3D volume localization using miniatures

Luka Šajn[1], Miroslav Radojevic[2] and Tomaž Dobravec[1]
[1]Faculty of Computer and Information Science
University of Ljubljana, Slovenija
[2]Biomedical Imaging Group of Erasmus MC
University Medical Center Rotterdam, Netherlands

***Abstract*—The prediction of the position of a given volume sample in a full body atlas, also known as a volume localization, is a part of an initial stage of image retrieval in most of the dedicated CAD systems. In this paper we present two methods for volume localization, namely histogram matching and classifier regression. Since the histogram matching method ignores the spatial orientation, it is used when the orientation of the volume cubes are not the same. On the other hand the classifier regression is much faster and can be used as a quick estimation and as a tool to reduce the scope of the initial problem. Both presented methods were tested on a dataset with 3962 volumes of a human body atlas. The accuracy and the speed of execution was compared and is presented in this paper.**

I. Introduction

3D volume localization problem is a problem of estimating the atlas coordinates of a given sample 3D volume (i.e. object, miniature). This is usually done by comparing the characteristics of a sample volume with the characteristics of other volumes with known coordinates. The coordinates of the most similar volume are returned as a result, as shown in Fig. 1.

To achieve this, one can use the method which compares the sample volume with the other volumes from the dataset and the coordinates are estimated as the interpolation of the coordinates of the most similar volumes, weighted by the degree of similarity. Another approach attempts to estimate the coordinates using classifier regression. Part of the problem is to find the matching volumes - this task can be viewed as the contents based image retrieval (CBIR) problem. One of the models to solve the CBIR is to treat the images as collections or histograms of the features [2] and compare the distances (dissimilarities) between the feature vectors. The resulting feature is the one with the lower distances, hence higher similarity. Potential features used for grey-level images can be the very pixels, features that correspond to human visual perception, texture features such as global descriptors or Gabor features, or local descriptors such as SIFT or corner detectors [2]. Shape features such as Fourier descriptors, moment invariants and finite element models were surveyed in [12], graph based shape features were presented in [1]. A method for retrieving 3D datasets based on Local Binary Patterns (LBP) features [4] was introduced in [11] and compared with features such as 3D Wavelet Transforms and 3D co-occurence matrices. In this paper we use and compare two methods for the coordinate prediction, namely the histogram matching, which is presented in Section II-A, and the classifier regression, presented in Section II-B. The results of the comparison are given in Section III. We conclude the paper in Section IV.

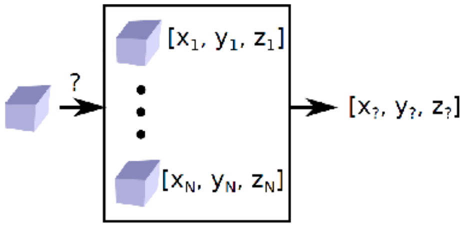


Fig. 1. Estimating the coordinates of the most similar volume

II. Methods

Two methods for the coordinate prediction are used: histogram matching and the classifier regression.

