

## RECOGNITION OF HUMAN POSE FROM IMAGES BASED ON GRAPH SPECTRA

A. A. Zakharov<sup>a\*</sup>, A. E. Barinov<sup>a</sup>, A. L. Zhiznyakov<sup>a</sup>

<sup>a</sup> Murom Institut Vladimir State University, CAD Department, , 602264, Orlovskaya 23, Murom, Russian Federation, aa-zakharov@ya.ru

Commission VI, WG VI/4

**KEY WORDS:** Image Recognition, Human Pose, Spectral Graph Matching

### ABSTRACT:

Recognition of human pose is an actual problem in computer vision. To increase the reliability of the recognition it is proposed to use structured information in the form of graphs. The spectrum of graphs is applied for the comparison of the structures. Image skeletonization is used to construct graphs. Line segments are the nodes of the graph. The end point of line segments are the edges of the graph. The angles between adjacent segments are used to set the weights of the adjacency matrix. The Laplacian matrix is used to generate the spectrum graph. The algorithm consists of the following steps. The graph on the basis of the vectorized image is constructed. The angles between the adjacent segments are calculated. The Laplacian matrix on the basis of the linear graph is calculated. The eigenvalues and eigenvectors of the Laplacian matrix are calculated. The spectral matrix is calculated using its eigenvalues and eigenvectors of the Laplacian matrix. The principal component method is used for the data representation in the space of smaller dimensions. The results of the algorithm are given.

### 1. INTRODUCTION

Recognition of human pose is an actual problem in many areas: security, human-machine interface, medicine, transport, simulators. Marker and markerless approaches used for motion capture (Field, M., Stirling, D. A., Naghdy, F., Pan, Z., 2009). Marker motion capture system are optical passive, optical active, magnetic, mechanical, inertial, acoustic. Common disadvantage of marker motion capture systems is the impossible to use the sensors when viewed real scenes. This feature limits the use of marker techniques. Thus, the development of methods and software tools to capture and recognition of human pose using markerless methods is an important task. Questions of pose recognition based on optical methods insufficiently studied. Using computer vision is limited by the following factors: the small size of the scenes, the presence of noise and optical effects, varying textured background, overlapping objects.

To increase the reliability of the recognition it is proposed to use structured information in the form of graphs. The spectrum of graphs is applied for the comparison of the structures. The advantage of structural methods is that they give an opportunity to analyze a big set of elements on the basis of a small quantity of simple components and rules of forming the graphic model. Also structural methods allow to describe the characteristics of the object excluding its reference to another class that increases the reliability of the recognition.

### 2. SURVEY OF ALGORITHMS OF HUMAN POSE RECOVERY

Taxonomy of human pose recovery consists of the following: appearance, viewpoint, spatial relations, temporal relations and behavior (Perez-Sala, X., Escalera, S., Angulo C., Gonzalez, J,

2014). Appearance can be defined as image evidences related to human body and its possible poses. The following features are used to describe models: silhouettes and contours, intensity, color, texture, depth, optical flow, logical relations. The following classifiers are used to detect human pose: discriminative classifiers, generative classifiers, templates, interest points.

Viewpoint determines the number of images to pose recovery. Viewpoint approach is discrete and continuous. Discrete approach uses single image. Continuous approach uses multiple images.

Spatial models encode the configuration of the human body (Johnson, S., Everingham, M., 2011). Spatial models are of two types: hard (skeleton, bone lengths), a soft way (pictorial structures, grammars).

Temporal models use a sequence of images to reduce the search space. Tracking is applied to ensure the coherence among poses over the time. Motion model is used to describe a variety of human actions (walking, running).

The term behavior includes actions and gestures. Recognition of actions is based on the images.

We use single image to recognize human pose. The spatial model is described as a skeleton. We use region-based approach to extract the skeleton of the image. We segment the image for the shock graph and shape matching (Thome, N., Merad, D., Miguët, S., 2008). The comparison of the graphs consists in searching the compliances between the structures of the graphs on the basis of more or less strict restrictions.

