

MRI Tissue Classification with Neighborhood Statistics: A Nonparametric, Entropy-Minimizing Approach

Tolga Tasdizen¹, Suyash P. Awate¹, Ross T. Whitaker¹, and Norman L. Foster²

¹ School of Computing, University of Utah

² Department of Neurology, University of Michigan

Abstract. We introduce a novel approach for magnetic resonance image (MRI) brain tissue classification by learning image neighborhood statistics from noisy input data using nonparametric density estimation. The method models images as random fields and relies on minimizing an entropy-based metric defined on high dimensional probability density functions. Combined with an atlas-based initialization, it is completely automatic. Experiments on real and simulated data demonstrate the advantages of the method in comparison to other approaches.

1 Introduction

Segmentation of magnetic resonance images (MRI) of the brain is an important problem in biomedicine; it has a number of applications including diagnosis, surgical planning and monitoring therapy. One of the fundamental tasks in brain MRI segmentation is the classification of volumetric data (3D images) into gray matter, white matter and cerebral-spinal fluid (CSF) tissue types. This classification is of great interest in the study of neurodegenerative disorders such as Alzheimer’s Disease. It also has other applications such as the generation of patient-specific conductivity maps for EEG source localization. Manual segmentation of high-resolution 3D images is an extremely time consuming and subjective task; hence, automatic and semi-automatic brain tissue classification methods have been studied extensively in the field of biomedical image processing. Recent developments in automatic brain tissue classification have led to a class of systems that incorporate the following strategies:

1. Parametric statistical models of single-pixel image intensity for each tissue class,
2. Markov random field (MRF) type models of spatial smoothness,
3. Bias field correction, and
4. Digital brain atlas information.

In this paper, we propose a novel approach that combines the intensity and spatial smoothness models (items 1-2) using an unsupervised learning approach that incorporates nonparametric statistics of local neighborhoods. The proposed method is compatible with state-of-the-art segmentation methods that use probabilistic brain atlases [1, 2] and bias field correction [3], but it uses an information-theoretic, data-driven approach to incorporate image neighborhood information. Validation studies on simulated data demonstrate that this approach offers significant advantages over the state-of-the-art.

2 Related Work

This paper introduces a new approach for removing the effects of imaging noise in tissue classification using a statistical framework. Others have used non-linear diffusion for image denoising as a pre-processing step [4]. However, for probabilistic algorithms, it is intuitive to incorporate spatial smoothness constraints directly into the segmentation process via Markov random field (MRF) models [5, 6, 3, 7, 8]. These methods modify single-pixel tissue-probabilities with energies defined on local configurations of segmentation labels. Spatially smooth segmentations are assigned lower energies and therefore are more likely. However, such MRF models can over regularize the fine structured borders, e.g. the interface between gray and white matter; therefore, it is often necessary to impose additional, heuristic constraints [6, 3]. Active contour models [9, 10] have also been used to impose smoothness constraints for segmentation. These methods typically attempt to minimize the area of the segmentation boundary, an approach that also can over regularize interfaces. The method proposed in this paper formulates the segmentation problem in information-theoretic terms using probability density functions (PDFs) defined on the space of image neighborhoods. In contrast to MRF-type approaches to tissue classification regularization, which formulate neighborhood probabilities on discrete segmentation labels, the method relies on the discovery of regular patterns in the input data.

Lee *et al.*[11] analyze the intensity statistics of 3×3 pixel patches in optical images, in the corresponding high-dimensional spaces, and find the data to be concentrated in clusters and low-dimensional manifolds exhibiting nontrivial topologies. Motivated by this observation, we use nonparametric density estimation. Consequently, we impose very few assumptions about the statistical structure of image neighborhoods. Popat *et al.*[12] were among the first to use nonparametric Markov sampling in images. They attempt to capture the higher-order nonlinear image statistics via cluster-based nonparametric density estimation and apply their technique for image restoration, image compression and texture classification. However, their method takes a *supervised* approach for learning neighborhood relationships. The proposed method builds on the work in [13], which lays down the foundations for unsupervised learning of higher-order image statistics. However, that work proposes reducing the entropy of image-neighborhood statistics as a method for removing image noise.