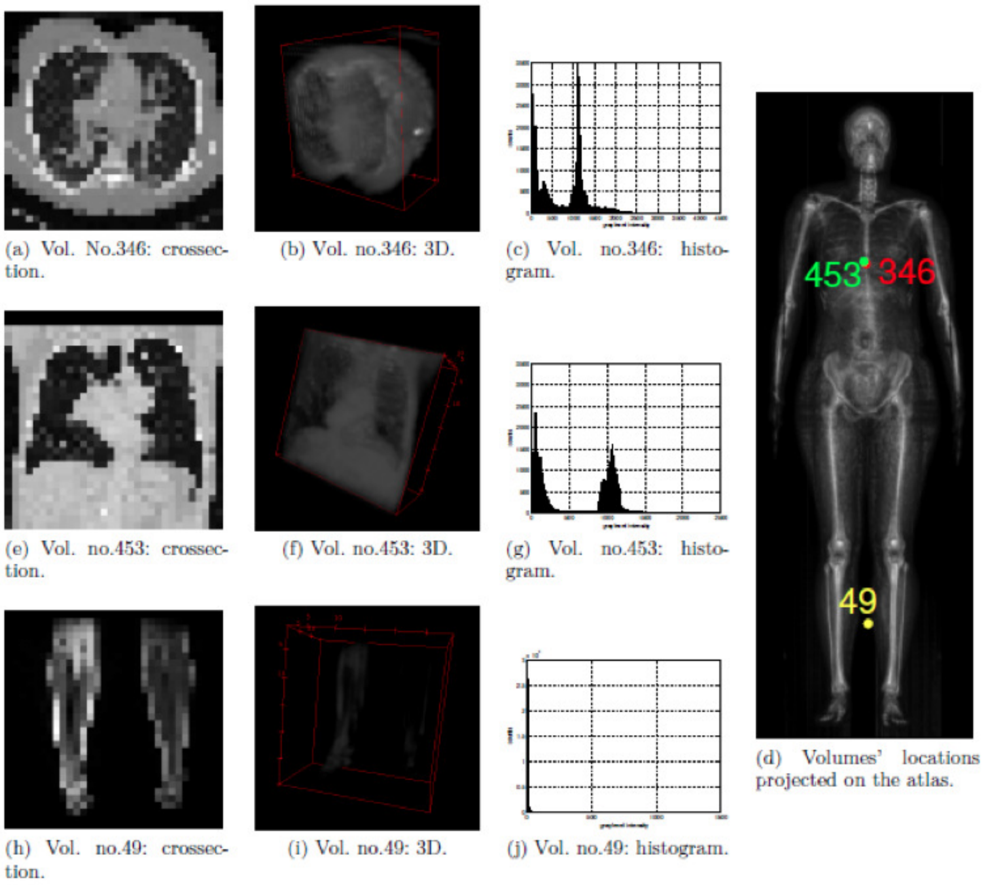
A. Histogram matching

The orientation of the sample volume is usually not the same as the orientation of the volumes in the dataset. Hence, it is necessary either to encounter for their proper alignment, or to use a measure independent to the orientation of the volumes. Such measures include histogram, rotation invariant image moments and local features. The first method presented in this paper compares the histogram of the samples with the histograms of all the other miniatures. The drawback of using histograms is the loss of spatial intensity distribution information. Some other rotation invariant feature extraction method such as SIFT or LBP in 3D [4] would be potentially suitable to overcome the problem of describing the spatial relations between voxels.

Figure 2 depicts three different volume samples in both 3D and cross-section version, their histograms and the detected locations within the human body atlas using the histogram matching method.

1) Histogram dissimilarity measures: There are several cross-bin and bin-to-bin dissimilarity measures to distinguish among histograms being compared [6]. Generally, cross-bin distances such as Earth Movers Distance (EMD) tend to be more robust and more discriminative than bin-to-bin measures when comparing two histograms [9].

Let L and M represent histograms being compared and each value $M(i)$ the count of the number of observations that fall



(a) Vol. No.346: crossection.

(b) Vol. no.346: 3D.

(c) Vol. no.346: histogram.

(d) Volumes' locations projected on the atlas.

(e) Vol. no.453: crossection.

(f) Vol. no.453: 3D.

(g) Vol. no.453: histogram.

(h) Vol. no.49: crossection.

(i) Vol. no.49: 3D.

(j) Vol. no.49: histogram.

Fig. 2. Three volumes with their histograms and detected location within human body atlas.

into one of the disjoint intensity intervals. If n be the total number of observations and k the total number of bins, then the following holds:

$$n = \sum_{i=1}^{k} M(i)$$

Among the most popular bin-to-bin measures let us mention: histogram intersection, L_1 and L_2 norm, χ^2, and Jeffreys Divergence with calculation formulas (in respective order)

- $h(L,M) = \frac{\sum_j min(L(j),M(j)))}{\sum_j M(j)}$
- $L_1(L,M) = \frac{\sum_j abs(L(j)-M(j))}{N}$
- $L_2(L,M) = \frac{\sum_j (L(j)-M(j))^2}{N}$
- $\chi^2(L,M) = \frac{\sum_j (L(j)-M(j))^2}{\sum_j L(j)+M(j)}$
- Jf$(L,M) =$ $\sum_j \left(L(j)log\frac{2L(j)}{L(j)+M(j)} + M(j)log\frac{2M(j)}{L(j)+M(j)} \right)$

An important representative of the cross-bin histogram comparison method is called the Earth Movers Distance (EMD) [8], [9]. Fast implementation of the cross-bin histogram comparison EMD was used in our tests [8].

