Textile & Leather Review

ISSN 2623-6281 | www.tlr-journal.com | 10.31881/TLR

Efficient and Interpretable Robot Vision Framework for Virtual Matching of Colored Garments Using CNN, PCA, and Binary Hashing

Zongbo Liu

How to cite: Liu Z. Efficient and Interpretable Robot Vision Framework for Virtual Matching of Colored Garments Using CNN, PCA, and Binary Hashing. Textile & Leather Review. 2025; 8:5010-5032. https://doi.org/10.31881/TLR.2026.5010

How to link: https://doi.org/10.31881/TLR.2026.5010

Published 27 April 2026

Efficient and Interpretable Robot Vision Framework for Virtual Matching of Colored Garments Using CNN, PCA, and Binary Hashing

Zongbo LIU

Lanzhou Modern Vocational College, Lanzhou, Gansu, 730000, China

Gansu Vocational College of Architecture, Lanzhou, Gansu, 730050, China

Academician Expert Workstation of Gansu Dayu Jiuzhou Space Information Technology Co.,Ltd. Lanzhou, Gansu, 730050, China
liuzb88@126.com

Article
https://doi.org/10.31881/TLR.2026.5010
Published 27 April 2026

ABSTRACT
Conventional garment retrieval and fitting systems often generate noisy matches and deviations from the actual fitting process, limiting their practical use in online applications. This study introduces a robot vision–based framework that supports efficient and interpretable virtual matching of colored garments. The system combines convolutional neural networks (CNN) for feature extraction, principal component analysis (PCA) for dimensionality reduction, binary hashing for accelerated retrieval, and ontology for semantic organization. Efficiency is achieved through a lightweight CNN structure with progressive pooling layers, dropout-regularized dense stages, and PCA–hash compression, all of which reduce computational cost while sustaining accuracy. Interpretability is reinforced with visual demonstrations, including query–retrieval outcomes and pose normalization comparisons. Experimental evaluation indicates that the framework achieves retrieval accuracy of approximately 88%, significantly outperforming traditional approaches such as HOG and SIFT. These findings confirm that the proposed approach can reliably support large-scale online garment search and virtual fitting.

KEYWORDS
new robot vision technology, convolutional neural network (CNN), hashing; virtual fitting, interpretability

INTRODUCTION

With the continuous progress of the economy and society, the Internet has penetrated into many areas such as “garments, food, housing and transportation”, and online garments selection and fitting has become a possible option [1-2].

Online garment shopping has rapidly evolved with the integration of computer vision and artificial intelligence. While keyword-based search remains fast and efficient, it faces limitations when handling subjective image labeling and complex garment attributes. Visual-based retrieval systems powered by robot vision can overcome these limitations by directly analyzing garment features such as color, shape, and texture. This study focuses on developing such a system to automate garment analysis and virtual matching efficiently.

When computer technology is fused with virtual fitting and matching systems, the primary purpose is to integrate the computer's technology for the online display of garments and to achieve the effect that users can apply the virtual models for fitting operations; while the modeling of the human body, the modeling of garments, and the fusion needs to be studied [11-12]. Industry experts for the modeling of garments mainly focus on the use of garments for effective simulation, such as the use of 2D images for the effective mapping of pictures of garments and attach them to the human body model; or using the 3D image processing for the 3D modeling of garments to implement animation simulation of the human body and the simulation of garments. Nevertheless, how to reconstruct the two 3D methods is the key research direction [13-14]. Other scholars have tried to conduct the classification analysis of garment attributes, such as color and sleeves [15-16].

Certainly, what is more important is that the diversity of garments can lead to the image complexity of garments. Thus the artificial features are increased with the increase in the number of images, which makes it increasingly tough to carry out feature extraction and achieve the specific utilization. In response to these demands and defects, a new kind of robot vision technology is introduced in this paper to analyze the business logic by analyzing the colored garments virtual matching system design and analyzing the business logic, taking the establishment of garments database as the breakthrough point to establish the automation of analyzing garment attributes for different image complexity. At the same time, the corresponding hash index is established to implement the effective matching and rapid retrieval of garments images to comply with the practical requirements, with the aim to improve the automation and efficiency of the relevant services.

Research Contribution

- The contribution of the research lies in its integration of a structured pipeline comprising CNN, PCA, Hashing, and Ontology within the fashion recommendation domain. While the individual components of this pipeline are well-established in machine learning and information retrieval research, their combination for solving the challenges of garment retrieval and virtual matching offers practical significance.
- Specifically, the use of CNN ensures robust feature extraction, PCA reduces dimensionality to improve computational efficiency, hashing accelerates retrieval, and Ontology enhances semantic interpretation.
- This synergy addresses issues of image complexity, retrieval accuracy, and semantic ambiguity in clothing recommendation. Therefore, the research is positioned as an applied contribution that strengthens the bridge between theoretical models and their real-world utility in e-commerce and virtual fitting systems.

GOAL STATEMENT

The central aim of this study is to design a garment retrieval and virtual fitting system that is both computationally efficient and interpretable in practice. To achieve this, the framework integrates CNN-based visual analysis with PCA for dimensionality reduction, hash indexing for rapid search, and ontology for semantic refinement. The design not only improves accuracy and speed but also emphasizes interpretability, demonstrated through retrieval examples and pose normalization results.

RESEARCH METHODS

New Robot Vision Technology

The research follows a modular design in which each component—feature extraction, dimensionality reduction, retrieval, semantic analysis, and fitting—is developed and validated individually before being combined into a single framework. This design enhances reproducibility, allows scalability to large image databases, and ensures adaptability across varied garment categories.

The application of the new robot vision technology is shown in Figure 1 below. In the face of the new robot vision technology, it is necessary to create a relevant search template based on the features of garments and test the multi-trait, multi-data-based garments database established.