3 Method

A random field/process [14] is a family of random variables $X(\Omega; T)$, for an index set T , where, for each fixed $T = t$, the random variable $X(\Omega; t)$ is defined on the sample space Ω . If we let T be a set of points defined on a discrete Cartesian grid and fix $\Omega = \omega$, we have a specific realization of the random field as a deterministic function $x(t)$ – the image. In the case of 3D MRI data, t is a three-vector and T represents the set of pixels in the 3D image. Let $N_t \subset T$ be the set of pixels in the neighborhood of t . If we associate with T a family of neighborhoods $N = \{N_t\}_{t \in T}$ such that $u \in N_t$ if and only if $t \in N_u$, then N is called a neighborhood system for the set T . An example of such a neighborhood is a $3 \times 3 \times 3$ cube of pixels centered at t . We define a random vector $Z(t) = \{X(t)\}_{t \in N_t}$, denoting its realization by $z(t)$, corresponding to the set of intensities at the neighbors of pixel t . If the image is real-valued, then $z(t) \in \mathbb{R}^m$ where m is the number of pixels in the neighborhood.

3.1 Neighborhood Entropy

Let $p_k(Z = z)$ be the probability of observing the image neighborhood z given that the center pixel of the neighborhood belongs to the tissue class k . This PDF is called the *likelihood* function. The total entropy associated with a set of K tissue PDFs is

$$H = - \sum_{k=1}^K \int_{\mathbb{R}^m} p_k(Z = z) \log p_k(Z = z) dz, \quad (1)$$

where the integration is performed over \mathbb{R}^m , the domain of the random vector Z . Let the sets $\{T_k\}_{k=1}^K$ denote a mutually exclusive and exhaustive decomposition of T into K tissue classes. We assume that the random process X , limited to pixels belonging to a single class T_k , is stationary ergodic. Entropy is the expectation of negative log-probability, and for such processes it can be approximated with the sample mean [15]

$$H(T_1, \dots, T_K) = - \sum_{k=1}^K \left(\frac{1}{|T_k|} \sum_{t \in T_k} \log p_k(z(t)) \right). \quad (2)$$

Piecewise stationarity is a better model for MRI, which is not truly stationary. In practice, the stationary ergodic assumption can be relaxed as shown in Section 3.2.

We consider the optimal decomposition (segmentation) to be the sets $\{\hat{T}_k\}_{k=1}^K$ for which H is minimum. If the PDFs $p_k(Z)$ are known, H can be minimized by assigning any pixel t to the class with the highest likelihood for that particular realization $z(t)$:

$$\hat{T}_k = \{t \in T \mid p_k(z(t)) \geq p_i(z(t)), \forall i \neq k\}. \quad (3)$$

In practice, the likelihood functions are not known a priori and have to be estimated as well. We propose to iteratively estimate the likelihood functions and update the segmentation sets $\{T_k\}_{k=1}^K$ until H converges. This approach is similar to the estimation of means and variances for Gaussian PDFs in the Expectation Maximization algorithm. However, Z is not well represented by a Gaussian PDF. For instance, consider the image neighborhood for a white matter pixel deep inside the white matter mass and for one close to the interface with gray matter. These two neighborhoods are drawn from a PDF with multiple modes. Hence, we use a nonparametric density estimation approach.

3.2 Nonparametric Multivariate Density Estimation

Entropy optimization on image neighborhoods entails the estimation of PDFs in sparsely populated, high dimensional spaces (the so-called *curse of dimensionality*). Despite theoretical arguments suggesting that density estimation beyond a few dimensions is impractical, empirical evidence is more optimistic [12, 13]. We use Parzen-window density estimation [16] with an m -dimensional isotropic Gaussian interpolation kernel, G_m , with standard deviation σ for all dimensions. The density estimate for class k is

$$p_k(Z(t) = z(t)) \approx \frac{1}{|A_k(t)|} \sum_{t_j \in A_k(t)} G_m(z(t) - z(t_j), \sigma), \quad (4)$$

where $A_k(t)$ is a small subset of T_k , chosen randomly. For a truly stationary random process, it is sufficient to form a global sample $A_k(t)$ and use it to evaluate the PDF for

all $z(t)$. For a piecewise stationary model, we estimate the PDF locally by choosing the locations in $A_k(t)$ from a spatial sampling Gaussian PDF centered at t .