2) Weighting the coordinates: We have performed the tests by calculating all of the inter-histogram dissimilarities. After this calculation, ten most similar volumes were queried out of the base and used to estimate the coordinate. Estimation was based on weighted averaging where weights sum up to unit and each weight corresponds to the similarity level d_i and follows negative exponential function with slope determined by normalized variance of 10 selected dissimilarities.
The coordinates of the ten best matches (ranked with i) were weighted by exponentially decaying weights, derived from the variance of their dissimilarity measures. The resulting coordinates $[x_{est}, y_{est}, z_{est}]$ were calculated by

$$[x_{est}, y_{est}, z_{est}] = \sum_{i=1}^{10} w_i[x_i y_i z_i]$$

where $w_i = e^{-\lambda i}$ and $\lambda = max\left(\frac{d_i - mean(d_i)}{var(d_i)}\right)$.

Such weighting was used to overcome the influence of the outliers on the estimation since outliers can cause higher variance in general as shown in Fig. 3.

B. Classifier regression

Another approach to predict the coordinate values uses classifier regression. Classifiers such as 10-NN and Random Ferns

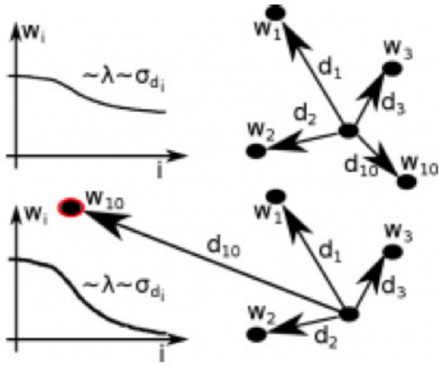


Fig. 3. The weights of the outliers are smaller than the weights of the neighboring candidates.

[7] use the image features and carry out the regression instead of usual crisp classification. PRTools Matlab library [3] is used for 10-NN implementation while Piotrs Matlab Toolbox [10] is used for Random Ferns implementation. Classifier scores are used to interpolate coordinates of those that were classified as being close to the test sample in a way typical for each classifier. Namely, averaging of the 10 nearest neighbors coordinates will be 10-NN regression. Similarly, in the case of SVM classifier usage, coordinates of the most similar ones separated by SVM will be interpolated. Features used for classifier regression are: mean, standard deviation, median, volume (actual number of pixels in the volume), centroid (1^{st} order image moments), central moments (2^{nd} order image moments), and the voxels themselves can be used as features, too. In cases when voxel intensities are used (Random Ferns classification), binary features need to be extracted - hence number of combinations can grow high. Therefore, volumes were resized using Gaussian pyramid before calculating the binary features.

III. Results

We tested the methods on a dataset which consist of 3926 volumes (miniatures representing body regions). Each of the volumes was resampled to 32 × 32 × 32 image cube. Voxel intensities of the image cube are real values ranging from 0 to 4095. Each volume corresponds to certain spatial regions of the body as it was shown with dots (center points of the region) in Fig. 5. Volumes are actually resembling different regions of the body and they are taken from different orientations (transverse, coronal, sagital).

Results are presented for both approaches introduced in the Section II. ”Histogram comparison” turns out to be quite precise at finding similar patches. However, the choice of the dissimilarity metric is very important. As mentioned, cross-bin dissimilarity is preferred since it is more robust. The performance of different metrics when matching histograms for query volume (object nr. 1540) shown together with its top 10 most similar patches from the database. Process finds 10 most similar volumes and interpolates their coordinates.