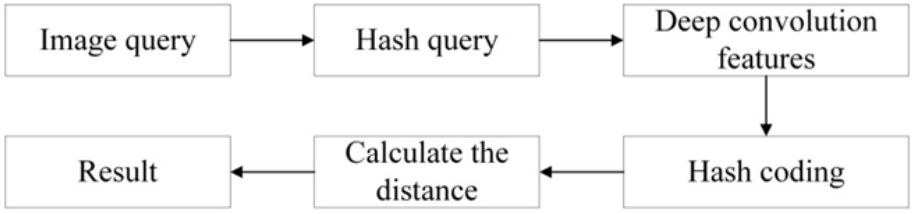

Figure 1. Schematic diagram of the garments retrieval model

With regard to the classification models, they are classified mainly by using the convolutional neural network (CNN), and a database for the classification of garments is established in the context of garment attributes based on the training of the corresponding model. Subsequently, the established database is used to implement the training framework analysis and the specific parameter adjustment of the corresponding neural network.

Establishment of the B_DAT Garment Database

As far as the publicly available data set is concerned, it contains a small amount of image attributes, which cannot be used to implement the image extraction and analysis of multiple attribute features;

in addition, the efficiency of classification retrieval can be reduced due to too little image data. The specific number of selected samples is displayed in Figure 2 and Figure 3 as the following:

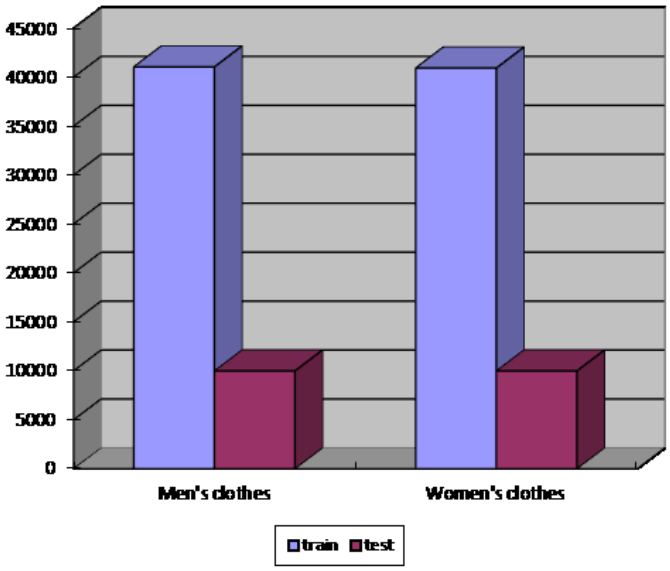

Figure 2. Distribution of the training samples and the testing samples combined

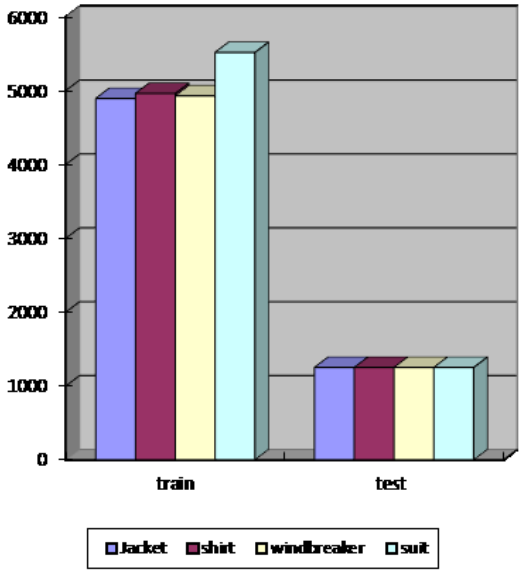

Figure 3. Training of men's clothes and distribution of the testing samples

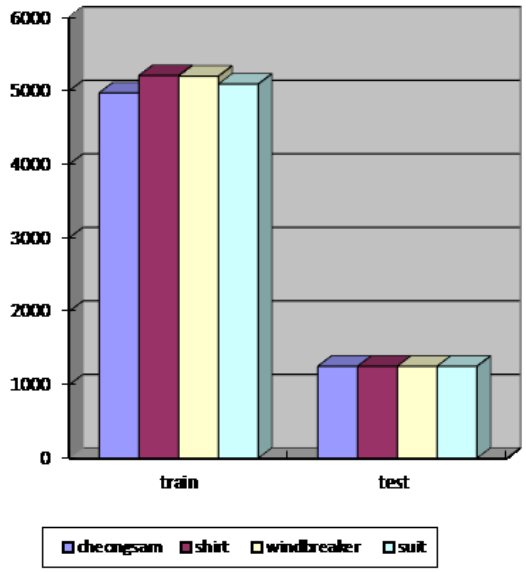

Figure 4. Distribution of the training samples and the testing samples for women's clothes

It can be observed from the results in Figure 4 above that the samples in the database can be relevant images with a simple background, as well as the original images of the model dressing, or even the clothes alone.

The B_DAT dataset developed for this research contains 18,500 garment images, comprising 9,000 men's and 9,500 women's garments, covering multiple categories such as T-shirts, shirts, jackets, trousers, and dresses. Each image includes both clean-background and model-wearing conditions. The dataset was constructed from publicly available e-commerce image sources and annotated for color, sleeve type, and garment category. Due to licensing limitations, B_DAT is currently proprietary but can be shared with academic collaborators upon reasonable request for research purposes.

Learning of the Features of Garment Attributes

For the purpose of achieving better analysis and application results of garments images, attempts are made in this paper to introduce deep learning related algorithms. As far as deep learning is concerned, its essence is to implement effective feature extraction by creating massive machine learning models and huge amounts of data to achieve classification and prediction accuracy. Thus, deep learning algorithms have been extensively applied in various computer vision tasks such as classification, segmentation, detection, and analysis [17, 18, 29, 30]. These networks, particularly convolutional neural networks (CNNs), have demonstrated strong capability for hierarchical feature extraction, making them effective for garment attribute understanding. The effective analysis of deep learning is achieved by using the basic shallow fundamental data. In this paper, the advanced formal features are

included to carry out the relevant analysis in response to practical demands, and the specific framework and Figure 5 as the following. In the figure, the input image is 256*256, and the fully connected layer is reconciled on the basis of the framework as well as the specific model.

