Entropy-Optimized Texture Models

Sebastian Zambal, Katja Bühler, and Jiří Hladůvka

VRVis Research Center for Virtual Reality and Visualization

Abstract. In order to robustly match a statistical model of shape and appearance (e.g. AAM) to an unseen image, it is crucial to employ a robust model fittness measure. Dense sampling of texture over the region covered by the shape of interest makes comparison of model and image in principle robust. However, when merely texture differences are taken into account, problems with low contrast regions, fuzzy features, global intensity variations, and irregularly varying structures occur.

In this paper we introduce a novel entropy-optimized texture model (ETM). We map gray values of training images such that pixels represent anatomical structures optimally in terms of information entropy. We match the ETM to unseen images employing Bayes' law.

We validate our approach using four training sets (hearts in basal region, hearts in mid region, brain ventricles, and lumbar vertebrae) and conclude that ETMs perform better than AAMs. Not only they reduce the average point-to-contour error, they are better suited to cope with large texture variances due to different scanners and even modalities.

1 Introduction

Statistical shape models exploit prior knowledge of shape to perform robust segmentation. Active Shape Models (ASMs) [1] are based on Principal Component Analysis (PCA) of a set of corresponding training shapes. In the matching process, the model iteratively deforms attracted by image features (e.g. edges).

Additionally to shape, Active Appearance Models (AAMs) [2] incorporate texture variations (represented again by PCA). In the following we motivate our new approach to texture representation for statistical models by resuming some critical aspects of texture modeling in AAMs.

Low contrast and fuzzy structures. Model matching is understood as finding the model instance which exhibits the minimal difference to the unseen image. Hence it is very important to employ an effective difference measure which consistently reflects how well a given model instance matches the image. In the context of AAMs, typically measures based on texture differences are used. The assumption is made that large texture differences correspond to large misalignment of the model. However, low contrast (e.g. lung and myocardium in cardiac MRI) and fuzzy/irregular structures (e.g. spongy bone, trabeculae) make it difficult to objectively compare texture differences which are measured in different image regions. It may thus happen, that regions of low contrast are over-ruled by regions showing fuzzy structures.

PCA-based texture modelling. For many data sets we observed that a large number of PCA texture parameters are required to sufficiently well cover variations in a training set. For the data investigated in this paper we observed that the sorted eigenvalues of shape decrease much faster than those of texture. This suggests that PCA is not the optimal choice for texture modeling.

Texture normalization. To improve the quality of AAMs, a texture normalization step is usually applied before statistical analysis. The goal is to keep irrelevant texture variations (e.g. global variations in brightness or gamma) out of the model. In the scope of AAMs different methods have been proposed (e.g. different types of non-linear texture normalization [3]). However, in general it is hard to predict which type of texture normalization leads to good results for which kind of data.
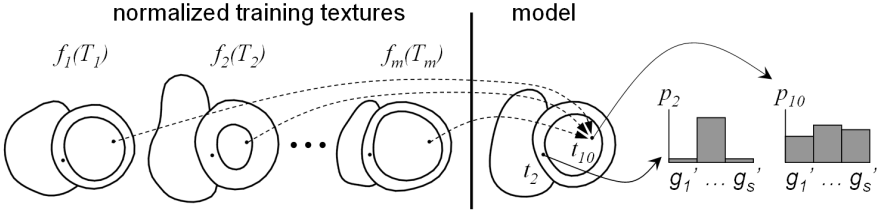
Mutual Information. A very popular matching criterion for medical image registration is mutual information (MI) [4]. The great advantage of this measure is that it makes it even possible to register images which were acquired using different imaging methods. Mutual information is based on entropy terms and in its original form limited to registration of a pair of images. Thus, although it would be reasonable to integrate mutual information into the AAM framework this cannot be done straight forwardly.

Bayesian Framework. Bayesian reasoning makes use of posterior probabilities to quantify how well a given model explains unknown data. Based on Bayes' law prior probability is combined with the likelihood of an observation to derive the posterior probability of the model. Bayesian reasoning has a sound mathematical background and is well-established for pattern recognition tasks. Thus it seems worth to incorporate it into the framework of statistical models of shape and appearance.

To tackle the above issues we borrow ideas about entropy from groupwise registration and formulate a novel probabilistic texture model. In section 2 we derive an optimization function for texture normalization which has similarities with a recently proposed function for *shape* correspondence optimization [5]. In section 3 we match the model to unseen images using Bayesian inference. We demonstrate the robustness of our entropy-optimized texture model (ETM) on four different 2D training sets: vertebra, brain ventricles, mid cardiac, and basal cardiac slices (section 4). Based on this validation we conclude in section 5 that the proposed ETM outperforms AAMs and even copes with texture variations due to different imaging modalities (different MRI scanners, T1/T2/FLAIR, CT/MRI).

