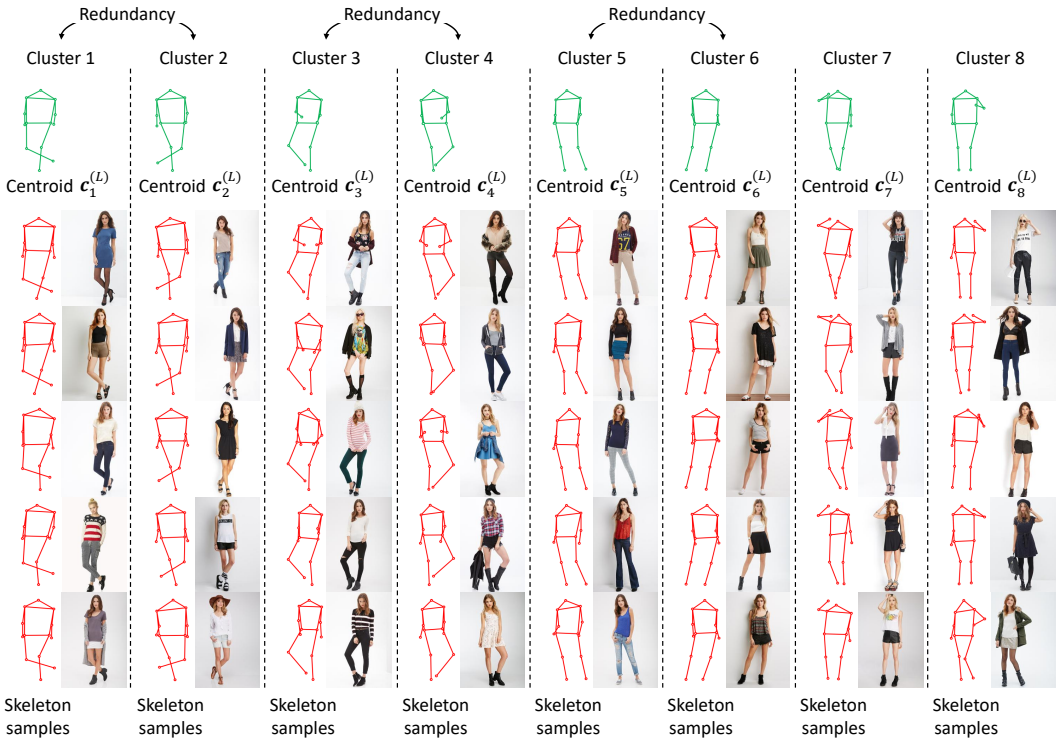


Figure 2: Representative postures and five nearby samples for $K = 8$: (a) k-means and (b) LSAK. In (b), original and reflected sample representations are red and blue, respectively; centroids use the display-only reflection in Section 4.3.1.

Table 2: Mean Pairs for Original (N entries), Duplication, and Mirror augmentation ($2N$ entries), with $\varepsilon = 0.9$.

K	Method	Original	Duplication	Mirror augmentation
8	k-means	2.94	3.04	3.05
	k-means++	2.94	2.99	3.02
	Faster k-medoids	2.00	2.00	2.00
	LSAK (ours)	0.00	0.00	0.12
16	k-means	4.75	4.82	5.80
	k-means++	4.52	4.74	5.47
	Faster k-medoids	2.00	2.00	3.00
	LSAK (ours)	0.69	0.58	0.89

was recomputed for each condition. We evaluated k-means, k-means++, Faster k-medoids, and LSAK with $K \in \{8, 16\}$ and $\varepsilon = 0.9$, using the remaining settings in Section 4.2.1. This comparison covers these four methods; the main comparison with all five baselines is given in Table 1. Table 2 reports means over 100 trials and repeats the Original results for reference.

As shown in Table 2, exact duplication does not eliminate line-symmetrical centroid pairs in the conventional baselines. Mirror augmentation also leaves redundant pairs and can increase their number: for k-means at $K = 16$, Pairs increases from 4.75 to 5.80. LSAK yields lower Pairs than all three baselines in every condition, although mirror augmentation increases its Pairs from 0.00 to 0.12 for $K = 8$ and from 0.69 to 0.89 for $K = 16$.

Mirror augmentation supplies the original and reflected samples as separate entries to an unchanged algorithm. In contrast, LSAK selects a representation during assignment and averages the orientation-aligned samples during centroid updates. The results show that appending reflected samples does not substitute for this procedure in the evaluated settings. Duplication controls for repeated observations, not robustness to new postures, noise, or viewpoint changes.

4.5 Sensitivity to the Pair-Detection Threshold

We examined $\varepsilon \in \{0.5, 0.7, 0.9, 1.1, 1.3\}$ for k-means, k-means++, and LSAK at $K \in \{8, 16\}$. This sweep focuses on the two closest k-means baselines; Table 1 compares all five baselines. The threshold controls pair detection, not clustering. As shown in Table 3, LSAK yields lower Pairs than both k-means and k-means++ throughout the tested range. Values at $\varepsilon = 0.9$ match Table 1. Pairs is not always zero: at $\varepsilon = 1.3$, LSAK reaches 0.58 and 1.62 for $K = 8$ and $K = 16$, respectively.

Table 3: Mean Pairs at five detection thresholds. Values at $\varepsilon = 0.9$ match Table 1.

		Pair-detection threshold ε				
K	Method	0.5	0.7	0.9	1.1	1.3
8	k-means	2.50	2.74	2.94	3.00	3.03
	k-means++	2.39	2.67	2.94	3.03	3.10
	LSAK (ours)	0.00	0.00	0.00	0.00	0.58
16	k-means	2.70	3.99	4.75	5.25	5.46
	k-means++	2.90	3.78	4.52	5.12	5.25
	LSAK (ours)	0.10	0.44	0.69	1.01	1.62

Table 4: Mean iteration counts and computation times over 100 trials under the settings in Section 4.6.

