

Enforcing Monotonic Temporal Evolution in Dry Eye Images

Tamir Yedidya¹, Peter Carr¹, Richard Hartley¹, and Jean-Pierre Guillon²

¹ The Australian National University, and National ICT Australia *

² Faculty of Medicine and Health Sciences, Lions Eye Institute, Australia

Abstract. We address the problem of identifying dry areas in the tear film as part of a diagnostic tool for dry-eye syndrome. The requirement is to identify and measure the growth of the dry regions to provide a time-evolving map of degrees of dryness. We segment dry regions using a multi-label graph-cut algorithm on the 3D spatio-temporal volume of frames from a video sequence. To capture the fact that dryness increases over the time of the sequence, we use a time-asymmetric cost function that enforces a constraint that the dryness of each pixel monotonically increases. We demonstrate how this increases our estimation’s reliability and robustness. We tested the method on a set of videos and suggest further re-search using a similar approach.

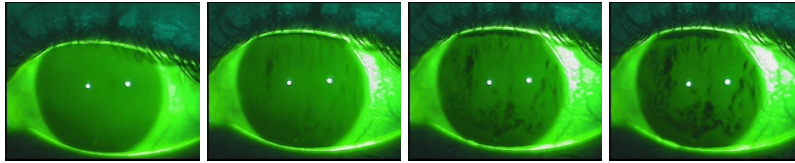


Fig. 1. A sequence of images showing how dryness forms over the iris. The first image is immediately after a blink and then the images are every 4 seconds. The intensity of regions over the iris is related to dryness: the darker the area, the drier it is.

1 Introduction

The pre-ocular tear film in humans does not remain stable for long periods of time [1]. When blinking is prevented, the tear film ruptures and dry spots appear over the cornea. This phenomenon is known as Dry Eye Syndrome [2]. The Fluorescein Break Up Time (FBUT) test was designed by Norn [3] to detect dryness. A small amount of fluorescein is instilled in the patient’s eye. Then, the tear film is viewed with the help of a yellow filter in front of a slit-lamp (see Fig. 1). A video of the front of the eye is recorded between two consecutive blinks. As time passes after the blink, dark areas form on the iris, indicating the lack of fluorescence and the rupture of the tear film. The degree of blackness of these areas is related to the degree of thinning of the tear film. When a dark area of a certain size first appears on the iris, the time elapsed since the blink is recorded as the Break Up time (BUT). If the eyes are kept open, the area of the break will increase in size and breaks may appear in new areas over the cornea. This is the most commonly used test by clinicians to evaluate dry eyes [4].

* National ICT Australia is funded by the Australian Government’s Backing Australia’s Ability initiative, in part through the Australia Research Council

In this paper, we present a graph-cut approach for automatic detection of dryness. We transform the video (after alignment) into a spatio-temporal 3D volume, so a relationship between successive images is defined. The 3D image volume is modeled as a 3-dimensional multi-label Markov Random Field (MRF) in which the label assigned to each pixel represents the degree of dryness. A graph-cut approach benefits from lesser sensitivity to spatial noise and misalignment of the eye images. In addition, we introduce the idea of enforcing temporal monotonicity. This reflects the condition that dry spots on the iris can only become darker (drier) in temporally successive images as seen in Fig. 1. To enforce the increasing dryness condition, we define asymmetric edge weights in the temporal direction, specifying an infinite (or very large) cost to assigning decreasing labels to a pixel in consecutive frames. The associated energy minimization problem is solved using the alpha expansion algorithm [5].

Previous Work. Assessment of dry-eye was reported by us in [6], but in that work we did not use any sort of spatial or temporal constraints. We compare the new results with our previous method.

There has been some work on 3-dimensional segmentation using graph-cuts. Recently Bokyov *et al.* [7] described a global N-D graph-cut segmentation approach that can be used to segment the kidney from a 3D MRI. They are interested in identifying three regions of the kidney and conduct three independent binary segmentations sequentially. However, they do not segment all three regions simultaneously using a multi-label approach. Another example for an application that uses 3D volume binary graph-cuts is for the segmentation of brain tumors [8].