2   Entropy Texture Model: Construction

Similar to AAM, the input for the model we propose is a set of m training images T_i, annotated with a fixed number of landmark points in correspondence, consistently triangulated, resampled by n texels and quantized to r_i gray levels.

normalized training textures   model

$f_1(T_1)$ $f_2(T_2)$ $f_m(T_m)$

p_2 p_{10}

t_2 t_{10}

$g_1' \cdots g_s'$ $g_1' \cdots g_s'$

Fig. 1. The entropy texture model: normalized corresponding pixels are modeled by probability distributions

While identical in shape description by mean and principal components, our model differs from AAMs in the representation of texture. Each of the n model texels t_j captures the statistical variations observed at corresponding training pixels.

In order to keep unspecific intensity variations out of the model we propose to normalize the training images using generic intensity mappings. The task is to optimize m intensity maps f_i that quantize the training images from r_i to s gray levels.

$$f_i : \mathbb{Z}_{r_i} \to \mathbb{Z}_s, \quad s \ll r_i \qquad i = 1 \dots m \tag{1}$$

In the following we design the cost function to assess mappings f_i. Each model's texel t_j observes m occurrences of the s possible mapped gray values at corresponding training texels. Such a set of observations can be interpreted as probability density function (PDF) and we denote it p_j. $p_j(g_k')$ is the probability that at corresponding texel t_j a mapped gray value g_k' is observed.

The predictability of mapped gray values expressed by PDFs varies across model texels. Figure 1 illustrates the situation for two texels. While PDF p_2 exhibits a single high peak, p_{10} is rather equally distributed. This suggests that predictions of gray values for texel t_2 will be more reliable than for texel t_{10}.

Inspired by Balci *et al.* [6] we propose to favor reliability (similar to p_2) and to penalize uncertainty (similar to p_{10}) by minimizing the entropy of the corresponding PDFs: $H(p_j) = - \sum_{k=1}^{s} p_j(g_k') \log_2(p_j(g_k'))$. Considering all n model texels yields a cost function that measures the quality of mappings f_i:

$$H^{model} = \frac{1}{n} \sum_{j=1}^{n} H(p_j) \quad \to \min \tag{2}$$

H^{model} yields minimum when the mappings degenerate such that all mapped training pixels show the same gray value. In this case the model exhibits maximal specificity.

However, this specificity has to be put into relation to the training images which contain no information once degenerated. This motivates for a compensation term which drives the mapped training textures $f_i(T_i)$ towards maximum information

(a) original   (b) optimized
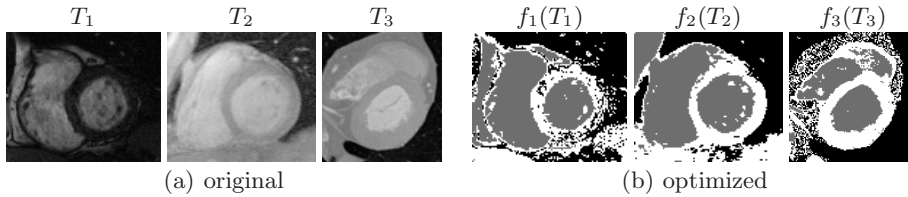
Fig. 2. A sample training set (two different MRI scanners and a CT) before (a) and after (b) mapping to three gray levels ($s = 3$)

content. We measure the information content by *image entropy* $H(f_i(T_i))$ and aim at its maximization across the entire normalized training set:

$$H^{tex} = \frac{1}{m} \sum_{i=1}^{m} H(f_i(T_i)) \quad \rightarrow \max \tag{3}$$

H^{tex} reaches a maximum when the texture transformations f_i maximize the information content of the textures T_i. This is equivalent to histogram equalization.

We combine the entropy terms of model (eq. (2)) and training textures (eq. (3)). The goal is maximization in terms of texture mappings f_i:

$$\{f_1^*, \ldots, f_m^*\} = \operatorname*{argmax}_{\{f_1, \ldots, f_m\}} \quad (H^{tex} - H^{model}) \tag{4}$$

There is no need for an ad-hoc weighing of the two entropy terms since they are commensurate, i.e., measured in bits per texel and bits per pixel respectively. An objective function similar to (4) has recently been proposed for *shape* correspondence optimization [5]. In this work similarly two energy terms are used: one for individual shapes and one for a shape ensemble. This makes our entropy-optimization of texture a direct analogy to Minimal Description Length (MDL) of shape with the difference that we do not stick to Gaussian distribution as description language for texture.