We propose to use the comparison methods on the basis of the spectral graph theory. The spectrum of the graph represents the vector of the eigenvalues of the connectivity matrix of the graph arranged in the decreasing or increasing order. The spectral methods are based on the following property: the

\* Corresponding author

eigenvalues and the eigenvectors of the connectivity matrix of the graph are invariant concerning the permutation of the vertexes in a matrix (Chung, F. R. K., 1997). Therefore, if the two graphs are isomorphic, their connectivity matrices will have identical eigenvalues and eigenvectors. However, the opposite statement is not always right: there are graphs with different structures but with the identical spectrum (cospectral graphs).

The spectral representation of the graphs has the following advantages: it is invariant with respect to the order of the vertexes; it contains the information on the structure of the graph; it describes the structure of the graph in the form of numerical values, which is convenient for calculations; the calculation problem of the eigenvalues is well studied and is solved for the polynomial time (Zvetkovic, D., 2012).

Index based on spectrum graphs used to recognize human pose (Thome, N., Merad, D., Miguet, S., 2008). Topological signature vector is calculated as the sum of eigenvalues of the matrix (Shokoufandeh, A., Dickinson, S.J., Siddiqi, K., Zucker, S.W., 1999). Hermitian matrix is used to represent the human pose in the article (Haseeb, M., Hancock, E. R., 2012.). The angles between the joints are used to obtain the spectrum graphs.

We use the Laplacian matrix to obtain a spectrum graph. The principal component method is used for the data representation in the space of smaller dimensions.

### 3. RECOGNITION OF HUMAN POSE BASED ON GRAPH SPECTRA

#### 3.1 A human pose model based on the graphs

The set of the segments located on the plane is obtained as a result of the image skeletonization. A segment is characterized by the coordinates of the end points. In the resulting graph the edges are represented by the segments and the vertices characterize the value of the angle between the edges. But it is difficult to represent the similar graph in the matrix form as it is impossible to reflect the weight of the vertices. Therefore, for the purpose of convenience it is proposed to transform the boundary representation into a linear graph, in which the vertices are represented by linear segments and edges connect adjacent segments of the graph and characterize the angle between them (Fig. 1). A Laplacian matrix is constructed using the values of the angles between adjacent segments.

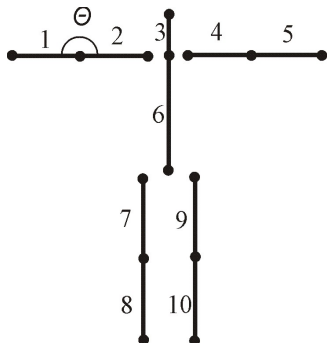


Figure 1. A Human Pose model based on the graphs

The adjacency matrix  $A$  of a weighted graph  $G=(V,E)$  is a  $|V| \times |V|$  matrix whose element with row index  $i$  and column index  $j$  is

$$A(i,j) = \begin{cases} w_{ij}, & \text{if } (i,j) \in E, \\ 0, & \text{otherwise.} \end{cases}$$

The degree matrix of a weighted graph  $G$  will be denoted, and is the diagonal matrix such that

$$D(i,i) = \sum_j A(i,j).$$

The matrix  $L=D-A$  is called the Laplacian matrix of graph  $G$ .

The set of the segments located on the plane is obtained as a result of the image skeletonization. A segment is characterized by the coordinates of the end points  $v_i=(x_i,y_i)$ .

Let us assume that there are two adjacent segments  $e_i$  and  $e_j$  with the nodes  $v_{i1}$ ,  $v_{i2}$  and  $v_{j1}$ ,  $v_{j2}$  respectively. The nearby nodes of segments in this case are  $v_{i2}$  and  $v_{j1}$ . The angle between the vectors, on which the segments  $e_i$  and  $e_j$  are situated, is calculated using the following formula:

$$\theta_{ij} = \arccos \left( \frac{v_{i2}v_{i1} \cdot v_{j1}v_{j2}}{|v_{i2}v_{i1}| \cdot |v_{j1}v_{j2}|} \right).$$

The adjacency matrix with the index of lines  $i$  and the index of columns  $j$  has the following form:

$$A_{ij} = e^{\theta_{ij}},$$

where  $\theta_{ij}$  – the angle between the vectors  $\overline{v_{i2}v_{i1}}$ ,  $\overline{v_{j1}v_{j2}}$ .