Parzen-window density estimation involves the following parameters: the size of the set $A_k(t)$, the standard deviations of the spatial sampling Gaussian and the interpolation Gaussian kernel. The first two are not critical, and they can easily be fixed for a wide range of MRI. In all of our experiments, we fix the standard deviation of the spatial sampling Gaussian to 15 pixels. This also automatically determines the minimum number of required samples in $A_k(t)$ to be 1000 as explained in [13]. On the other hand, the standard deviation σ of the interpolation kernel in equation (4) is critical for successful density estimation in high dimensional spaces. The optimal choice for this parameter depends on the sparsity of the data which varies with various factors including the amount of noise present in an image. We choose σ to minimize the entropy of the associated PDF via a Newton-Raphson optimization scheme [13]. This choice for σ is consistent with our entropy minimization segmentation formulation.

3.3 MRI brain tissue classification algorithm

The algorithm segments four tissue classes: gray matter, white matter, CSF and non-brain tissue. Each class is represented with a non-parametric PDF. Since, the true PDFs are not known *a priori*, we use an iterative learning procedure. Another component in MRI brain tissue classification is the correction of intensity inhomogeneities. Various solutions to this problem have been proposed. The approaches in [20, 21, 3] are the most interesting because they propose a combined approach that iteratively updates the segmentation labels and the bias field correction. The focus of this paper is not on bias field correction; however, to show that our segmentation method is compatible with previous techniques, we have implemented a simplified version of the polynomial least-squares fit correction described in [3]. In summary, the segmentation algorithm carries out the following steps per iteration:

1. estimate likelihood functions from the current $\{T_k\}_{k=1}^K$ and bias field,
2. update $\{T_k\}_{k=1}^K$ according to the new likelihoods,
3. estimate the bias field from the new $\{T_k\}_{k=1}^K$.

The iterations are carried out until the reduction in entropy H computed from equation (2) drops below 0.1% of the current value of H .

The algorithm requires an initial partition of classes, which can be obtained by registering the MRI with a brain atlas. Digital brain atlases can be used to provide segmentation labels for a reference dataset [17, 18] or prior tissue probabilities computed from a population of subjects [1, 2, 19]. In this paper, we use the ICBM probabilistic atlas [19], which provides probabilities of gray matter, white matter and CSF classes for each pixel. We register this atlas with the MRI by computing the affine transformation that maximizes a mutual information metric. The registered probability maps are then used to form an initial segmentation using equation (3). As in [1, 2], we also treat these atlas probability maps as prior probabilities. However, we have found that while the priors help in the discrimination between brain tissues vs. non-brain tissues, they don't offer significant benefits for our method when choosing between the brain tissues. Hence, we sum the CSF, white matter and gray matter priors into a single brain-tissue prior map. The non-brain-tissue prior map is then obtained by subtracting this map from unity.

4 Results and Validation

We validate the proposed approach on simulated images with known ground truth. We use 1 mm isotropic T1-weighted images from the BrainWeb simulator [22] with varying amounts of noise and bias field. The neighborhood system is chosen to include the six adjacent pixels in the three Cartesian directions in addition to the center pixel. Hence, the PDFs exist in a seven dimensional space.

Van Leemput *et al.* [1] use the Dice metric [23] to evaluate the performance of their state-of-the-art Expectation Maximization and MRF based approach on images from the BrainWeb simulator. For comparison purposes, we use the same metric. Let \hat{T}_k denote the ground truth set of pixels in tissue class k , then the Dice metric for class k is defined as $2|\hat{T}_k \cap T_k|/(|\hat{T}_k| + |T_k|)$, where $|\cdot|$ denotes set size.