We used Simulated Annealing to optimize mappings f_i subject to the cost function (4). Represented by lookup tables, the texture mappings f_i are initialized to linearly remap the training images to s levels. During optimization the lookup tables are updated in each iteration. After optimization the mapped training textures represent structures with minimal uncertainty at maximum information content (figure 2(b)).

Figure 3 shows entropies of the optimized distributions of four models we use for validation in section 4. While bright texels exhibit strong variations (high entropy), dark texels represent high predictability (low entropy).

3 Entropy Texture Model: Matching

In contrast to AAMs, only shape parameters $S = \{$position, scaling, rotation, and statistical shape deformation$\}$ are optimized. The cost function describing model

(a) heart LV+RV   (b) heart LV   (c) vertebra   (d) brain ventricles
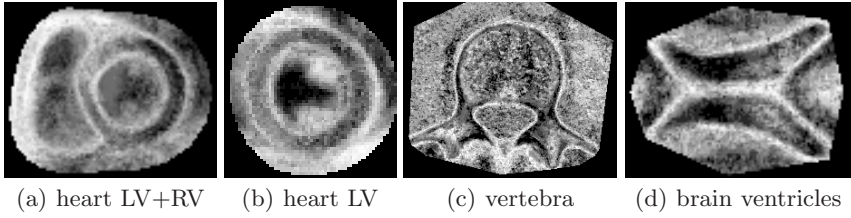
Fig. 3. Entropy (brightness) of model PDFs after optimization

matching accuracy is based on model PDFs and sub-texture $U = (u_1, \ldots, u_n)$ of the unseen image currently overlapped by the model.

Mapping of the Unseen Image'S Texture. As our model texels capture PDFs of s normalized gray values, we first need to normalize U accordingly: we seek for a function $f_u : \mathbb{Z}_{r_u} \rightarrow \mathbb{Z}_s$ that maps the texture to s gray levels.

We want to find a suitable intensity mapping for each of the r_u gray values. For a fixed gray value $\hat{u} \in \{1, \ldots, r_u\}$ all model texels observing it vote according to their PDFs: out of the s possibilities the target gray value leading to the maximal likelihood is assigned. Formally:

$$f_u(\hat{u}) = \operatorname*{argmax}_{g'_k} \prod_{u_j = \hat{u}} p_j(g'_k) \qquad k = 1 \ldots s \tag{5}$$

Cost Function. To assess how well the shape instance S fits to the unseen image, we compare model PDFs $p(.)$ to the normalized texels $U' = f_u(U) = \{f_u(u_j)\}_{j=1}^n$. To convey this, we consider maximizing the posterior probability $P(S|U')$ of the shape instance S, with the observed normalized texture U' given. According to Bayes' law [7] the posterior is proportional to the product of a likelihood and a prior (left side of eq. (6)).

The *likelihood* $P(U'|S)$ of the observed normalized texture U', with the shape instance S given, is (under the *naive* Bayesian assumption) the product of probabilities of the normalized texels.

The *prior* probability $P(S)$ of shape instance S is calculated feeding its parameters into the multivariate Gauss distribution that corresponds to the PCA shape space (please refer to [7] for details).

Putting both terms together yields the cost function which is maximized subject to shape parameters S:

$$P(S|U') \quad \propto \quad P(U'|S)\,P(S) \quad = \quad \left(\prod_{j=1}^n p_j(f_u(u_j)) \right) \, P(S) \tag{6}$$

Strategy and Implementation Issues. In contrast to AAM there are no texture parameters to optimize. The search space is only spanned by shape parameters and typically has not more than 10 to 15 dimensions. This fact makes exhaustive search strategies similar to [8] affordable and we employ such

| (a) | (b) | (c) | (d) | (e) |
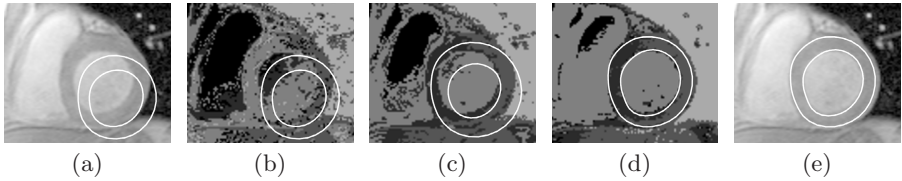
Fig. 4. Matching of a model comprising five target values ($s = 5$): original image with initialized contour (a), maximum likelihood estimate for f_u for initial contour (b), after 4 iterations (c), and after 8 iterations (d). Final match over the original unseen (e).

a scheme in this work. Figure 4 illustrates how a model with five gray levels evolves during matching. Shape and maximum likelihood texture mapping converge consistently.