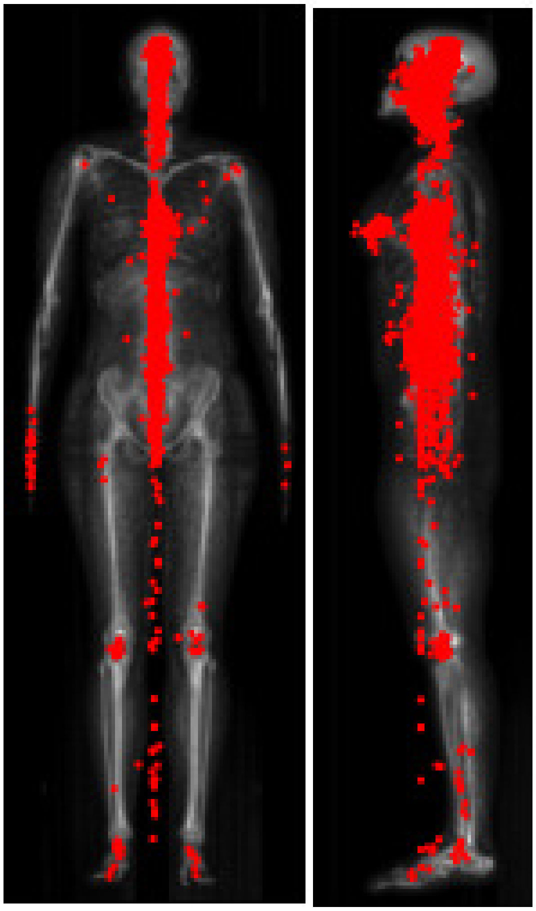


Fig. 4. Human atlas with 3926 volumes.

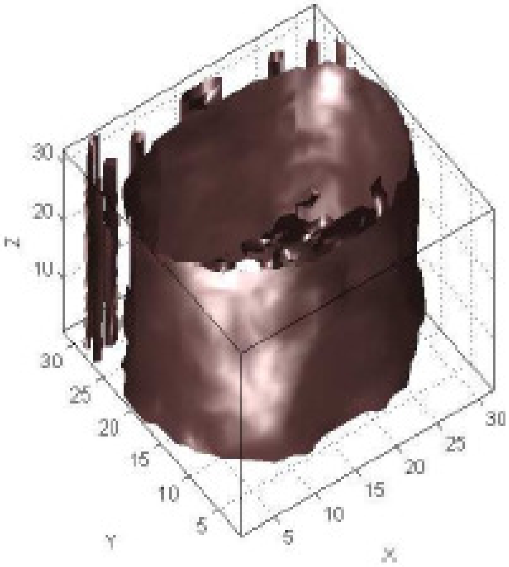


Fig. 5. Voxel representing a region of a human body.

Dissimilarity measure	h	L_1	L_2	χ^2	Jf	EMD
Estimated position error	16.5792	3.7436	129.8316	1.3632	1.2895	**1.2466**

TABLE I

ESTIMATED POSITION ERRORS FOR DIFFERENT HISTOGRAM COMPARISON MEASURES (INTERSECTION, L1 AND L2 NORM, χ^2, JEFFREY DIVERGENCE AND EMD) FOR THE OBJECT NR. 1540