The first validation experiment is performed on simulated T1-weighted data without any bias field and with intensity noise levels in the range 0% – 9%. The noise percentages are defined with respect to the mean intensity of each tissue class. Figure 1 plots the Dice metric for gray and white matter tissue classifications from the proposed algorithm and the corresponding values given in [1]. The Dice metric for combined brain tissues vs. non-brain tissues is consistently above 98% for both algorithms and is not shown in this paper due the lack of space. The proposed algorithm performs better at

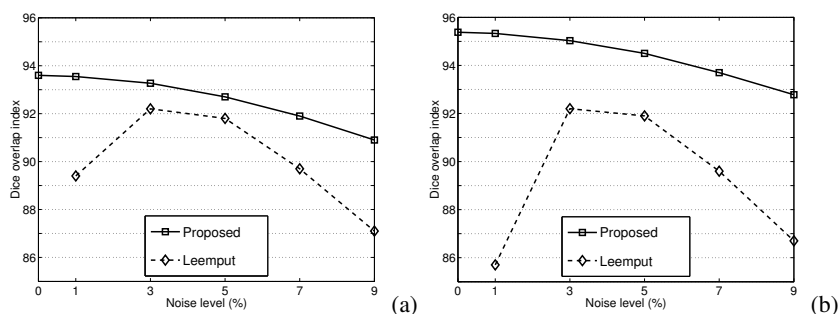


Fig. 1. Dice metric as a function of noise level: (a) gray matter, and (b) white matter. Solid and dashed lines plot the performance of the proposed algorithm and [1], respectively.

all noise levels for gray and white matter tissues. For 3% noise, which can be considered typical for real MRI, the performance gains are approximately 1.1% and 2.8% for gray and white matter, respectively. The proposed approach scales better with increasing noise amounts; the performance gain at 9% noise is 3.8% and 6.1% for gray and white matter, respectively. This property would be useful for segmenting clinical, fast acquisition MRI that can have high amount of noise. The use of image neighborhood information is critical to the success of the proposed method. We repeated the segmentation experiments at the 9% noise level using only the center pixel of the neighborhood. In this case, the method learns single-pixel intensity PDFs in a nonparametric manner. The Dice metrics for gray and white matter were 82% and 85%, respectively, a drop of 8% from the neighborhood algorithm. Figure 2 visually demonstrates this same drop in performance using an axial slice from the 3D image with 9% noise.

For low noise levels, the performance of the parametric Expectation Maximization algorithm drops dramatically [1] because pixels close to the interface between gray and white matter are systematically assigned to the class which happens to have the larger amount of natural variability, i.e. gray matter. In contrast, nonparametric density estimation does not suffer from this drawback as can be observed in Figure 1.

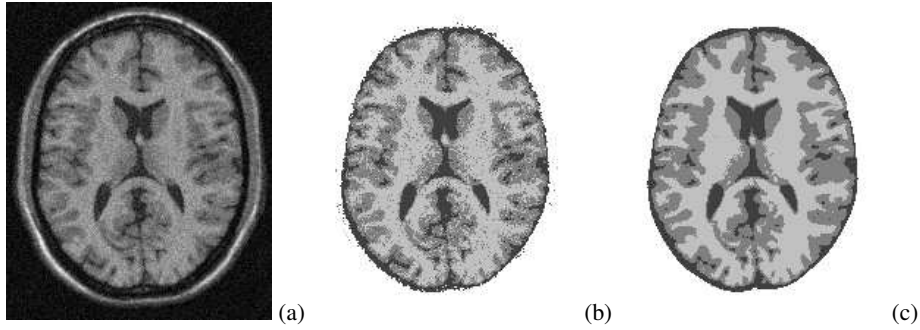


Fig. 2. (a) Simulated, noisy image. Tissue classification (b) without, and (c) with neighborhood information. Classification legend: CSF(black), gray matter(dark gray), white matter(light gray).

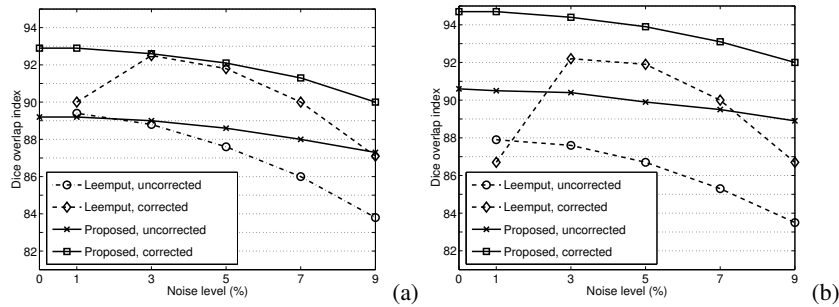


Fig. 3. Results on simulated data with 40% bias field. Dice metric as a function of noise level with and without bias field correction: (a) gray matter, and (b) white matter.

