UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI INGEGNERIA

AUTOMATIC VESSELS SEGMENTATION IN WIDE-FIELD RETINA IMAGES OF INFANTS WITH RETINOPATHY OF PREMATURITY THROUGH SUPERVISED CLASSIFICATION.

Docente relatore: Prof. Alfredo Ruggeri

Correlatore: Enea Poletti

Laureando: Gianluca Callegari

Corso di Laurea in Bioingegneria

Anno Accademico 2011/2012

Index

Summary

Retinopathy of prematurity (ROP) is a disease that can affect premature infants and which can lead to retinal detachment and visual loss if not correctly and promptly healed.

It is characterized by an abnormal development of retinal vasculature, which doesn't reach the periphery of the retina. In fact, the development of retinal network usually becomes complete around full term at 9 months, that is the reason why in premature infants the vessels are not completely developed.

Our purpose is to create a suite of algorithms able to automatically detect the vascular network in the retina and, basing on that, calculate the main parameters of the vessels, as extension, tortuosity and dilatation.

All the pre-existent algorithms we know have been created for the segmentation of adult image, taken with fundus camera. In our program we use images of the infant eye, that have a lower quality and that are taken with a RetCam.

In order to achieve our target, we filter the images at our disposal and, with the results, we create a set of features that a support vector machine (SVM) classifier uses to make a prediction of the vessels in the images.

The results that we obtained can be very useful for the follow up of the extension of the vessels in the retina, that is one of the principal hallmark of ROP; moreover some features can be used for the extraction of other parameters as tortuosity or dilatation of the vessels.

Chapter 1

RETINOPATHY OF PREMATURITY (ROP)

ROP (retinopathy of prematurity) represents one of the leading cause of blindness in infant in developed and emergent countries.

It is fundamental being able to detect ROP promptly in order to diagnose and treat it early, in several cases solving the problem or avoiding serious complication.

1.1 Definition

Retinopathy of prematurity is a disease characterized by an abnormal retina vascularization which can take place in premature newborns. In these babies the growth of retinal blood vessels doesn't reach the whole development, so the peripheral area of the retina aren't vascularizated.

In fact, usually, blood vessels begin their development, starting from the optic disk, during the 15th gestation week continuing their growth and ramification until the 9th month of pregnancy.

Retinal blood vessels grow starting from cells called "spindle cells", which develop from the optic nerve through the ora serrata. Spindle cells reach the ora serrata at 29th week of gestation, whereas the blood vessels develop later and they reach the retinal periphery only during the last weeks of gestation. Consequently, in premature newborns the blood vessels accretion didn't have the time to reach all retina's areas and this lead to several kind of complications.

Retinopathy of prematurity is a complex disease caused by several factors and circumstances so the exact pathogenesis isn't completely clear yet.

1.2 Causes and risk factors

The two main important causes related to the development of this pathology are considered a low birth weight (less than 1500 grams) and a low gestation age (less than 32 weeks). Generally, the more premature is the baby, the deeper is the sickness.

Retinal blood vessels begin to develop 3 months after conception and complete their development at the time of normal birth. If an infant is born very prematurely, eye development can be disrupted. The vessels may stop growing or grow abnormally from the retina into the normally clear gel that fills the back of the eye. The vessels are fragile and can leak, causing bleeding in the eye. Scar tissue may develop and pull the retina low from the inner surface of the eye. In severe cases, this can result in vision loss.

Today, the risk of developing ROP depends on the degree of prematurity. Generally, the smallest and sickest premature babies have the highest risk.

Typically, all babies younger than 30 weeks gestation or weighing fewer than 1500 grams at birth are screened.

Luckily, the rate of ROP in moderately premature infants has decreased dramatically with better care in the neonatal intensive care unit. However, this same intensive care unit has led to high rates of survival of very premature infants who would have had little chance of survival in the past. So these very premature infants are at the highest risk of developing ROP.

1.3 Disease classification

The international ROP classification (ICROP), used in order to code correctly epidemiological data, avails of a specific pattern to classify lesions [1].

ICROP classification is based on several essential observations to do to describe retinopathy, as pathology's location, extension and stage and blood vessels conformation.