Asymmetric cost functions have not seen widespread use. For example, when employed for spatial geometric constraints [9], alpha-expansion was not able to find a good solution. In [10], the authors use an asymmetric cost to segment multiple surfaces in 3D CT images. Even though the surfaces are segmented simultaneously, they use a binary label set (and not a multi-label approach). To our knowledge, asymmetry has not been used before to enforce temporal constraints within volumetric images.

Motivated by the recent report of the international dry eye workshop (DEWS) [2], we apply the monotonic constraint to the dry eye problem. The report notes the lack of gold standard for diagnosis of dry eye and the need for more robust methods.

2 Formulation of the Problem

We formulate our problem as a second-order MRF. In this approach, each variable i must be assigned a label x_i from the set of labels $\mathcal{L} = \{0, 1, 2, \dots, \ell\}$. The most probable labeling $\mathbf{x}^*$ minimizes the associated energy function:

$$E(\mathbf{x}) = \sum_{i \in \mathcal{P}} E_i(x_i) + \sum_{(i,j) \in \mathcal{N}} E_{ij}(x_i, x_j). \quad (1)$$

Here, $\mathcal{P}$ is the set of pixels in the image and $\mathcal{N}$ is the set of pairs of pixels defined over the standard four-connectedness neighborhood.

The unary terms E_i are application dependent, and we employ a dryness measure similar to [6]. The pairwise terms E_{ij} enforce an *a priori* model. In our application, we expect the labels of neighboring pixels to be the same (or at least quite similar). However, large changes are also possible at edges. Therefore, we employ a function based on

the truncated linear distance (see Fig. 3(a)), which encourages local smoothness, while limiting the cost of large changes to a threshold T :

$$E_{ij}(x_i, x_j) = \lambda \min(|x_i - x_j|, T). \quad (2)$$

The alpha-expansion algorithm [5] can minimize functions of the form (2) as they obey the triangle inequality [5,11]. Although an optimal solution is not guaranteed, in practice the method performs quite well.

2.1 3D Graph Construction

Graph-cut minimization is not limited to 2D and is easily extended to 3D applications. The main advantage of a 3D approach to segmenting individual 2D slices is that the relationship between pixels at consecutive slices is considered. Moreover, it allows one to incorporate monotonic constraints (described in the next section) between slices, which would have been impossible otherwise.

Extending the 2D approach to 3D is based on redefining the neighborhood used in the pairwise term. While in the case of MRI segmentation, it is fairly clear what the individual slices are, we offer an approach based on spatial and temporal progression. Even though the image modality is 2D in the case of the FBUT test, it can be perceived as a 3D approach to capture the global relationship between image frames. Denoting image t in a video of length $n + 1$ as I_t , each image is considered as a horizontal slice in the 3D graph (or MRF), creating a graph based on spatial and temporal changes. Therefore slice number 0 in the graph is the image immediately after the blink and slice n is the last image in the sequence. Every other slice is related to the time passed since the blink. The construction of the graph is based on a 6-connectedness neighborhood $\mathcal{N}$, and an example of a 3D MRF showing the 6-connectedness neighborhood is depicted in Fig. 2. Each voxel in the MRF $(x, y, t), t = \{0, \dots, n\}$ is connected to its four immediate neighbors in the same image and to the corresponding pixel (x, y) in the previous and next frames: $(x, y, t - 1)$ and $(x, y, t + 1)$. Another way to look at the neighborhood of a voxel is: $\mathcal{N} = \{left, right, up, down, next, previous\}$. The energy function is still built only from quadratic terms, as each voxel can be seen as being part of a maximum of 6 pairwise cliques. Each voxel is now also dependent on two voxels which are temporally different. This allows the addition of time based constraints. Denoting the set of pixels of frame t by $\mathcal{P}_t$, the new set of pixels is now defined over the whole image sequence: $\mathcal{P} = \mathcal{P}_0 \cup \dots \cup \mathcal{P}_n$. The hidden nodes of the MRF are the labels assigned to each voxel from the set $\mathcal{L}$.