Any $p_j(.)$ being zero degenerates the equations (5) and (6). In order to avoid this we assign a small compensation probability of $1/(2m)$ and renormalize the distribution p_j.

To avoid numerical instability due to the large number of multiplications in (5) and (6) we maximize sums of logarithms of the probabilities instead.

4   Validation and Results

In order to validate our approach we compared the segmentations achieved by ETM ($s = 5$) to those achieved by standard AAM. For cross validation we split training sets into reasonable subsets (e.g. images from same patient or same scanner). Every image was matched as an unseen by a model built from complementary subsets. Model matching was performed by initializing each of the models close to the correct contours (hence no automatic initialization of the model was done). Matching accuracy was measured in terms of average point-to-contour distances. To include enough information about object boundary regions into the models, additional landmark points outside the objects were defined. In the following we describe the data sets and the results in detail.

Heart LV+RV, basal region: contains short axis slices from the base of the heart showing few or no papillary muscles. The training set comprises data from four different sources: CT and three different MR protocols. Annotations define the contours of inner and outer boundaries of the left ventricle (LV) and the inner contour of the right ventricle (RV). ETM performed better in 23 out of 38 cases, reducing the average point-to-contour distance from 4.36 to 3.36 mm.

Heart LV, mid region: includes 42 short axis MR images from the mid region of the heart exhibiting large texture variations due to papillary muscles. Images stem from two different MR scanners.The average pixel size is 1.39mm. Annotations define the contours of the inner and outer boundaries of the left ventricle (LV). ETM performed better in 38 out of 42 cases, reducing the average point-to-contour distance from 4.02 to 2.38 mm.
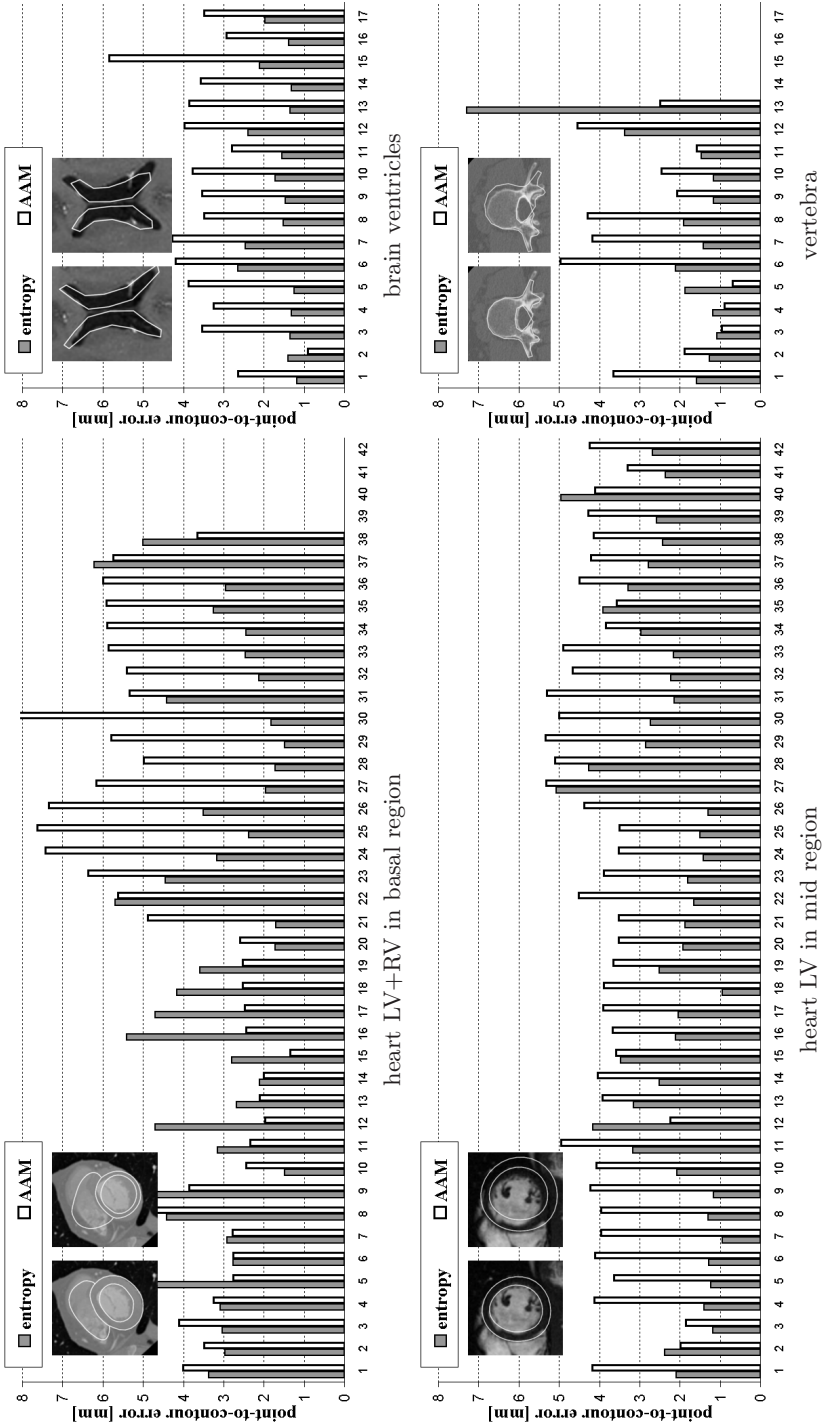
Fig. 5. Average point-to-contour errors [mm] achieved by ETMs and AAMs. Example thumbnails show final segmentations.