1.3.1 Localization

Considering the localization, the ocular fundus has been subdivided in three concentric zones centred on the optic disk:

Zone 1, the posterior zone of the retina, is defined as the circle with a radius extending from the optic nerve to double the distance to the macula. This zone is the most centrally located and ROP develops in this area if the retina in this area is most underdeveloped. Disease in zone 1 is more severe compared with disease limited to zones 2 or 3.

Zone 2 is an annulus with the inner border defined by zone 1 and the outer border defined by the radius that is the distance from the optic nerve to the nasal ora serrata.

Zone 3 is the residual temporal crescent of the retina.

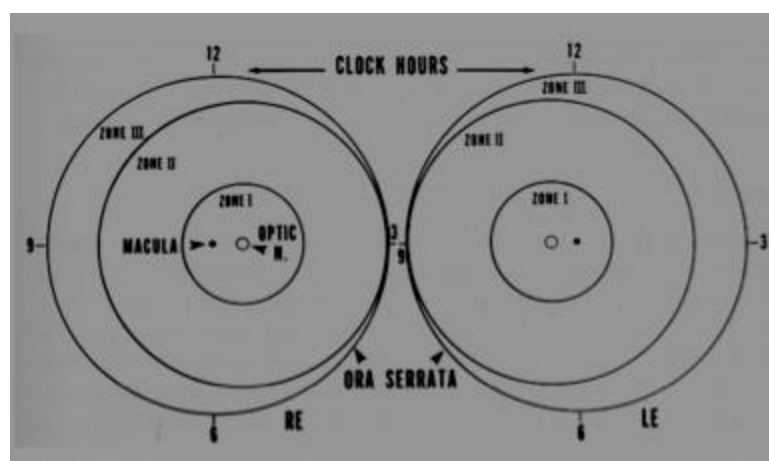We can see the zones of the retina in Figure1.

Figure 1 Retina zones

1.3.2 Extension

Considering the extension, it is evaluated basing on the horary localization as if the top of the eye was 12 on the face of a clock, that is basing on the "number of hours" characterized by pathology signs. As the observer looks at each eye, the 3-o'clock position is to the right and nasal in the right eye and temporal in the left eye, and 9-o'clock position is to the left and temporal in the right eye and nasal in the left eye.

1.3.3 Pathology's stages

There are 5 stages that are usually used to describe the abnormal vascular response at the junction of the vascularized and avascular retina. We show for each stage an image represents it.

ROP STAGE 1: examining the premature newborns ocular fundus the image can show a peripheral area of the retina lacking of blood vessels, really well defined and bordered by an evident demarcation line from the rest of the retina, which is well vascularized. This line is a thin but definite structure that separates the avascular from the vascularized retina. The demarcation line that divides the two areas (one vascularized, the other without blood vessels) can disappear during the following weeks, and the blood vessels can develop in a normal way reaching also the peripheral area of the retina.

Figure 2 ROP stage 1

ROP STAGE 2: demarcation line assumes the morphology of a ridge, this is the hallmark of stage 2 ROP. It arises in the region of the demarcation line but it has height and width and extends above the plane of the retina. The ridge may change from white to pink and vessels may leave the plane of retina posterior to the ridge to enter it.
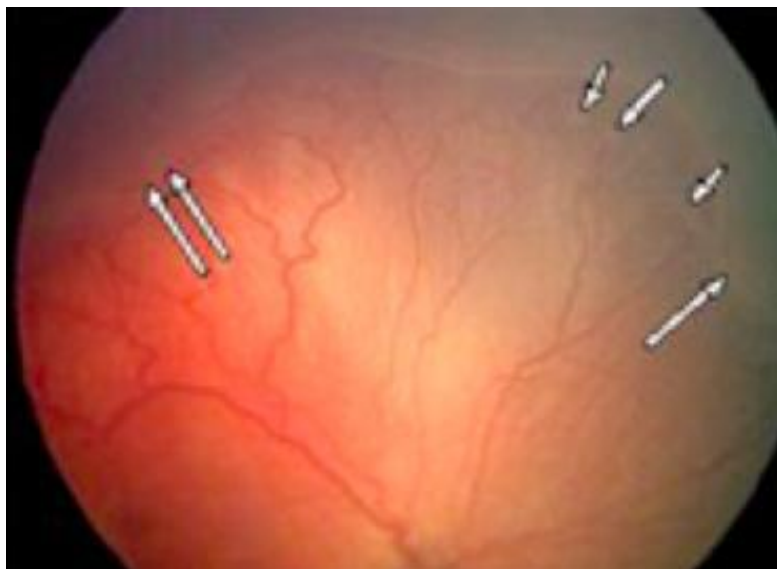