2.2 Monotonic Constraint

Multi-label problems usually have an inherent meaning to the ordering of the labels. In the case of the FBUT test, the labels represent the estimated thickness of the tear film. The labeling 0 represents no thinning of the tear film and the final label ℓ corresponds to a complete absence of fluid, or a break-up of the tear film. Other labels depict the different degrees of thinning of the tear film. Again, we expect the labels to change

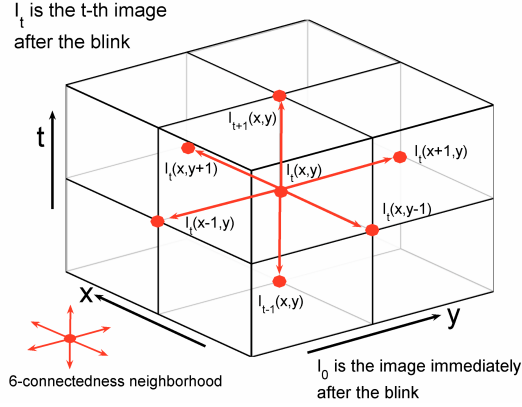


Fig. 2. Explaining the 3D MRF. The x and y axis are the image planes; the t plane is the temporal plane. A pixel $I_t(x, y)$ (or the voxel (x, y, t)) is the pixel (x, y) in the t -th image after the blink. The figure shows the 6-connectedness approach (by the red arrows), where a pixel is connected to its 4 immediate neighbors in the same plane and to two pixels corresponding to the same location at times $t - 1$ and $t + 1$. Each voxel in the MRF can be assigned any label from the set $\mathcal{L}$.

gradually, so we employ a distance metric for the pairwise cost in the temporal axis:

$$f_{ij}(x_i, x_j) = \gamma |x_i - x_j|. \quad (3)$$

As long as the patient does not blink, the thickness of the tear film can not increase between consecutive images. Formally, the label x_j of a particular pixel at time t_j must be less than the corresponding label x_i of the same pixel at time $t_i = t_j + 1$. We enforce this monotonic dryness condition directly into the pairwise energy term:

$$E_{ij}(x_i, x_j) = \begin{cases} \infty & \text{if } t_i = t_j + 1 \text{ and } x_i < x_j \\ f_{ij}(|x_i - x_j|) & \text{otherwise.} \end{cases} \quad (4)$$

We use the truncated linear term of (2) when i and j are a spatial pair ($t_i = t_j$), and the monotonic function of (4) when they are a temporal pair ($t_i = t_j + 1$).

The monotonic function (4) sets an infinite cost to any labeling $\mathbf{x}$ where a pair of labels (x_i, x_j) for a particular pixel at times t_j and $t_i = t_j + 1$ decreases — i.e., $x_i < x_j$. Although we associate an infinite cost for violating monotonicity, in general, a finite cost can be employed.

Fig. 3 shows two examples of pairwise functions which can be minimized using alpha-expansion. Part (a) is a cost function based on (2). The maximum penalty for assigning different labels is bounded by T . Part (b) is a cost function based on (4). If the change of labels is negative, the cost is infinity; Otherwise, the penalty is linear and not truncated. When $x_i = x_j$ the function is assigned 0, however it is not mandatory. In our algorithm, we use the first function for spatially neighboring voxels as a large change between labels should happen at edges. The second function is used for temporally neighboring pixels, where changes in labels (dryness) are usually gradual.

The inclusion of a monotonic constraint makes the pairwise terms asymmetric: the cost of changing from label α to β can be different from changing from β to α , or mathematically $E_{ij}(\alpha, \beta) \neq E_{ij}(\beta, \alpha)$. Alpha-expansion requires the cost function to be metric, however this definition [5] does not include symmetry.