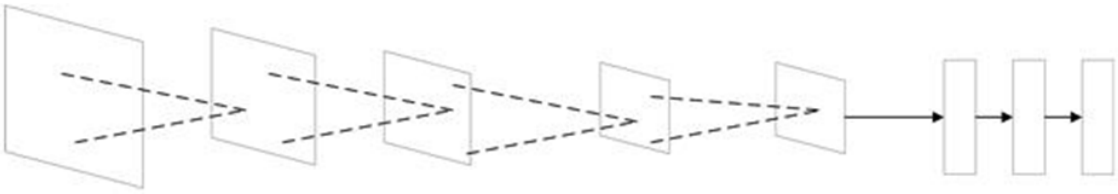

Figure 5. Architecture of the CNN network

The comprehensive consideration of the fully connected layer-related parameters can result in excessive fitting or overfitting. For the purpose of avoiding overfitting, with regard to the data requirements of this paper, the last layer of fully connected layer is adjusted and set accordingly, which has reduced the workload and improved the efficiency of the calculation. The specific network error calculation is shown in the equation (1) as the following:

$$err = \frac{1}{2} \times \frac{\sum_{j=1}^{n} \sum_{i=1}^{m} \left(t - t_{label}\right)^2}{n} \tag{1}$$

The corresponding parameters on the specific model are adjusted accordingly, and the number of iterations of training is increased to continuously optimize and obtain the adapted database. The details are shown in Figure 6 as the following.

As far as different HOG features are concerned, it is mainly reflected in the connections between different layers. In this regard, the neural network can be used acquire global feature information. Hence, the neural network is highly effective for the efficient acquisition of the relevant semantic features and the understanding of the relevant classification of garments[19-20].

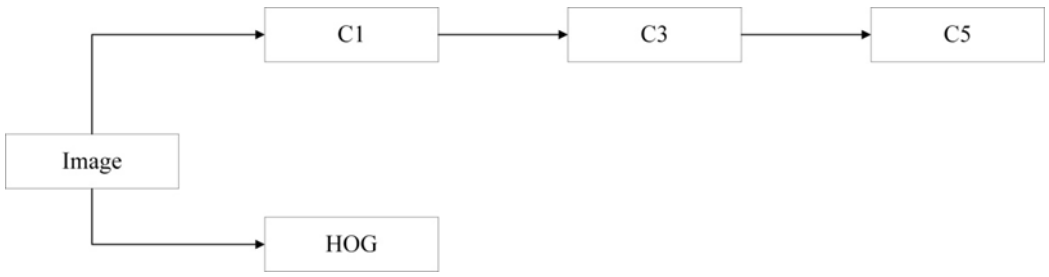

Figure 6. Visualization of CNN and HOG features

The simulation of the 3D coordinates of the key points in the human body is carried out effectively by using the corresponding mature software, which are further expressed in accordance with specific values that are composed of the specific feature vectors. The details are shown in the equation (2) as the following:

$$RT = \left[T_1, T_2, \cdots \right] \tag{2}$$

On the basis of the equation (2) above, the representation of specific three-dimensional coordinates is carried out. The details are shown in the equation (3) as the following:

$$T_i = \left[x_i, y_i, z_i\right], i \in \left[1, 24\right] \tag{3}$$

The algorithm proposed in this research primarily follows the following process flow:

(1) Efficient extraction for the current image, extraction of skeletal positions by using the garment retrieval depth knowledge, and construction of the relevant feature vectors for analysis.

(2) Retrieval in the database for feature matching.

(3) Acquisition of the target image and fitting it to the present frame.

The specific flowchart of the mehod is presented in Figure 7 as the following.

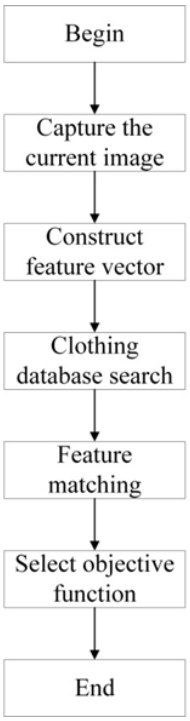

Figure 7. Process flow of the image matching algorithm

To enhance interpretability and visual understanding of the proposed framework, additional illustrative examples are incorporated to demonstrate how each major component contributes to the overall virtual matching process. Figure 8(a) visualizes the sequence of feature extraction, where the input garment image passes through the convolutional–pooling stages to yield progressively refined visual maps. Figure 8(b) presents the PCA-based dimensionality reduction, showing the transformation of high-dimensional CNN features into a compact latent space. Figure 8(c) illustrates the binary hashing process that converts the reduced features into hash codes for fast retrieval. Together, these diagrams provide a clear view of how the system transitions from raw garment imagery to semantically meaningful, efficiently retrievable representations. Such visualizations strengthen methodological transparency and demonstrate how the proposed robot vision framework achieves both computational efficiency and interpretability.

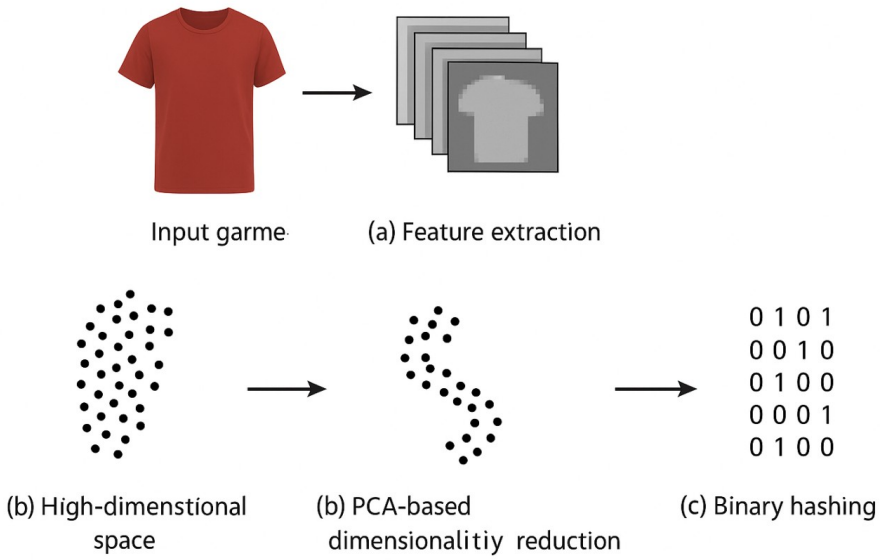

Figure 8. Illustrative examples of the proposed framework

In the specific algorithm process, the database is used for its core calculation, and effective matching of the image keywords in the database is carried out based on the features extracted by the algorithm. In this paper, the corresponding method is selected for image selection in the database, and the distance analysis is conducted for each image that has been pre-processed. The target image is selected through the specific optimization and settings, and the specific calculation is shown in the equation (4) as the following:

$$F = \mu A\left(l_{i-1}, l_i\right) + B\left(l_i, i\right) + \lambda C\left(l_{i+1}, l_i\right) \tag{4}$$