#### 3.2 Vector space embeddings of graphs via graph matching

Graphs from a some domain are mapped to vector spaces of predefined prototype graphs (Riesen, K., Jiang, X., Bunke, H., 2010).

**Definition (Graph Embedding).** Let us assume a graph domain  $H$  is given. If  $T=\{g_1, \dots, g_N\} \subseteq H$  is a training set with  $N$  graphs and  $P=\{p_1, \dots, p_n\} \subseteq T$  is a prototype set with  $n$  graphs, the mapping  $\varphi_n^P: H \rightarrow R^n$  is defined as the function  $\varphi_n^P(g)=(d(g,p_1), \dots, d(g,p_n))$  where  $d(g,p_i)$  is any graph dissimilarity measure between graph  $g$  and the  $i$ -th prototype graph.

We use for graph embedding a symmetric polynomial. A symmetric polynomial is a polynomial  $S(x_1, x_2, \dots, x_n)$  in  $n$  variables, such that if any of the variables are interchanged, the same polynomial is obtained. For a set of variables  $x_1, x_2, \dots, x_n$  the elementary symmetric polynomials can be defined as:

$$S_n(x_1, x_2, \dots, x_n) = \prod_{i=1}^n x_i.$$

A symmetric polynomial is invariant under permutation of the variable indices (Wilson, R. C., Hancock, E. R., & Luo, B., 2005).

The elementary symmetric polynomials can be computed from using the Newton-Girard formula

$$S_r = \frac{(-1)^{r+1}}{r} \sum_{k=1}^r (-1)^{k+r} P_r S_{r-k},$$

where  $P_n(x_1, x_2, \dots, x_n) = \sum_{i=1}^n x_i^n$  – the power symmetric polynomial functions.

The eigendecomposition of a Laplacian matrix is  $L = U\Lambda U^T$ , where  $\Lambda = \text{diag}(\lambda_1, \lambda_1, \dots, \lambda_{|V|})$  is the diagonal matrix with the eigenvalues in increasing,  $U = (u_1, u_2, \dots, u_{|V|})$  is the matrix with the ordered eigenvectors as columns. The Laplacian spectrum is the set of eigenvalues  $(\lambda_1, \lambda_1, \dots, \lambda_{|V|})$ .

The spectra and eigenvectors are used to compare graphs.

We construct a complex spectral matrix

$$\Psi = (\sqrt{\lambda_1} U_1 | \sqrt{\lambda_2} U_2 | \dots | \sqrt{\lambda_n} U_n).$$

Polynomial coefficients are calculated based on the Newton-Girard and variable of matrix  $\Psi$ :  $S_1(\psi_{1,i}, \dots, \psi_{n,i}), \dots,$

$$S_n(\psi_{1,i}, \dots, \psi_{n,i}).$$

Thus, the structures of each pose defined by the vector  $S_n$ . Space must be reduced to compare vectors. The principal component is used for this.

### 3.3 An algorithm of human pose recognition based on graph spectra

Step 1. The graph on the basis of the vectorized image is constructed. The angles between the adjacent segments are calculated.

Step 2. The Laplacian matrix on the basis of the linear graph is calculated:

$$L_{ij} = e^{\theta_{ij}}.$$

Step 3: The eigenvalues and eigenvectors of the Laplacian matrix are calculated.

$$L = U\Lambda U^T.$$

Step 4. The spectral matrix is calculated using its eigenvalues and eigenvectors of the Laplacian matrix

$$\Psi = (\sqrt{\lambda_1} U_1 | \sqrt{\lambda_2} U_2 | \dots | \sqrt{\lambda_n} U_n).$$

Step 5. Elementary symmetric polynomials are calculated on the basis of the  $\Psi$  matrix. The columns form a vector which characterizes the structure of the model.

Step 6. The principal component method is used for the data representation in the space of smaller dimensions.

Step 7. The method of K-means is applied to clustering the points.

### 3.4 The results of the algorithm operation

Studies of the algorithm were performed using 30 images of human poses. As test example, the image in Figure 2 was used.