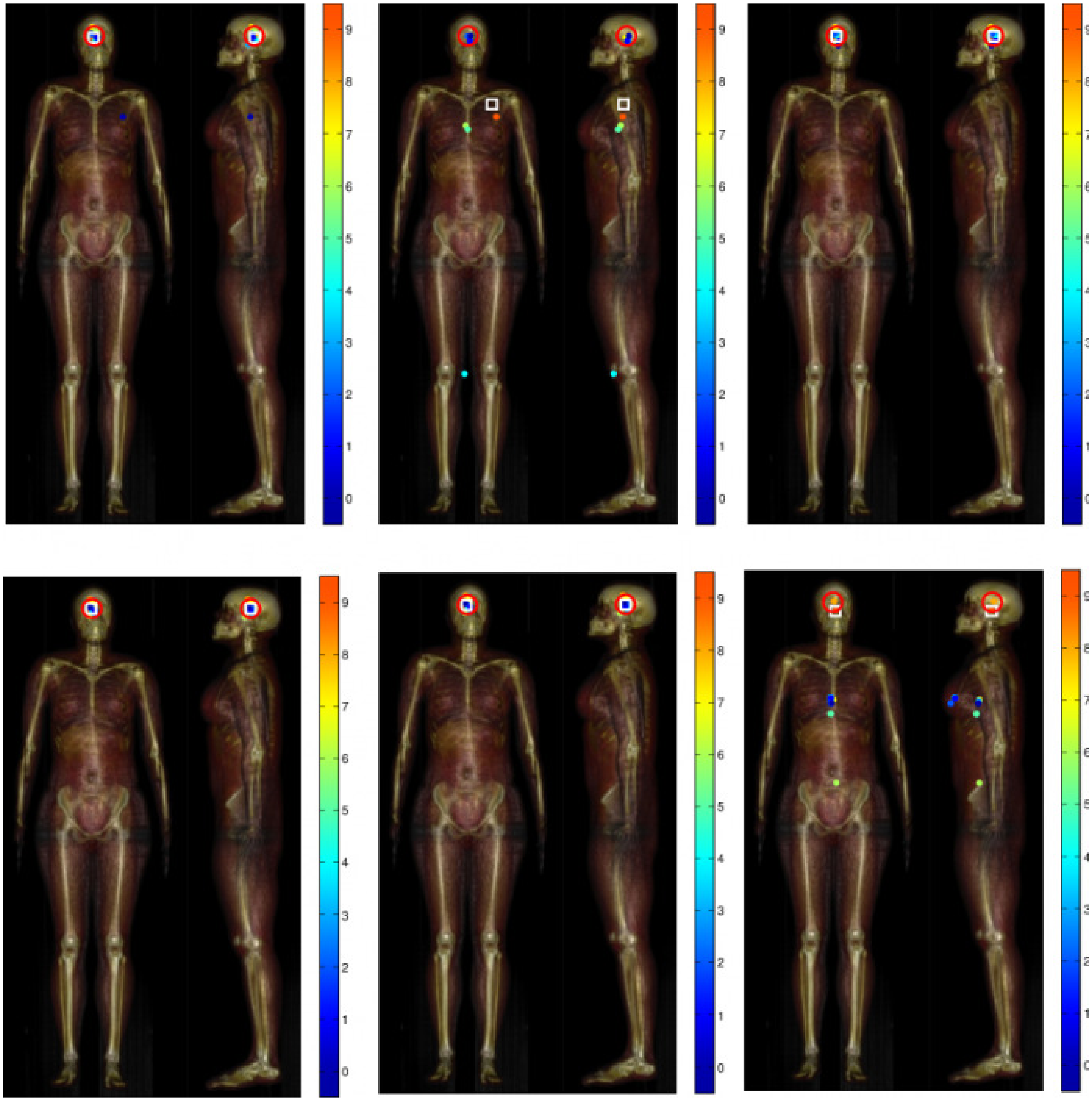


Fig. 6. Results of volume matching using histogram comparison for dissimilarity measures L1 norm, L2 norm, EMD, χ^2, Jeffrey divergence, and Intersect (from left to right, top to bottom).

EMD dissimilarity measure turns out to be the one that performs the best both in terms of accuracy and robustness, however it is slower and features used for matching are dependent on the database. Downside of its usage is computational cost, nevertheless fast implementation of EMD [8] used in this experiment provides satisfactory performance. What is still a potential problem with this approach is the fact that it is dependent on the data from the training set - since it is based on comparing the input with the training dataset. In that sense - classifier regression based approach with invariant statistical/shape/moment based features would be interesting. Figure 6 show the results of volume matching using histogram comparison for dissimilarity measures: L1 norm, L2 norm, EMD, and χ^2, Jeffrey divergence, and intersection (from left to right, top to bottom). Predicted value is represented with white square, while the true value is described with red circle. Top 10 matching positions were shown as dots in jet color space - ranging from red ones that represent high ranking till the blue ones with low ranking. Estimation fails in case metric was not properly chosen. EMD turns out to be the one that performs the best both in terms of accuracy and robustness.