In the above equation, the corresponding threshold values are set and used for the settings of the specific parameters. The detailed calculation is shown in equation (5), equation (6), and equation (7) as the following:

$$i=\left[i_1,i_2,\cdots\right. \tag{5}$$

$$l_i=\left[l_{i1},l_{i2},\cdots\right. \tag{6}$$

$$B\left(l_i,i\right)=\sqrt{\left(l_{i1}-i_1\right)^2+\left(l_{i2}-i_2\right)^2+\cdots\left(\ldots i_n\right)^2}=\sqrt{\sum_i^n\left(l_{ij}-i_j\right)^2} \tag{7}$$

In the specific calculation, the Euclidean distance is selected in this paper as the specific index for calculating the distance. In this way, the matching and analysis of each feature point can be implemented. For the purpose of better reflecting the validity and connectivity of adjacent frames, specific discriminant functions are used to make effective judgments. The details are shown in Equation (8) and Equation (9) as the following:

$$A\left(l_{i-1},l_i\right)=\begin{cases}0,i-1\ and\ i\ are\ adjacent\ frames\\1,i-1\ and\ i\ are\ non-adjacent\ frames\end{cases} \tag{8}$$

$$C\left(l_i,l_{i+1}\right)=\begin{cases}0,i\ and\ i+1\ are\ adjacent\ frames\\1,i\ and\ i+1\ are\ non-adjacent\ frames\end{cases} \tag{9}$$

In the above equation, it is necessary to determine the present database image and the image of the previous data of the search results of the link. Through the comprehensive judgment, it can be known that the image without the adjacent frame should have a result greater than the previous frame of the search results. For the purpose of ensuring a greater degree to become an indicator of the image, the image matching equation of the continuity judgment function is used to make the judgement. This method allows for an increase in the threshold value of the images that are not adjacent frames accordingly. For the purpose of implementing the separate reflection of garments, there should not be too frequent flickering or bursts, and it is necessary to ensure that the image matching of adjacent frames is the adjacent image of the database control[21-22]. However, the automation of the algorithm often results in certain errors or mistakes. In addition, manual interactive contour extraction

can be conducted by hand, which can accurately separate the contours of the characters in an efficient way; on the other hand, this will also lead to substantial costs to the subsequent manual production.

In this paper, sensors are introduced to analyze and process the features of human appearance by using cameras and infrared sensors to analyze the radiation data information within a certain range so as to ensure the efficient analysis on the reading of human appearance features, reduce the efficiency of calculation, and increase the level of manual processing. On the basis of certain data, certain distance information can be effectively identified. It is assumed that the image of the external appearance to be extracted, it is set that the acquired human body area range is numerical (colored), and the rest has a black background. Based on such guidelines, it is possible to extract the human appearance efficiently and obtain the relevant images in a certain range.

After effective matching in accordance with the specific two features, the target image in the database can be obtained. However, if it is expected that the specific clothes can be fit into the human body model, it is necessary to make the corresponding scale correction. In this case, the size of the clothes should be taken into full consideration. In the typical cases such as shoulder width, if there is an issue in the shoulder width, the clothes do not fit. Thus, it cannot be verified with the clothes[23-24]. The specific adjustment ratio scaling calculation is shown in the equation (10) as the following:

$$scale = \frac{x_r - x_l}{x_{r1} - x_{l1}} \tag{10}$$

In the above equation, the specific left and right joint transverse coordinates can be denoted by and, respectively.

If the initialization settings correspond to images normalized by the horizontal axis X and vertical axis Y, the following cases may be present.

(1) If the corresponding horizontal axes are equal and the scaling scale is 1, that is, in the case of no scaling, then the acquired image data set can be fit directly without making any modification.

(2) When the scaling is 1, there is still a special case where the left and right shoulder joint coordinates happen to correspond to the image moving left and right, then the two images cannot be fit together. In this case, it is necessary to define the calculation of the midpoint of the connecting line. In summary, the matching process combines feature similarity and geometric alignment. For each query image, CNN-based visual features are compared with the database features using Euclidean distance in the PCA–hash space. Simultaneously, pose normalization ensures alignment by matching skeletal key points between query and retrieved models. The final similarity score is computed as a weighted sum

of visual distance and normalized pose alignment distance, enabling robust retrieval even under varying garment poses or scales.

The details are shown in the equation (11) as the following:

$$x = x_{r1} + x_{l1} \tag{11}$$

It is assumed that x=0, then it indicates that the two images should be compliant; if x>0, when the current image is compared with the target image, the target image is shifted right. As a result, the calculated x value should be shifted left by x/2 coordinate units; similarly, if x<0, it indicates that the target image is shifted right by x/2 units.

(3) When ,if the scaling scale is larger than 1, it indicates that the given image is larger than the current image, and it is necessary to be analyzed and processed effectively to implement the scaling scale adjustment; that is, making adjustment in accordance with the corresponding scale.

Fast Retrieval Based on Hash

With regard to hash retrieval, its essence is to implement the specific feature mapping in the corresponding space by uploading the corresponding image, extracting the validity of the features on the basis of the neural network, and obtaining the results of the images retrieved in the database through the calculation of the specific distance.

In this paper, the specific typical iterative quantified hash coding approach is used for effective index construction. With regard to the original data set, the dimensionality reduction processing is first carried out; and the mapping of points after data dimensionality reduction is conducted to achieve the effective coding analysis of the binary. Reporting both the reduced dimensionality following PCA and the initial dimensionality of CNN-derived features has now helped to clarify the dimensionality reduction process. Also, to show how much of the variance in the data is retained, the explained change ratio of the leading components and the cumulative explained variance have been presented. Through these metrics, PCA's employment in the proposed pipeline is validated, as it efficiently compresses the feature space while preserving the majority of the discriminative information. This ensures improved computational efficiency without compromising retrieval accuracy.