Figure 3 ROP stage 2

ROP STAGE 3: extraretinal fibrovascular proliferation or neovascularization extends from the ridge into the vitreous. This extraretinal proliferating tissue is continuous

with the posterior aspect of the ridge, causing a ragged appearance as the proliferation becomes more extensive.

The severity of a stage 3 lesion can be subdivided into mild, moderate, or severe depending on the extent of extraretinal fibrovascular tissue infiltrating the vitreous.
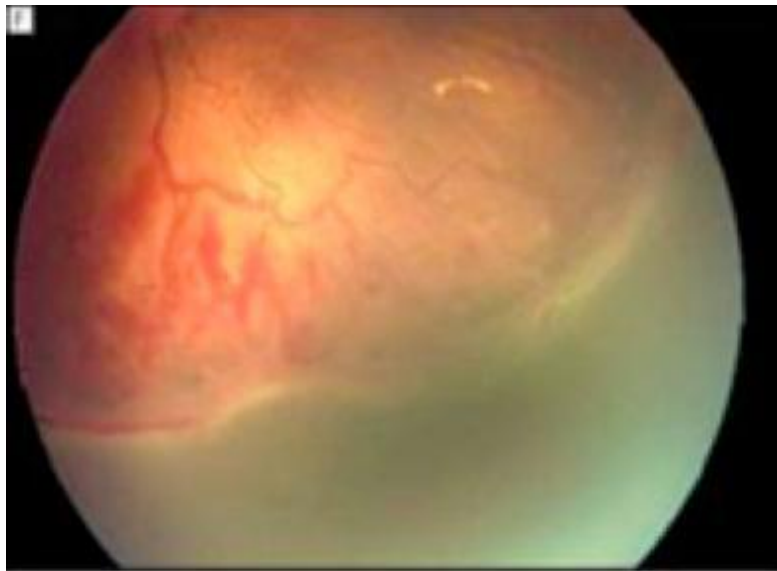
Figure 4 ROP stage 3

ROP STAGE 3 PLUS : if the development of the disease is very fast and the posterior vessels are characterized by tortuosity and dilation associated to retinal hemorrhages, the pathology is classified as stage 3 plus which represents the point of no return for this kind of disease, the moment in which is indispensable to consider the possibility of a surgery intervention. ROP in stage 3 that requires treatment is usually called threshold disease.

Figure 5 ROP stage 3 plus

ROP STAGE 4: it is divided into extrafoveal (stage 4A) and foveal (stage 4B) partial retinal detachments. Stage 4 retinal detachments are generally concave and most are circumferentially oriented. The extent of retinal detachments depends on the number of clock hours of fibrovascular traction and their degree of contraction.

Figure 6 ROP stage 4

ROP STAGE 5: it is characterized by a total retinal detachment. Retinal detachments are generally tractional and may occasionally be exudative.

Figure 7 ROP stage 5

1.3.4 Plus disease

Along with the changes described earlier at the leading edge of the abnormally developing retinal vasculature, additional signs indicating the severity of active ROP may occur. These include increased venous dilatation and arteriolar tortuosity of the posterior retinal vessels and may later increase in severity to include iris vascular engorgement, poor papillary dilatation and vitreous haze. This important constellation of signs in the original classification was referred to as plus disease. Subsequent multicentered clinical trials have used a "standard" photograph to define the minimum amount of vascular dilatation and tortuosity required to make the diagnosis of plus disease. This definition has been further refined in the later clinical trials in which the diagnosis of plus disease could be made if sufficient vascular dilatation and tortuosity are present in at least 2 quadrants of the eye [1]. Plus-disease may be present at any stage.

1.3.5 Pre-plus disease

Pre-plus disease is defined as vascular abnormalities of the posterior pole that are insufficient for the diagnosis of plus disease but that demonstrate more arterial

tortuosity and more venous dilatation than normal. Over time, the vessel abnormalities of pre-plus may progress to plus disease as the vessels dilate and become more tortuous [1].