K	Method	Num. iterations	Time [ms]
8	k-means	21.74	62
	k-means++	21.07	69
	Faster k-medoids	2.00	42
	SBKM	21.86	2,708
	MPSK	4.71	637
	LSAK (ours)	21.81	83
16	k-means	22.57	106
	k-means++	19.69	111
	Faster k-medoids	3.00	39
	SBKM	21.41	3,036
	MPSK	4.64	768
	LSAK (ours)	23.77	168

4.6 Computational Cost

We measured the number of iterations and computation time until stopping, using an Intel Core i7-7800X CPU with 64 GB of memory. The implementation used Python 3.10.12 and Cython-compiled extensions with GCC 11.4.0 and compiler optimizations disabled. Table 4 reports means over 100 trials under the stopping criteria and other settings in Section 4.2.1.

For $K = 8$ and $K = 16$, LSAK required 21.81 and 23.77 iterations and 83 and 168 ms, respectively. The corresponding k-means times were 62 and 106 ms, giving runtime ratios of approximately 1.34 and 1.58. As shown in Table 4, LSAK was faster than SBKM and MPSK for both cluster counts, but slower than k-means, k-means++, and Faster k-medoids. These comparisons describe the tested implementations and compilation settings.

4.7 Limitations

The evaluation is limited to one fashion-image dataset, two cluster counts ($K = 8, 16$), and 2D skeletons with complete annotations. Generalization to

other domains, viewpoints, missing keypoints, and noisy or swapped keypoints remains untested. The comparisons establish performance relative to the evaluated baselines, not superiority over untested pose-clustering or embedding methods. Pairs measures reflection-related redundancy rather than general semantic clustering quality or non-reflected near-duplicates. SSE is computed using the working representation of each method and is descriptive, not a common reflection-invariant criterion. After dataset centering, the working-space reflection corresponds to $R(\boldsymbol{x}-\boldsymbol{\mu})+\boldsymbol{\mu}$ in the original coordinates, whereas Pairs applies R to the returned centroids. These coincide when $R\boldsymbol{\mu}=\boldsymbol{\mu}$ and otherwise differ by $\boldsymbol{\mu}-R\boldsymbol{\mu}$. Runtime depends on the implementation, and direct user evaluation remains future work.

5 CONCLUSIONS

We presented LSAK, a coordinate-space specialization of transformation-invariant clustering for representative 2D postures. The method combines cluster-dependent reflection selection with orientation-aligned mean updates and evaluates the redundancy of the resulting centroid set at fixed K. On the evaluated dataset, LSAK reduced mean Pairs to 0.00 ($K=8$) and 0.69 ($K=16$), compared with five-baseline averages of 2.6 and 3.7. The reduction persisted under sample duplication, mirror augmentation, and every tested pair-detection threshold from 0.5 to 1.3. LSAK also yielded the lowest reported SSE for both cluster counts, and qualitative results illustrated additional posture patterns within a fixed number of representatives. Together, these results support the main objective of suppressing line-symmetrical centroid redundancy while preserving compact representative clusters.

Future work will focus on three directions. First, we plan to extend the formulation to 3D skeletons and investigate the influence of viewpoint changes on symmetry-based posture similarity. Second, we will evaluate robustness to missing keypoints and pose estimation errors, and explore distance functions that account for keypoint confidence. Third, we plan to conduct user studies to assess the practical value of the extracted representative postures. By measuring task completion time and accuracy in posture search and comparison tasks, we will examine whether reduced centroid redundancy helps users investigate postures more efficiently.

We would like to thank Dr. Michiko Inoue and Dr. Shunsuke Yoneda for their valuable suggestions during this research.

REFERENCES

Arthur, D. and Vassilvitskii, S. (2007). *k*-means++: The Advantages of Careful Seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1027–1035.

Chung, K.-L. and Lin, J.-S. (2007). Faster and More Robust Point Symmetry-Based *K*-Means Algorithm. *Pattern Recognition*, 40(2):410–422.

Chung, K.-L. and Lin, K.-S. (2006). An Efficient Line Symmetry-Based *K*-Means Algorithm. *Pattern Recognition Letters*, 27(7):765–772.

Frey, B. J. and Jojic, N. (2003). Transformation-Invariant Clustering Using the EM Algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(1):1–17.

Jenicek, T. and Chum, O. (2019). Linking Art through Human Poses. In *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*, pages 1338–1345.

Jiang, Y., Yang, S., Qiu, H., Wu, W., Loy, C. C., and Liu, Z. (2022). Text2Human: text-driven controllable human image generation. *ACM Transactions on Graphics*, 41(4):162:1–162:11.

Liu, Z., Luo, P., Qiu, S., Wang, X., and Tang, X. (2016). DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1096–1104.

MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297.

Schubert, E. and Rousseeuw, P. J. (2019). Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms. In *Proceedings of the 12th International Conference on Similarity Search and Applications*, pages 171–187.

Su, M.-C. and Chou, C.-H. (2001). A Modified Version of the *K*-Means Algorithm with a Distance Based on Cluster Symmetry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):674–680.

Xu, Y., Zhang, J., Zhang, Q., and Tao, D. (2022). ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation. *Advances in Neural Information Processing Systems*, 35:38571–38584.

Yang, Z., Zeng, A., Yuan, C., and Li, Y. (2023). Effective Whole-Body Pose Estimation with Two-Stages Distillation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 4210–4220.