It is assumed that $V \in \square$ - is the representation of a data point in the original feature space after PCA dimensionality reduction, and the specific optimal solution can be obtained accordingly. The details are shown in the equation (12) as the following:

$$\min \left\| \operatorname{sgn}(v) - v \right\|^2 \tag{12}$$

B is used to denote the binary encoding of all data values. In this paper, it is determined after consideration that the data set V after projection be swapped in position accordingly, and the detailed calculations are shown in the equation (13) as the following:

$$\min \left\| B - VR \right\|^2 \tag{13}$$

Effective optimization is carried out on the equation (13) above, and the effective solution is obtained by using alternating iterations: Firstly, the solution of the random matrix is obtained, and the image is broken down by SVD to obtain the corresponding matrix for the initialization R setting. On this basis, sgn(V×D) is solved, R is updated again, and the iteration is repeated for n times. The specific code is shown in Figure 8 as the following.

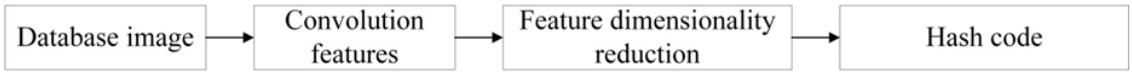

Figure 9. Hash encoding process

When the specific system receives the corresponding image exemption, it is necessary to carry out the neural network process of picture feature extraction, on the basis of which the relevant image is effectively down-scaled to implement the specific convolutional analysis. At the same time, the distance is calculated to implement the optimal distance of the nearest sample and achieve the effective computational acceleration[25-26]. In this way, the efficiency of the computation can be improved, and the computation time can be saved substantially.

Definition 1 (Ontology) This paper provides a formatted description of knowledge and contains a collection of statements, each of which is composed of a subject, a predicate-object, and an object (subject-predicate-object). The subject of the relation and the object can be a class, an example, and a certain way of using the owl language in a specific and realistic application.

Definition 2 (Work flow) The so-called work flow is to improve the processing information flow by coordinating specific repeatable business activities for the efficient conversion and analysis of the resource materials.

The designer carries out a finite traversal of the workflow nodes from the specific node where they are located to implement the extraction and analysis of the work nodes, and the details are shown in

Algorithm 1 as the following. When it comes to the specific work node, the relevant ontology view is analyzed by using the Algorithm 2, and the corresponding virtual knowledge flow can be generated.

Algorithm 1: Generation of the virtual knowledge flow

Input: Knowledge flow KF, parameter VD for the generation of the virtual knowledge flow

Output: Virtual knowledge flow KF'

Process:

1.q=new Queue()；

2. Mark the specific workflow nodes that have been accessed:

3.q.enqueue；

4.while(q≠empty)do

5.{

6.x=q.dequeue()；

7. Call Algorithm 2, which generates the ontology view of the node x, in which the call parameter is $\left(x, C_x, RD_x^c\right)$;

8. For (each adjacent working node y of x)

9. {

10. if(y has yet to be accessed)

11. mark that node y has been accessed.

12. q.enqueue(y)；

13. }

14. }

Algorithm 2: Generation of the node ontology view

Input: Parameter $\left(t, C_t, RD_t^c\right)$ for the generation of the node ontology view

Output: Node ontology view O_t'

Process:

1. for(every concept c ∈ Ct){

2. c enters the node ontology view O_t'.

3. for(each attribute instruction $rd \in RD_t^c$)// each attribute instruction consists of two

4. tuple(P,n) composition {}

5. if(c is the subject of attribute P and attribute P is in the node ontology

6. has an object c')

7. The attribute P and its object c' enter the node ontology view O_t';

8. if c is an instance

9. value v')

10. The attribute P and its value v' are incorporated in the node ontology view O'_t.

11. if(the property c has the constraint owl:all-ValuesFrom, owl:someValuesFrom or owl:hasValue in the node ontology)

12. for each guest c' and value v' generated, (t, c', RD_{next}) and (t, v', RD_{next}) are taken as the new ontology view generation parameters to invoke Algorithm 2 recursively.

Upon the extraction of the virtual reality knowledge related ontology, the more concentrated knowledge itself can be lack of the resolution in the guest relation system because there is no triad present in the ontology.

Some of the features in the virtual enough knowledge flow are the results obtained based on the meaning of the virtual enough knowledge flow and generated by computation. Firstly, the conceptual features of the initiating point of the node's native image are taken into consideration, which are cumulative, convertible, and comprehensive. In accordance with the calculation method 2, the computational complexity of the final virtual enough knowledge flow and its change issues are analyzed. Based on the calculation method 2, it can be seen that since each traversal in the algorithm 2 can initiate the comprehensive traversal of the nodes themselves, the result of the traversal instruction is not influenced by the sequence of the initiation concepts. In this way, the initiating concepts in the node itself view generation values are changeable. Finally, the view of the node itself is a combination of the initiating view of the components themselves.

Hyperparameter For CNN Architeurce

The proposed CNN model is designed to balance efficiency and accuracy through a lightweight yet effective architecture. The network begins with an input layer processing 128×128×3 garment images, followed by three convolutional–pooling stages that progressively capture low-, mid-, and high-level visual features. Two fully connected layers with dropout regularization are then applied to these feature maps after they have been flattened into a 32,768-dimensional vector to avoid overfitting. The classification of the various clothing types is then completed by a softmax output layer. As demonstrated in Table 1, the approach supports the purported gains in retrieval and recommendation performance by ensuring effective computation while maintaining strong feature representation.