The second validation experiment is performed on simulated T1-weighted data with 40% bias field, i.e. the multiplicative factor is in the range 0.8-1.2 over the brain area. Figure 3 plots the Dice metric with and without bias field correction for the proposed algorithm and the corresponding values given in [1]. These results point out the importance of performing bias field correction. Also, as in the previous experiment, the proposed method performs better at all noise levels for both tissue types. We use a 2nd degree polynomial least squares fit to the observed multiplicative factors between white matter pixel intensities and the white matter mean intensity. The Dice metric values obtained with this correction method are approximately 0.5% worse than the values shown in Figure 1 for data with no bias field. Van Leemput *et al.* use a 4th degree polynomial fit using all tissue types and obtain results that are effectively the same as their results for unbiased data. We expect that with better bias correction methods, such as the one used in [1], this difference can be made very small for our algorithm as well.

Figure 2 illustrates that our approach can remove the effects of noise from the classification results without over regularizing the interfaces between different tissue types.

We have also tested the algorithm on a 1 mm isotropic real T1-weighted image. Figure 4 shows coronal and axial slices from this data with corresponding tissue classifications. In both of these cases, the fine structure of the interface between gray and white matter is preserved without the need of placing any additional constraints on the neighborhood PDFs learned by the algorithm.

The disadvantage of the proposed algorithm is the computational speed. For volumetric data with dimensions $181 \times 217 \times 181$, one complete iteration of the algorithm takes approximately 90 minutes to compute on an *Intel 2.7Ghz* processor. The algorithm requires 4-7 iterations to converge depending on the noise level in the data. The Parzen window density estimation is the bottleneck in computational speed; therefore, we plan to address computational issues by using fast density evaluation algorithms [24].

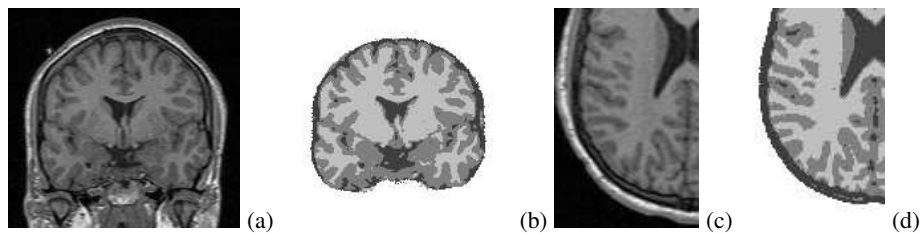


Fig. 4. Real data: (a) coronal slice, (b) classification, (c) close-up view of an axial slice, and (d) classification. Classification legend as in Figure 2.

5 Conclusion

In this paper, we introduced a segmentation method that uses entropy minimization to learn nonparametric statistics of local neighborhoods from noisy data in an unsupervised manner. This segmentation framework is used in conjunction with the ICBM brain atlas and bias field correction methods from the literature in an automatic MRI brain tissue classification application. Validation studies on simulated 3D images compare favorably to a state-of-the-art parametric algorithm based on MRFs. Validation studies on real data is also necessary, and will be performed as a continuation of this work. Experiments on real and artificial data also demonstrate that noise is effectively removed without over regularizing interfaces between different tissue types.

The algorithm easily extends to larger neighborhood systems and multi-modal data. Preliminary experiments with 2D images demonstrate that the classification performance can be slightly improved in these cases. Also, we have not considered the partial voluming effect, which has been studied in several papers in the literature. In future work, we plan to generalize the entropy defined in equation (1) from tissue labels to mixture percentages; hence, treating the partial voluming effect in an explicit manner.

Acknowledgments

This work was supported in part by NIH NCRR P41 RR12553-04, NSF EIA 0313268, NIH U01-AG024904, the Louise Madsen Memorial Fund and the Rolan K. Schuhholz Research Fund at the University of Michigan.