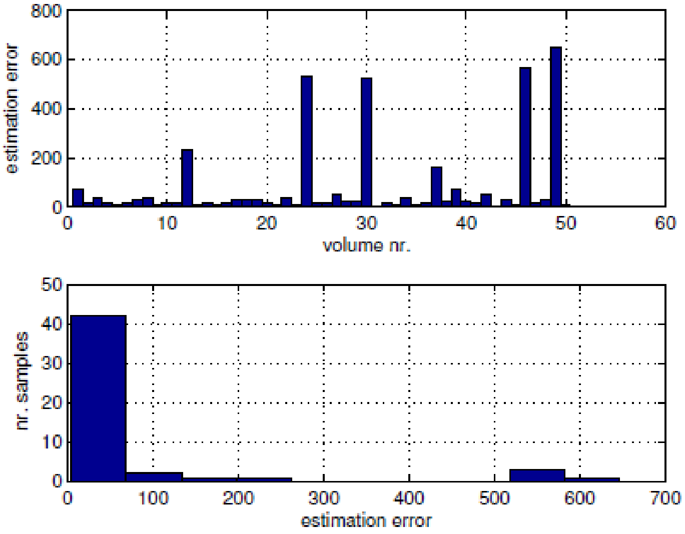


Fig. 7. Statistics for the absolute distance error between the position estimate and the actual position.

Finally, histogram matching was performed individually on 50 randomly chosen objects while the rest were used for matching with the test sample each time. Figure 7 shows the statistics for the absolute distance error between the position estimate and the actual position. Majority of discrepancies fall into 0-50 voxel distance range which suggests that the object was targeted. Usually, distances up to 50 absolute voxel values do properly guess the object since the main task is rough localization of the body region.

”Classifier regression” was based on independent features (such as volume mean, standard deviation, image moments, volume). As stated before - classifier score was the basis for the data regression. 10-NN classifier was used to find the neighbors in feature space and average the coordinates associated with them in order to carry out the regression. This process works much faster since it takes short time to extract the features and there is no need for comparison. It accomplishes the retrieval but the performance is lower as shown in the figure 8.

Regression accuracy is worse than the one obtained using histogram comparison and one of the causes is probably plain averaging of the feature-space-nearest-neighbor coordinates (no weighting). The features used can influence the choice of the nearest neighbors. Finally, Random Ferns [7] were tested on position regression. They performed slightly slower than 10-NN regression, but had an overall lower average error calculated over 50 randomly chosen miniatures. Voxel intensities were used as the source for making binary features that this classification algorithm uses. Usage of statistic/shape features that were supplied to 10-NN did not perform as well as voxel intensities did. Binary features were chosen to be -1 or +1 for each possible pair of voxels. In case the first one from the pair was higher - feature was set to +1 and -1 was for the opposite case. Each miniature needed to be resized to $8 \times 8 \times 8$ using Gaussian pyramides and even after the reduction - quarter of the total number (every 4^{th}) of the voxels was used for matching in the experiment due to processing time constrains. Results proved to be promising in spite of all the constrains.