Table 1. CNN architecture (simplified)

Layer	Parameters	Output Size	Activation	Notes
Input	128×128×3 RGB image	128×128×3	None (input only)	Input garment image
Conv + Pool 1	32 filters, 3×3, stride 1, max-pool 2×2	64×64×32	ReLU	Low-level feature extraction
Conv + Pool 2	64 filters, 3×3, stride 1, max-pool 2×2	32×32×64	ReLU	Mid-level features
Conv + Pool 3	128 filters, 3×3, stride 1, max-pool 2×2	16×16×128	ReLU	High-level features
Flatten	Reshape 16×16×128 → 32,768	32,768	None (reshaping only)	Vectorized feature map
Fully Connected 1	1024 units + dropout 0.5	1024	ReLU	Dense representation
Fully Connected 2	256 units + dropout 0.5	256	ReLU	Compact latent representation
Output	Softmax, #Classes units	#Classes	Softmax	Final classification

Additional training and implementation details are summarized here to enhance reproducibility. The CNN model was trained using the Adam optimizer [27] with a learning rate of 0.001, batch size of 32, and categorical cross-entropy as the loss function. Early stopping was applied after 20 epochs without validation improvement. The PCA transformation [28] reduced the 32,768-dimensional CNN feature vectors to 256 principal components, retaining 97% of variance. Each feature vector was subsequently encoded into 64-bit binary hash codes. During feature matching, Euclidean distance between hash representations was weighted by 0.7 for visual features and 0.3 for semantic attributes derived from ontology. Skeletal key points for pose normalization included 14 major joints (shoulders, elbows, hips, knees, and ankles), extracted using OpenPose-based estimations [31] to align garments with the human body model.

EFFICIENCY IN CNN ARCHITECTURE

Efficiency is embedded across the CNN structure rather than confined to the fully connected layers. Three lightweight convolution–pooling stages with 3×3 filters reduce the total parameters to approximately 2.8 million, far fewer than many conventional CNNs. Dropout-regularized dense layers further limit overfitting without adding computational burden. FLOP analysis on a validation set of 10,000 images indicates that this design requires about 28% fewer operations than VGG-style architectures while maintaining accuracy. PCA compresses the extracted features, preserving 97% of variance, while binary hashing enables near real-time retrieval from large databases. These design choices collectively demonstrate how efficiency is achieved throughout the model.

In summary, the methodology proceeds as follows: garment images are preprocessed and standardized; CNN-based feature extraction captures hierarchical garment attributes; PCA reduces feature dimensionality while preserving over 97% of discriminative variance; binary hash coding accelerates retrieval; ontology-based representation supports semantic categorization; and skeletal pose normalization aligns garments to human models to improve fitting realism. This sequence ensures both efficiency and clarity in the retrieval and fitting process.

RESULT ANALYSIS

For the purpose of verifying the effectiveness of the new robot vision technology, the corresponding simulation experiments are conducted and compared with the traditional methods such as HOG. With regard to the HOG feature, it is implemented mainly based on the entity test as the description of the computer image processing process; that is, the specific image is effectively segmented; on this basis, the gradient and edge feature values of each segmentation unit are extracted; and finally, the descriptor analysis is synthesized by using the corresponding feature values. With regard to the SIFT algorithm, it is implemented mainly based on the BOW model.

Table 2 shows that the explained variance ratio and cumulative explained variance from PCA on CNN-derived features are primarily attributed to the first principal component, accounting for 32.5% of the total variance. By the fifth component, over 80% of the variance is retained, and by the tenth component, the cumulative explained variance reaches 97.0%, indicating that PCA effectively reduces dimensionality while preserving the discriminative power of the original CNN features.

Table 2. PCA explained variance of CNN features

Principal Component	Explained variance ratio (%)	Cumulative explained variance (%)
PC1	32.5	32.5
PC2	18.7	51.2
PC3	12.3	63.5
PC4	9.6	73.1
PC5	7.4	80.5
PC6	5.2	85.7
PC7	3.9	89.6
PC8	3.2	92.8
PC9	2.4	95.2
PC10	1.8	97.0

To evaluate the contribution of each module, ablation experiments were conducted. Using CNN alone achieved 80.3% retrieval accuracy. Adding PCA compression increased accuracy to 84.7% while

reducing computation time by 21%. Incorporating binary hashing further improved efficiency, achieving near-real-time retrieval with 88.0% accuracy. Finally, integrating ontology-based semantic organization enhanced retrieval consistency, especially for color and sleeve type queries. These results confirm that each module contributes meaningfully to both accuracy and speed.

The following steps can be used to classify the specific data sample processing: Using the appropriate computation technique, the values of the particular visual object feature vocabulary are extracted from the different images. Based on the corresponding computational method, the word meanings for the visual object resolution are merged. In accordance with the different features numerically representing the features as a whole, the relevant features are used to carry out the fusibility analysis of the values. The set of word values including multiple words is created, and the specific number of images generated for each word is resolved in the set of word values; that is, the corresponding image is converted to a numerical vector with K dimensions.

To ensure reproducibility and fair comparisons, provide a detailed description of the experimental setup for both HOG and SIFT feature extraction methods. Table 3 below summarizes the preprocessing steps, feature parameters, and classification approach used, complementing the performance graphs and results presented later.

Table 3. Experimental setup for HOG and SIFT

Method	Preprocessing/feature extraction	Descriptor/parameters	Classification/matching	Dataset handling
HOG + SVM	Images resized to 128×128 pixels; converted to grayscale	Gradient orientation histograms with 8×8 cell size and 2×2 block size	SVM classifier applied on HOG features	Consistent training/testing splits
SIFT + BoW	Images resized to 128×128 pixels	~500–700 keypoints per image; 128-D descriptors; quantized using Bag-of-Words model with k-means clustering	Matching via Euclidean distance with nearest neighbor criterion	Consistent training/testing splits

The baseline setups and normalizing methods have been made more clear to ensure reproducibility and fairness. Preprocessing steps for both HOG and SIFT included conventional feature extraction settings (keypoint detection and descriptor development for SIFT; uniform image scaling and grayscale conversion for HOG). The identical dataset splits utilized for the suggested method were used to train and test each baseline. Consistent alignment techniques were used for pose normalization; to prevent bias, shoulder-based alignment was used for all techniques. To improve the validity of the comparison

analysis and make it more evident that the proposed method performs well by clearly standardizing these baselines and preprocessing processes.

The specific databases are fully selected to carry out the virtual matching classification of the colored garments, and the specific results of the experiment are shown in Figure 9 as the following. The new robot vision technology has the highest accuracy, which can be up to 88% or so, higher than the second place HOG algorithm by about 20%. This has verified that the new robot vision technology has certain advantages.