References

1. Leemput, K.V., Maes, F., Vandermeulen, D., Seutens, P.: Automated model-based tissue classification of mr images of the brain. *IEEE Tr. Med. Imaging* **18** (1999) 897–908

2. Pohl, K.M., Wells, W.M., Guimond, A., Kasai, K., Shenton, M.E., Kikinis, R., Grimson, W.E.L., Warfield, S.K.: Incorporating non-rigid registration into expectation maximization algorithm to segment mr images. In: Proc. MICCAI. (2002) 564–571
3. Leemput, K.V., Maes, F., Vandermeulen, D., Seutens, P.: Automated model-based bias field correction of mr images of the brain. *IEEE Tr. Med. Imag.* **18** (1999) 885–896
4. Gerig, G., Kubler, O., Kikinis, R., Jolesz, F.A.: Nonlinear anisotropic filtering of mri data. *IEEE Tr. Med. Imaging* **11** (1992) 221–232
5. Kapur, T., Grimson, W.E.L., Wells, W.M., Kikinis, R.: Segmentation of brain tissue from magnetic resonance images. *Med. Im. An.* **1** (1996) 109–127
6. Held, K., Kops, E.R., Krause, B.J., Wells, W.M., Kikinis, R., Muller-Gartner, H.W.: Markov random field segmentation of brain mr images. *IEEE Tr. Med. Imaging* **16** (1997) 878–886
7. Pachai, C., Zhu, Y.M., Guttman, C.R.G., Kikinis, R., Jolesz, F.A., Gimenez, G., Froment, J.C., Confavreux, C., Warfield, S.K.: Unsupervised and adaptive segmentation of multispectral 3d magnetic resonance images of human brain: a generic approach. In: Proc. MICCAI. (2001) 1067–1074
8. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain mr images through a hidden markov random field model and the expectation maximization algorithm. *IEEE Tr. Med. Imaging* **20** (2001) 45–57
9. Kelemen, A., Szekely, G., Gerig, G.: Three-dimensional model-based segmentation of brain mri. In: Proc. IEEE Workshop Biomedical Image Analysis. (1998) 4–13
10. Atkins, M., Mackiewicz, B.: Fully automatic segmentation of the brain in mri. *IEEE Tr. Med. Imaging* **17** (1998) 98–107
11. Lee, A., Pedersen, K., Mumford, D.: The nonlinear statistics of high-contrast patches in natural images. *Int. J. Comput. Vision* **54** (2003) 83–103
12. Papat, K., Picard, R.: Cluster based probability model and its application to image and texture processing. *IEEE Tr. Image Processing* **6** (1997) 268–284
13. Awate, S., Whitaker, R.: Higher-order image statistics for unsupervised, information-theoretic, adaptive, image filtering. SCI Institute Technical Report UUSCI-2005-004, University of Utah (2005) To appear in CVPR'05.
14. Dougherty, E.R.: *Random Processes for Image and Signal Processing*. Wiley (1998)
15. Viola, P., Wells, W.: Alignment by maximization of mutual information. In: *Int. Conf. Comp. Vision*. (1995) 16–23
16. Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*. Wiley (2001)
17. Fischl, B., Salat, D.H., Busa, E., ALbert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M.: Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* **33** (2002)
18. Collins, D.L., Zijdenbos, A.P., Barre, W.F.C., Evans, A.C.: Animal+insect: Improved cortical structure segmentation. In: *Proc. An. Symp. Inf. Proc. in Med. Img. Volume 1613*. (1999)
19. Rex, D.E., Ma, J.Q., Toga, A.W.: The loni pipeline processing environment. *NeuroImage* **19** (2003) 1033–1048
20. Wells, W.M., Grimson, W.E.L., Kikinis, R., Jolesz, F.A.: Adaptive segmentation of mri data. *IEEE Tr. Med. Imaging* **15** (1996) 429–443
21. Guillemaud, R., Brady, M.: Estimating the bias field of mr images. *IEEE Tr. Med. Imaging* **16** (1997) 238–251
22. Cocosco, C., Kollokian, V., Kwan, R.S., Evans, A.: Brainweb: Online interface to a 3d mri simulated brain database. *NeuroImage* **5** (1997)
23. Dice, L.R.: Measures of the amount of ecologic association between species. *Ecology* **26** (1945) 297–302
24. Yang, C., Duraiswami, R., Gumerov, N.A., Davis, L.: Improved fast gauss transform and efficient kernel density estimation. In: *Int. Conf. Comp. Vision*. (2003) 464–471