1.3.6 Aggressive posterior ROP

An uncommon, rapidly progressing, severe form of ROP is designated as AP-ROP. If untreated, it usually progresses to stage 5 ROP. The characteristic features of this type of ROP are its posterior location, prominence of plus disease, and the ill-defined nature of the retinopathy. Early in the development of AP-ROP, the posterior pole vessels show increased dilation and tortuosity in all 4 quadrants and it's often difficult to distinguish between arterioles and venules because of the significant dilation and tortuosity of both vessel types [1].

1.3.7 ROP phases

From this classification it emerges that the disease develops through a progressive process characterized by stages of different severity. In some cases the disease can spontaneously evolve until a complete healing and a natural blood vessels' maturation.

Patients at risk of ROP or affected by ROP are periodically controlled, following a protocol in order to localize the lesion and to monitor the evolution of the disease.

1.4 Diagnosis

Early diagnosis and screening assume a strategic role because they allow to control how the situation evolves and to decide the best moment for an intervention that, in most cases, can solve or strongly reduce the problem.

At the moment the diagnosis is made by observing the retina, after pupillary dilatation using eye drops, by indirect ophtalmoscopy. Examination of the retina of a

premature infant is performed to determine how far the retinal blood vessels have grown (the zone), and whether or not the vessels are growing flat along the wall of the eye (the stage).

The exam can also be realized by digital ophtalmoscopy RetCam, a technology which uses lenses characterized by a wide field of view (130°-160°), that allows the analysis of both the posterior pole and the peripheral area of babies' retina.

1.5 Treatment

Despite of the relevant progress about ROP pathogenesis, nowadays the only effective approach to stop or to reduce the development of the disease is surgical intervention.

Some new approaches are based on administration of substances that control angiogenesis and blood vessels proliferation but these methods are still under study. Timing is one of the most important factors that make the treatment successful in ROP, because the disease can advance very quickly and delayed treatment often reduces the chance of success. The rapidly progressing ROP is called Rush Disease and it is usually associated with very extensive or aggressive growth of abnormal blood vessels. As explained above, abnormal dilatation of retinal veins with florid abnormal new vessels is called Plus Disease.

The goal of the treatment is to destroy the retina that is deprived of retinal vessels. This helps to shrink the new vessels and prevents the formation of dense scars that usually follow. The dense scars cause traction on the retina. The result is distortion of the normal orientation of the retina and impairment of vision. In other cases, the retina detaches from the wall of the eye. The aim of the treatment is to extinguish the seeds of growing abnormal vessels in the eye, ultimately preventing the retina from detaching or from being severely distorted.

Currently the normal treatment in case of ROP in acute phase are [2] cryotherapy and laser therapy.

Cryotherapy has been used since the 1970 to treat ROP. By this method a regional retina destruction is done using a probe to freeze the desired areas. However, cryotherapy is no longer preferred for routine avascular retinal ablation in premature babies, due to the side effects of inflammation and swelling.

Laser Therapy enables the destruction of the avascular retina with the use of a solid state laser photocoagulation device. Laser therapy has shown to be as effective as cryotherapy but with less systemic side effects.

In case of a severe stage of ROP (stage for or 5) cryotherapy and laser therapy are not enough, so it must be done two possible kind of intervention, sclera buckling and vitrectomy.

Scleral buckling can be necessary if a detachment of the retina occurs. It is a procedure that places a band around the globe of the eye this brings the retina back into contact with the inner layers of the eye.

Vitrectomy is used to remove scar tissue from within the eye in the more severe cases of rop. During the vitrectomy, the lens of the eye is usually removed.

Due to the difficulties of performing these surgeries on the small eyes of premature infants, they are usually performed only by a small number of ophtalmologists with extensive experience in ROP.

1.6 Conclusions

Due to the complexity of this pathology, making decisions about the way to proceed in order to limit the damage of ROP development and at the same time avoiding or limiting difficult interventions, it is a challenging task. The possibility of having an automatic device that provides useful information to the clinician about the actual state of the eye under exam would represent an important aid and would improve

the quality of care of premature infants. The realization of a software that automatically detects the stage of ROP basing on RetCam images, would standardize and make more objective the way to proceed and would give more information to the ophtalmologists that could decide how to treat the pathology in an easier way before that complex interventions as vitrectomy or scleral buckling were necessary.