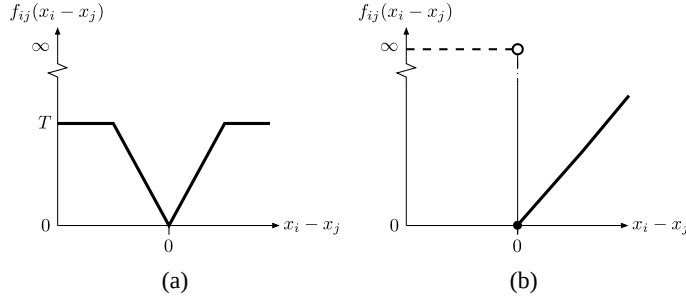


Fig. 3. The spatial (a) and temporal (b) pairwise functions. (a) Neighboring pixels within the same frame are encouraged to have the same label, unless the difference is quite big. In this case, medium and large differences are penalized equally. (b) Temporally, labels must not decrease as time progresses. Moreover, the increase (if any) should not be too large.

3 Application to Detect Dryness

The 2D segmentation approach [6] first detects the iris in each of the video frames. The images are aligned, such that the iris is located roughly at the same location in each image. The segmentation of the dry areas is based on analyzing the aligned video. A cost function examines differences in intensities for each of the pixels in the iris between the first and last images in the video. A dryness image is created, where each pixel is assigned an intensity value which is proportional to its degree of dryness and is denoted by $\tilde{I}(x, y)$. However, the degree of dryness is also computed at each individual slice, and we denote this pixel value by $\tilde{I}(x, y, t)$.

This approach produces good segmentation results and is very fast. Nevertheless, it has a few disadvantages:

1. Small errors in the alignment can completely bias the dryness result for a pixel.
2. The spatial relationships between neighboring pixels in the 2D image are not used.
3. There is no use of the knowledge regarding the temporal change.

3.1 Advantages of The 3D Approach

Given the aligned video created by the 2D segmentation approach, it is possible to incorporate the ideas discussed so far to improve the segmentation results. Instead of looking at individual pixels, and examining every single 2D image for the Break Up Time (BUT), we add the following assumptions:

1. Smoothness constraint - If a pixel becomes dry, it is likely that its neighbors also show a similar degree of dryness.

2. Using temporal knowledge - The video is considered as a 3D volume where each 2D frame is a slice in the 3D image. Segmenting the 3D volume takes into consideration the relationship between the pixel's values at all times.
3. Monotonicity constraint - Temporally, pixels should only become darker, as the amount of fluid in the tear film decreases as time passes. If a pixel becomes brighter it is probably caused by an error in the alignment process or because of shifting of the fluorescein after the blink and not related to the actual dryness.

3.2 Applying the Technique

We show now how the described approach can be easily adopted to the dryness problem. Given the aligned video created in the 2D approach, it is used to create a 3D graph based on temporal changes (see Sec. 2.1). The region of interest in each image is defined as only the pixels belonging to the iris. This region should not be image dependent as after the alignment the iris is resized to the same size at the same location.

The number of labels needed for segmenting dryness depends on the importance of distinguishing between the different degrees of thinning of the tear film. A reasonable choice is to use a set of 9 labels: $\mathcal{L} = \{0, 1, \dots, 8\}$. This number of labels generally produces suitably precise segmentations of the tear film.

The unary term is defined using the value $\tilde{I}(x, y, t)$ computed in [6] for every pixel for every image $t = \{0, \dots, n\}$. When using a multi-label algorithm, a value has to be assigned for each label $E_i(\mathbf{x}_i), x_i \in \mathcal{L}$. The value $\tilde{I}$ can be associated with the expected label x_i^* for each pixel. For example, the intensity range of $\tilde{I}$ can be divided into $|\mathcal{L}|$ equally spaced bins, where each bin is associated with a label. The unary term is then defined as a function h proportional to the difference from the expected label:

$$E_i(x_i) = [h(x_i - x_i^*)]^2. \quad (5)$$

The pairwise term uses linear distance metrics in both the spatial (2) and temporal (4) directions with parameters λ and γ manually tuned to 1. The spatial term is truncated, since large label discontinuities are expected as break-up areas can be local in shape. In the temporal domain, large discontinuities are not expected, so the regular linear distance metric is appropriate. The value of γ is related to the rate of temporally changing labels and can be tuned according to the number of slices in the 3D MRF. The clinical definition of tear film in the FBUT test states that the thickness of the tear film can not increase with time, thus directly encoding a monotonic restriction into (4).