Brain ventricle: comprises 17 transversal MR slices of 15 different patients. One patient is represented by a T1 image and a FLAIR image. Another patient is represented by a T2 image and a FLAIR image. Remaining patients are represented each by a T1 image. The average pixel size is 0.85mm. Annotations define the contours of brain ventricles. ETM performed better in 16 out of 17 cases, reducing the average point-to-contour distance from 3.53 to 1.67 mm.

Vertebra: consists of 13 CT slices from 4 different patients showing transversal sections of lumbar vertebrae. The average pixel size is 0.32mm. Annotations define the outer contour of the vertebra and the contour of the spinal canal. ETM performed better in 9 out of 13 cases, reducing the average point-to-contour distance from 2.67 to 2.06 mm.

On average our ETMs clearly outperformed AAMs. Results are summarized graphically in figure 5.

5   Conclusions and Future Work

We proposed a novel texture model for medical image segmentation. Its texture is described by probability distributions of individual texels. Moreover, these probabilities are optimized by information entropy terms. This allows to effectively cope with unspecific intensity variations in the training set caused e.g. by different modalities, scanner settings, and fuzzy anatomical structures. The model is matched to unseens in accordance with Bayesian reasoning.

There are two major messages we would like to conclude with. First, ETM perform better than AAM, as it is shown in the validation. Second, thanks to intensity normalizations that are integral part of model construction, some requirements on training texture (e.g. Gaussian distribution) can be relaxed. We have demonstrated this with training sets containing mixtures of CT/MR and MR T1/T2/FLAIR images.

There are several issues left for future work. To name a few, we intend to address extension to 3D volume textures, automatic initialization, and acceleration of matching in the near future.

References

1. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active Shape Models – Their training and application. Computer Vision and Image Understanding 61(1), 38–59 (1995)
2. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active Appearance Models. In: European Conference on Computer Vision, vol. 2, pp. 484–498 (1998)
3. Kittipanya-ngam, P., Cootes, T.F.: The effect of texture representations on AAM performance. In: International Conference of Pattern Recognition, pp. 328–331 (2006)
4. Viola, P., Wells, W.M.: Alignment by maximization of Mutual Information. International Journal of Computer Vision 24(2), 137–154 (1997)

5. Cates, J., Fletcher, P.T., Styner, M., Shenton, M., Whitaker, R.: Shape modeling and analysis with entropy-based particle systems. In: Information Processing in Medical Imaging, pp. 333–345 (2007)
6. Balci, S.K., Golland, P., Shenton, M., Wells, W.M.: Free-form B-spline deformation model for groupwise registration. In: Medical Image Computing and Computer–Assisted Intervention (Statistical Registration Workshop), pp. 23–30 (2007)
7. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, Heidelberg (2006)
8. de Bruijne, M., Nielsen, M.: Shape particle filtering for image segmentation. Medical Image Computing and Computer–Assisted Intervention 1, 168–175 (2004)