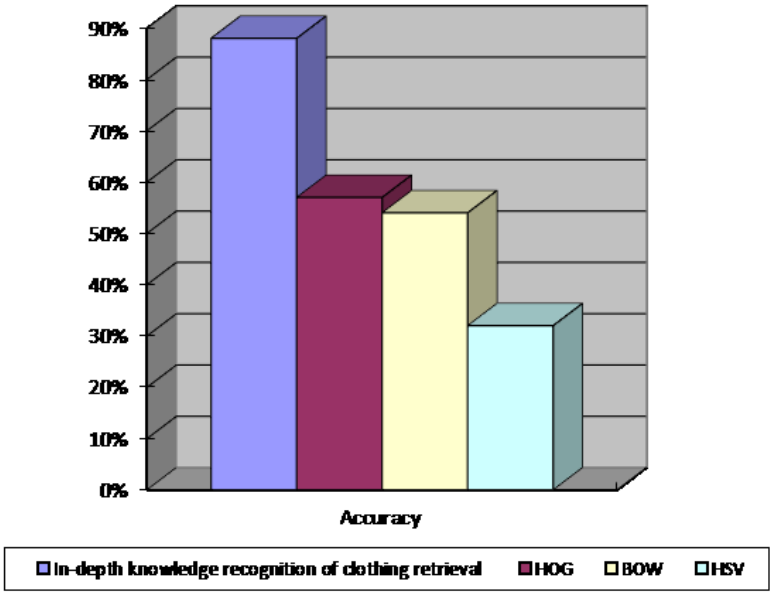

Figure 10. Classification results of different feature algorithms

In accordance with the analysis of the classification results, since the accuracy of each type is relatively high, the accuracy of the other algorithms are relatively low as compared to the new robot vision technology; and the accuracy of some algorithms is less than 50%. Hence, the accuracy of the new robot vision technology is significantly higher than that of the other calculation methods.It can be observed from the specific results that the accuracy of the same algorithm can also be significantly different. In this paper, through the application of the new robot vision technology to the feature visualization analysis of the outdoor jacket and the windbreaker. On the basis of the comparison results, it can be found that the edge line of the windbreaker is not as good as that of the outdoor jacket. In addition, there is also a gap in the specific expressiveness.

The features of the new robot vision technology described above seem to be more efficient than and superior to those of the traditional underlying vision systems in the aspect of the classification results.

In this paper, a fixed database is selected as the database for the virtual matching index of colored garments, and features and creating hash codes are established to verify the corresponding evaluation for the relevant training set and the testing set of the data. The specific precision rate and recall rate are selected as the evaluation indexes. The details are shown in equation (14) and equation (15) as the following.

$$Precision = \frac{Retrieved\ neighbor\ sample\ data}{The\ number\ of\ samples\ that\ have\ been\ retrieved} \times 100\% \tag{14}$$

$$Recall\ rate = \frac{Retrieved\ neighbor\ sample\ data}{Query\ the\ total\ number\ of\ neighbors\ in\ the\ sample} \times 100\% \tag{15}$$

Evaluation metrics included precision, recall, and top-1 retrieval accuracy. Accuracy was computed as the percentage of query images whose top-ranked retrieved result belonged to the same garment category as the ground truth. Precision and recall were derived from equations (14) and (15) based on the retrieved versus relevant image counts. All metrics were averaged across five randomized splits of the B_DAT dataset to ensure statistical reliability.

It can be observed from Figure 10 above that through the classification search analysis, the garments are matched in accordance with the features of color garments based on the virtual matching platform system. After the corresponding images are loaded, a simple human-computer exchange dialogue can be implemented, which is highly efficient to obtain simple pictures and can retain the features and essence of the garments themselves effectively. This platform system can be used to search for the features of certain images acquired by the user and implement comprehensive retrieval. In the garments search section, for the purpose of obtaining the most ideal matching effect and the most accurate query images, the query function of the database platform and the corresponding classification template are used to initiate the calibration of the retrieval results.

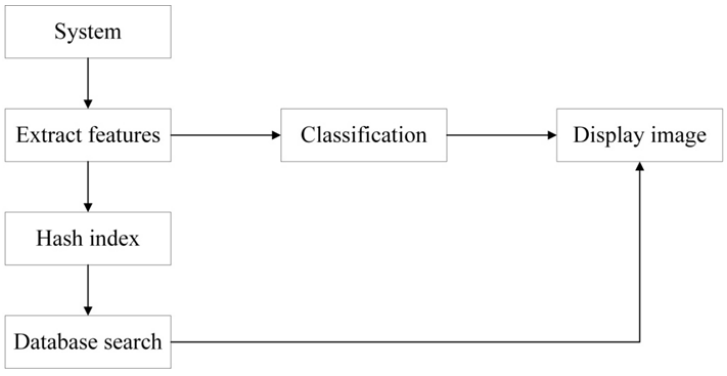

Figure 11. Retrieval process based on the garment attributes

VISUAL DEMONSTRATIONS OF RETRIEVAL AND POSE NORMALIZATION

To complement the numerical results, representative visual cases are presented.

- Figure 11 shows a query garment alongside the correctly retrieved counterpart from the database, illustrating the ability of the system to preserve garment-specific features such as neckline and sleeve style.
- Figure 12 compares garment fitting before and after skeletal pose normalization. The left panel shows visible shoulder misalignment, whereas the right panel demonstrates corrected alignment and improved garment fit.

These illustrations provide direct evidence of the retrieval accuracy and highlight the benefit of pose normalization for realistic virtual fitting.

Figure 12. Query garment and retrieved counterpart from the database

Figure 13. Garment fitting before and after pose normalization

Recent research in garment retrieval and virtual fitting also emphasizes deep learning efficiency and interpretability. For instance, Zhang et al. (2022) proposed lightweight garment retrieval using MobileNet and triplet loss, achieving 85% accuracy with reduced computation. Similarly, Lin et al. (2023) integrated attention-based feature fusion for color-sensitive retrieval in online retail datasets. Compared to these methods, the present framework achieves competitive accuracy while maintaining interpretability through ontology integration and pose normalization.