This finalizes the creation of the graph and it is solved using graph-cuts. The labeling for each voxel is its degree of dryness at the time. The labeling of the voxels at time n can be seen as a similar output to the dryness image computed by the 2D approach.

4 Results

To test our method, we used a database of 22 videos with a varying length (4-24 seconds), all having a break of the tear film. Fig. 4(a) shows the result for the sequence in Fig. 1. The brightest areas correspond to areas of maximum thinning. The top slice is the final segmentation result. The t-axis shows the progress of dryness through time.

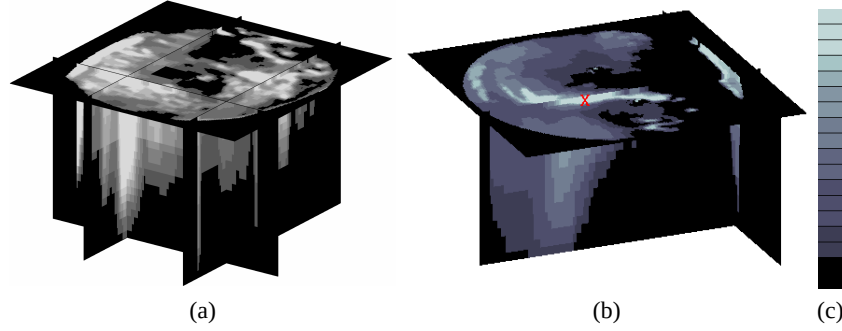


Fig. 4. The top slice shows the final segmentation result. The brighter the colors the more severe the dryness. The t -axis is the temporal axis (not to scale). The 3D view demonstrates how the dryness progresses at specific spatial locations. The monotonic constraint enforces pixels to have non-decreasing intensity. (a) Dryness image of the sequence from Fig. 1 (rotated counterclockwise for clarity). (b) A sequence where the dryness is mainly in the central and superior parts (rotated clockwise). (c) Temporal progress of the voxel highlighted by x in (b)

It can be seen how the monotonic constraint enforces the voxels to have only a non-decreasing intensity and that some of the voxels start showing dryness at a later stage but progress faster. Fig. 4(b) shows a similar cut for a different sequence where the dryness mainly develops in the central and superior areas. The area of dryness in the superior part is quite thin, but the smoothness constraint ensures it is a connected area.

In order to show the contribution of the monotonic constraint, we examined the average number of label changes between every two consecutive slices:

$$\mathcal{C} = 1/|\mathcal{P}_i| \sum_{x \in \mathcal{P}_i} \sum_{t=0}^{t=n-1} |x_i^{t+1} - x_i^t|. \quad (6)$$

We denote the label of pixel i at time t by x_i^t . When using the monotonic constraint, the upper bound for $\mathcal{C}$ is defined by the max number of labels: $\mathcal{C} \leq \ell$. Applying (6) to both methods on all 22 videos, we received an average of 0.906 and 0.523 for the 2D approach and 3D approach respectively. Clearly the new approach is much more robust and smooth. We note that in a few videos, most of the image pixels have no thinning of the tear film at all, so the change of labels is focused in a small number of pixels. Therefore, the difference in $\mathcal{C}$ between the approaches is quite meaningful, as in the 2D approach, individual pixels mainly near the eyelids or the iris's borders, had up to 68(!) label changes and a maximum of 8 when using the monotonic 3D approach. Fig. 5(a) shows another segmentation result using our approach. Parts (b) & (c) are temporal cuts, where the y-axis in these images is progression through time. Notice how near the left end side the monotonic constraint creates a smooth transition between labels with no fluctuations while there is a lot of noise in the other approach.

We asked a clinician to measure the BUT in each of the videos. We then automatically computed the BUT using [6] and then using the new method, considering a break of a pixel when it is assigned the highest label in $\mathcal{L}$. The average difference between

the clinician's BUT and the approaches was 2.4s and 2.34s for the old and new method respectively. Considering the high inter-observer variance, these results are on the acceptable range. However, the new method detected small break areas in 2 videos that were not discovered before. This is due to the tendency of our approach to reduce the number of incorrectly segmented pixels. Thus, the result is less sensitive to outliers allowing the use of a lower threshold for computing the BUT (see [6]).