In this work we developed a suite of algorithms that, basing on RetCam images, automatically track the retina blood vessels network. This represents a fundamental step for the realization of a software that is able to calculate indicative parameters as tortuosity and dilation.

Chapter 2

RETCAM IMAGES

So far we described the main features of ROP, underlying who were the infants that have to be checked and the importance of an early detection of the disease and of a periodic screening, to observe how the disease develop and to decide if a surgical intervention is necessary.

In most of the case premature newborns manifest ROP in different stages and indirect ophtalmoscopy doesn't represent a reliable and effective method to diagnose the disease correctly, because the significant area of the retina that as to be examined is the most peripheral. In order to get over these problems, experts use images taken from RetCam, which allows a more accurate way to analyze and categorize the images. Thanks to this instrument the images can also be collected, registered and interchanged.

2.1 RetCam: advantages

The hallmarks of ROP are the dilatation and the tortuosity of the vessels, and these features are usually well visible in the most peripheral area of the eye. In fact this is the zone in which retinal blood vessels grow during the last weeks of gestation.

In order to treat the baby in the best way, it's fundamental to monitor watchfully the retinal condition. The use of RetCam allows the specialists to obtain images of the eye that can be collected and compared at different times, leading to a more effective and reliable process of screening, diagnosis or follow-up of the patient.

Moreover RetCam enables the doctors to have wide-field retina images, due to the wide field of view of the instrument. In this way, ophtalmologists are able to

perform a more exhaustive and objective analysis of the patient condition [3], for this reason RetCam represents an essential device to early diagnosis.

RetCam technology also allows the realization of fluoroangiography, which can represent an important technique during the pre and post surgical follow-up.

2.2 RetCam: drawbacks

Ophtalmologists can come up against some problems using images taken from RetCam; these problems will take them lot of time.

If we compare the images from the RetCam to the images from the fundus camera, we can find various kind of drawbacks, for example: low resolution; low contrast; small dimension of blood vessels; non-uniform illumination; fuzzy areas; low noise-signal ratio; effects due to the fact that the single image is taken by a JPEG video, which manipulates and compresses the single frame, damaging the whole image.

Beyond that, infants eyes present a very flimsy retina that is extremely transparent and through which choroidal vessels are often easily confused as retinal vessels.

We can see these problems in Figure8.

Figura 8 Tipical RetCam Image

2.3 Why retcam?

So, why ophtalmologists use RetCam images even if they present so many drawbacks?

The reason is that RetCam images have a wide field of view (120°-130°), while the fundus camera images usually have a field of view that doesn't exceed 60°. For this reason RetCam is useful if we want to see a wide portion of the retina.

In fact, in order to monitor newborns retina and to diagnose the presence of ROP, its stage and its acuteness, the possibility of examining images of a large area of the eye are indispensable.

Our program represent the first attempt of the use of an automatic segmentation on RetCam images, indeed in the literature we can find segmentation of the fundus images [4], or of the angiography images [5], that is the reason why our program is a total innovation.