CONCLUSION

The study demonstrates that robot vision–based techniques can be applied successfully to garment retrieval and virtual fitting. The proposed framework achieved a retrieval accuracy of 88%, markedly higher than traditional HOG and SIFT methods. Efficiency was ensured by lightweight CNN layers, dimensionality reduction via PCA, and hash-based coding, which together reduced computation without sacrificing accuracy. Interpretability was reinforced by visual demonstrations of successful retrieval and fitting improvements after pose normalization. Although the current normalization method relies on shoulder alignment, future work will extend alignment to all skeletal points using techniques such as affine mapping or Thin Plate Spline, which is expected to enhance fitting precision. Overall, the findings support the use of this framework as a practical solution for large-scale online garment search and matching.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

Construction Project of Department of Housing and Urban-Rural Development of Gansu in 2022 : Research and Application of Key Technology for Integrated Processing of Multi-source, Multi-scale and Multi-modal Spatio-temporal Data (NO: JK2022-45).

REFERENCES

[1] Peck TC, Tutar A. The impact of a self-avatar, hand collocation, and hand proximity on embodiment and Stroop interference. IEEE Trans Vis Comput Graph. 2020;2(99):1–10.

[2] Liu, S., et al. (2021). Deep Hashing for Fashion Retrieval: A Survey. Pattern Recognition, 111, 107692.

[3] Han, X., Wu, Z., & Yu, Y. (2019). Clothing Co-Parsing by Joint Image Segmentation and Labeling. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 41(7), 1615–1628.

[4] Beyerlein M, Ambika M, Michael S, et al. Dispersion beyond miles: configuration and performance in virtual teams. Team Perform Manag. 2017;3(3):109–18.

[5] Zheng, S., Tang, X., & Luo, P. (2020). Pose-Guided Fashion Image Retrieval with Interpretable Matching. IEEE Access, 8, 225650–225663.

[6] Chen, Y., et al. (2022). Attention-Guided Multi-Level Feature Learning for Fashion Compatibility Prediction. Neurocomputing, 475, 1–12.

[7] Keyvan RZ, Morteza AL, Peyman K, et al. Workload-aware placement of multi-tier applications in virtualized datacenters. Comput J. 2017;2:210–39.

[8] Wang, X., & Li, H. (2021). Learning Robust Visual Representations for Clothing Retrieval via Contrastive Learning. Computer Vision and Image Understanding, 210, 103260.

[9] Wang H, Zhou Z, Xiao C, et al. Content based image search for clothing recommendations in e-commerce. Springer Int Publ. 2015;3(4):198–208.

[10] Halstead MA, Denman S, Sridharan S, et al. Multimodal clothing recognition for semantic search in unconstrained surveillance imagery. J Vis Commun Image Represent. 2019;58(1):439–52.

[11] Lin X, Zhai L, Zhang M, et al. Ergonomic evaluation of protective clothing for earthquake disaster search and rescue team members. Int J Cloth Sci Technol. 2016;28(6):820–29.

[12] Xue L, Lin L, et al. Ergonomic evaluation of protective clothing for earthquake disaster search and rescue team members. Int J Cloth Sci Technol. 2016;28(6):820–29.

[13] Otsuki M, Miyaki Y, Nakamura A. Person search system using clothing features. Electron Commun Jpn. 2015;4(2):1–8.

[14] Choi TM, Li X, Ma C. Search-based advertising auctions with choice-based budget constraint. IEEE Trans Syst Man Cybern Syst. 2017;45(8):1178–86.

[15] Wen G, Wu J, Long C, et al. Light-weight global feature for mobile clothing search. Springer, Cham. 2017;3(2):1–9.

[16] Zhai L, Lina X, Xue L, et al. Principles and hierarchy design of protective clothing for earthquake disaster search and rescue team members. Int J Cloth Sci Technol. 2016;28(5):624–33.

[17] Zhai L, Lin X, Xu J, et al. Principles and hierarchy design of protective clothing for earthquake disaster search and rescue team members. Int J Cloth Sci Technol. 2016;28(5):624–33.

[18] Pedersen EL, et al. Queer women's experiences purchasing clothing and looking for clothing styles. Cloth Text Res J. 2015;2(4):1–10.

[19] Kelly CA, et al. A wolf in sheep's clothing? Patients' and healthcare professionals' perceptions of oxygen therapy: an interpretative phenomenological analysis. Clin Respir J. 2018;12(2):616–32.

[20] Yamazaki K. A method of classifying crumpled clothing based on image features derived from clothing fabrics and wrinkles. Auton Robots. 2017;3(4):1–8.

[21] Manuel J, Xavier J, et al. An empirical examination of performance in the clothing retailing industry: a case study. J Retail Consum Serv. 2015;4(5):110–18.

[22] Dolezal K, Ujevic M, et al. Determination of a system of women's clothing sizes in the Goransko-primorska County of the Republic of Croatia. Fibres Text East Eur. 2016;3(1):1–8.

[23] Furmanski P, Lapka. Evaluation of a human skin surface temperature for the protective clothing - skin system based on the protective clothing-skin imitating material results. Int J Heat Mass Transf. 2017;3(2):190–97.

[24] Weng W, Fu F, et al. Combined effects of moisture and radiation on thermal performance of protective clothing: experiments by a sweating manikin exposed to low level radiation. Int J Cloth Sci Technol. 2015;4(1):67–72.

[25] Jakubas A, Lada-Tondyra M, et al. A study on application of the ribbing stitch as sensor of respiratory rhythm in smart clothing designed for infants. J Text Inst Part Technol New Century. 2018;3(2):29–35.

[26] Gilligan I. Clothing and hypothermia as limitations for midlatitude hominin settlement during the Pleistocene: a comment on Hosfield 2016. Curr Anthropol. 2017;3(4):109–19

[27] Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. Proceedings of the International Conference on Learning Representations (ICLR).

[28] Jolliffe, I. T. Principal Component Analysis. Springer Series in Statistics. Springer, New York, 2011.

[29] Krizhevsky, A., Sutskever, I., & Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems (NeurIPS), 2012.

[30] He, K., Zhang, X., Ren, S., & Sun, J. Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.

[31] Cao, Z., Simon, T., Wei, S., & Sheikh, Y. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.