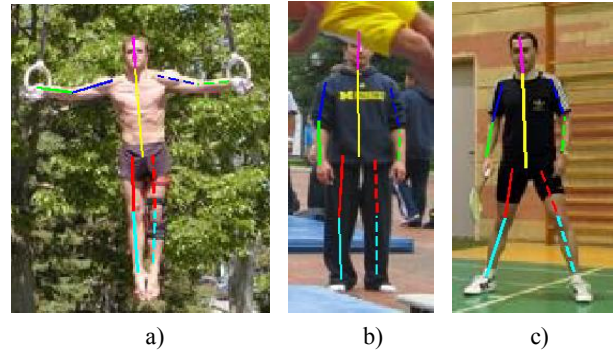


Figure 2. Example of human poses: a)pose 1, b)pose 2, c)pose 3

The feature sets are used for the data representation in the 2D-space by means of the principal component analysis. The method of K-means is applied to clustering the points in the 2D-space. The result of recognition of views is shown in Figure 3.

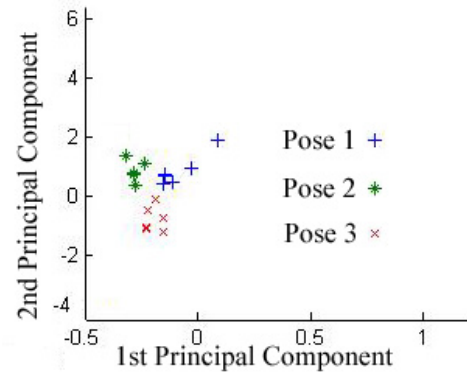


Figure 3. The data representation in 2D-dimensional space using the method of principal components

## 4. CONCLUSIONS

In this paper we proposed an approach to the recognition human pose using the spectral graph theory. The model of the image description is offered using the segments integrated in the graph. The information on the angle between them is used to create the graph. For obtaining the spectrum of the graph the Laplacian matrix is used. The results of the research showed that the proposed method will recognize the main types of the human poses. However, examples of simple poses are presented. Overlapping objects in the scene are often. Such human poses is difficult to recognize on the basis of single image. Further research will focus on the recognition of human pose by using multiple cameras on the basis of the presented approach.

## ACKNOWLEDGEMENTS

This work was supported by the RFBR grant 15-07-01612, project number 2918 in the framework of the basic tasks of the state of the Russian Ministry of Education.

## REFERENCES

Chung, F. R. K., 1997. Spectral graph theory, AMS, p. 207.

Field, M., Stirling, D. A., Naghdy, F., Pan, Z., 2009, Motion capture in robotics review, *IEEE International Conference on Control and Automation*, IEEE, Christchurch, New Zealand, pp. 1697-1702.

Haseeb, M., Hancock, E. R., 2012. Unsupervised Clustering of Human Pose using Spectral Embedding. In *Proceedings of Structural, Syntactic, and Statistical Pattern Recognition – Joint IAPR International Workshop, SSPR/SPR*, Berlin, Heidelberg. Springer-Verlag, pp. 467–473.

Johnson, S., Everingham, M., 2011. Clustered Pose and Nonlinear Appearance Models for Human Pose Estimation. In *Proceedings of the 21st British Machine Vision Conference (BMVC2010)*, pp. 1-11.

Perez-Sala, X., Escalera, S., Angulo C., Gonzalez, J., 2014. A Survey on Model Based Approaches for 2D and 3D Visual Human Pose Recovery. *Sensors*, 14(3), pp. 4189-4210.

Riesen, K., Jiang, X., Bunke, H., 2010. Exact and Inexact Graph Matching: Methodology and Applications. *Managing and Mining Graph Data Advances in Database Systems* Vol. 40, pp. 217-247.

Shokoufandeh, A., Dickinson, S.J., Siddiqi, K., Zucker, S.W., 1999. Indexing using a spectral encoding of topological structure. *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, Vol. 2, pp. 491-497.

Thome, N., Merad, D., Miguet, S., 2008. Learning articulated appearance models for tracking humans: a spectral graph matching approach. *Signal Processing: Image Communication*, Vol. 23(10), pp. 769–787.

Wilson, R. C., Hancock, E. R., & Luo, B., 2005. Pattern Vectors from Algebraic Graph Theory. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(7), pp. 1112–1124.

Zvetkovic, D., 2012. Spectral recognition of graphs. *Yugoslav journal of operations research*, no. 2, pp.145-161.