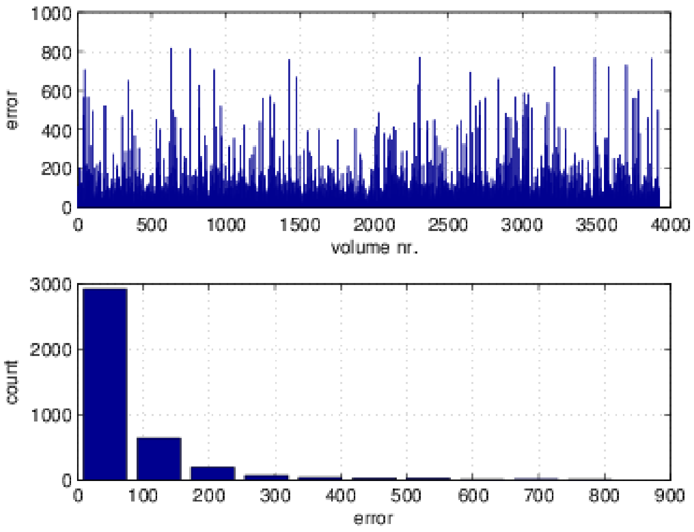


Fig. 8. Results of classifier regression (10-NN) based on independent features

IV. Conclusion

This paper summarizes the results obtained while developing 3D content based image localization framework. Two approaches have been examined - one based on comparing the histograms and another one based on the classifier regression. 10-NN and Random Ferns classifiers were used for the regression. The best performance in terms of accuracy was achieved by comparing histograms using the EMD (Earth Movers Distance) method. Classification regression using 10-NN proved to be slightly less precise but significantly faster, and the same conclusion can be made for Random Ferns that found their best performance by taking voxel values as features and compared the 3D patches for similarity. Average error on position prediction calculated on 50 random samples was lower than the one obtained using 10-NN. Nevertheless, Random Ferns classification was significantly constrained with reducing the feature number for the computational time purposes, therefore its performance can certainly improve. Future work can include clustering of the coordinates into crisp classes so that further evaluation of position prediction performance can be estimated in a more appropriate way. 10-NN regression can be improved by using weighted average of the nearest neighbors or some additional features. Random Ferns regression can perform better with more elegant way of reducing features it uses. Finally, local feature extraction methods such as 3D SIFT or 3D LBP can be used.

REFERENCES

[1] M.-C. Chang and B.B. Kimia. *Measuring 3d shape similarity by graph-based matching of the medial scaffolds.* Computer Vision and Image Understanding, 115(5):707 720, 2011. Special issue on 3D Imaging and Modelling.

[2] T. Deselaers, D. Keysers and H. Ney. *Features for image retrieval: an experimental comparison.* Information Retrieval 11:77107, 2008. 10.1007/s10791-007- 9039-3.

[3] R.P.W.Duin, P.Juszczak, P.Paclik, E.Pekalska, D.deRidder, D.M.J.Tax, and S.Verzakov. *Pr-tools4.1, a matlab toolbox for pattern recognition.* http://prtools.org, 2007.

[4] J.Fehr and H.Burkhardt. *3d rotation invariant local binary patterns. In Pattern Recognition, 2008.* ICPR 2008. 19th International Conference on, pages 1–4, dec. 2008.

[5] J. Flusser, T. Suk, and B. Zitov. *Moments and Moment Invariants in Pattern Recognition.* John Wiley & Sons, Ltd, 2009.

[6] F.D. Jou, K.C. Fan, and Y.L. Chang. *Efficient matching of large-size histograms.* Pattern Recognition Letters, 25(3):277 286, 2004.

[7] M. Ozuysal, M. Calonder, V. Lepetit, and P. Fua. *Fast keypoint recognition using random ferns. Pattern Analysis and Machine Intelligence*, IEEE Transactions, 32(3):448 461, march 2010.

[8] O. Pele and M. Werman. *Fast and robust earth movers distances.* In Computer Vision, 2009 IEEE 12th International Conference on, pages 460 467, 29 2009-oct. 2 2009.

[9] O. Pele and M. Werman. *The quadratic-chi histogram distance family.* Computer Vision ECCV 2010, volume 6312 of Lecture Notes in Computer Science, pages 749762. Springer Berlin / Heidelberg, 2010.

[10] D. Piotr. *Piotrs image and video matlab toolbox (pmt).* http://vision.ucsd.edu/ pdol- lar/toolbox/doc/index.html.

[11] Y. Qian, X. Gao, M. Loomes, R. Comley, B. Barn, R. Hui, and Z. Tian. *Content-based retrieval of 3d medical images.* In eTELEMED 2011, The Third International Conference on eHealth, Telemedicine, and Social Medicine, pages 712, 2011.

[12] Y. Rui, T.S. Huang, and S.F. Chang. *Image retrieval: Current techniques, promising directions, and open issues.* Journal of Visual Communication and Image Representation, 10(1):39 62, 1999.