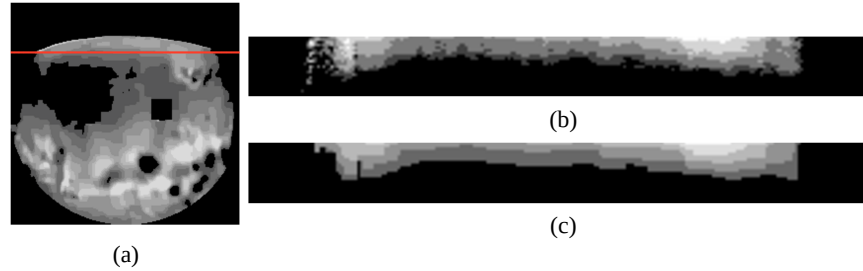


Fig. 5. Example of an x - t cut. (a) Segmentation result using the new approach. An x - t cut (not to scale) at the location of the line in (a) using the (b) 2D approach. (c) 3D approach. The y -axis in (b) & (c) shows temporal progress from bottom to top.

5 Conclusion and Further Research

In this paper, we demonstrated how an asymmetric graph-cuts approach, can be used to segment dryness. The inclusion of a temporally monotonic constraint improves the robustness of the results and reduces the sensitivity to outliers.

The approach presented in this paper can be extended to other medical applications. For instance in segmentation of OCT images of the retina, ordering of the different retinal layers may be enforced using spatially monotonic constraints. In fluoroscopic imaging involving perfusion of contrast agents, temporally monotonic increase and subsequent decrease of intensity may be enforced using an extension of these methods. Finally, spatial relationships and geometric properties such as convexity may be modeled using MRFs with asymmetric edge labels.

References

1. Holly, F.J.: Formation and rupture of the tear film. *Exp Eye Res.* **15** (1973) 515–525 [I](#)
2. Lemp, M.A., Baudouin, C., Dogru, J.B.M., Foulks, G.N., Kinoshita, S., Laibson, P., McCulley, J., Murube, J., Pfugfelder, S.C., Rolando, M., Toda, I.: The definition and classification of dry eye disease: Report of the international dewes. *The Ocular Surface* **5**(2) (2007) [I](#), [II](#)
3. Norn, M.: Desiccation of the pre corneal film. *Acta Ophthal.* **47**(4) (1969) 865–880 [I](#)
4. Korb, D., Craig, J., Dought, M., Guillon, J., Smith, G., Tomlinson, A.: *The Tear Film*. First edn. Butterworth Heinemann (2002) [I](#)
5. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *IEEE Trans. PAMI.* **23**(11) (2001) 1222–1239 [II](#), [III](#), [V](#)
6. Yedidya, T., Hartley, R., Guillon, J.: Automatic detection of pre-ocular tear film break-up sequence in dry eyes. *DICTA* (2008) 442–448 [II](#), [V](#), [VI](#), [VII](#), [VIII](#)

7. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient n-d image segmentation. *Int. J. Comput. Vision* **70**(2) (2006) 109–131 [II](#)
8. Wels, M., Carneiro, G., Aplas, A., Huber, M., Hornegger, J., Comaniciu, D.: A discriminative model-constrained graph cuts approach to fully automated pediatric brain tumor segmentation in 3-d mri. In: *MICCAI*. (2008) 67–75 [II](#)
9. Winn, J., Shotton, J.: The layout consistent random field for recognizing and segmenting partially occluded objects. Volume 1. (2006) 37–44 [II](#)
10. Kang, L., Xiaodong, W., Chen, D., Sonka, M.: Optimal surface segmentation in volumetric images: A graph-theoretic approach. *IEEE Trans. on PAMI* **28**(1) (2006) [II](#)
11. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? *IEEE Trans. PAMI* **26**(2) (2004) 147–159 [III](#)