The images below show the difference between a fundus camera image of an adult retina and a RetCam image acquired from an infant. These show the difficulty in automatic or non-automatic analysis of RetCam images. All the problems listed before are well-visible and, even at a first sight, we can understand how difficult is to extract the vessels.
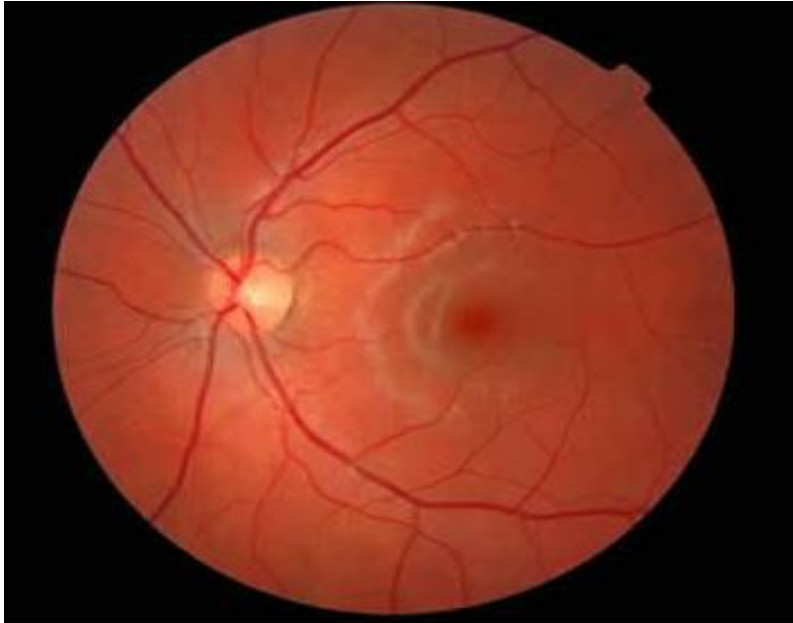
Figure 9 Fundus Camera Image

Figure 10 RetCam Image

2.4 Targets

The main purpose of this work is to overcome the inherent inaccuracies in qualitative evaluation using automated techniques. The final aim is to obtain an

automatic technique which will be able to calculate in a reliable way the parameters as arterial tortuosity and venous dilation, that are determinant to diagnose in an objective and efficient way.

The analysis of RetCam images is the first step to achieve a completely automated tracking of the retinal blood vessels, basing on RetCam images analysis and elaboration.

Using appropriate filter, we have extracted from these images some features that our classifier has used for training and testing set.

In the next chapter we will show what kind of features we have used.

Chapter 3

THE FEATURES

We have extracted from the images at our disposal a lot of features to create the testing and training set for our classifier. The features are our attempt to enhance the characteristics of the pixels represent the vessels; these characteristics make them different from pixels that represent the background.

We can divide the features into two different groups: the first one derived from a simple filtering of the image, while the second one derived from the extraction of the texture of the image.

3.1 Various kind of filtering

In the first step we have calculated the negative of each image and we have applied a morphological opening to it. Then we have summed The result to the negative. We can see the result in Figure11.
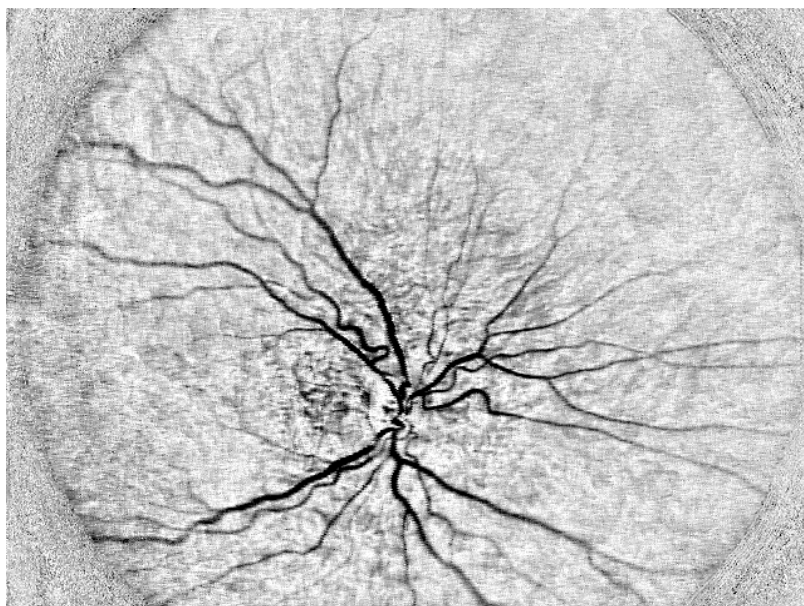

Figure 11 Filter image

In the second step, which is extremely important, we have found the mask of the image, that is the zone in the whole matrix (640x480 pixels) emptied by the retina. The mask looks like a circle and it will be useful because we will make train and test only on the pixels that belong to this particular area.

The mask also allowed us to fill the pixels outside of this zone with values taken from the pixels at the same distance to the borderline. This is necessary if we want to reduce the borderline effects in future predictions. We show a mask in Figure 12.
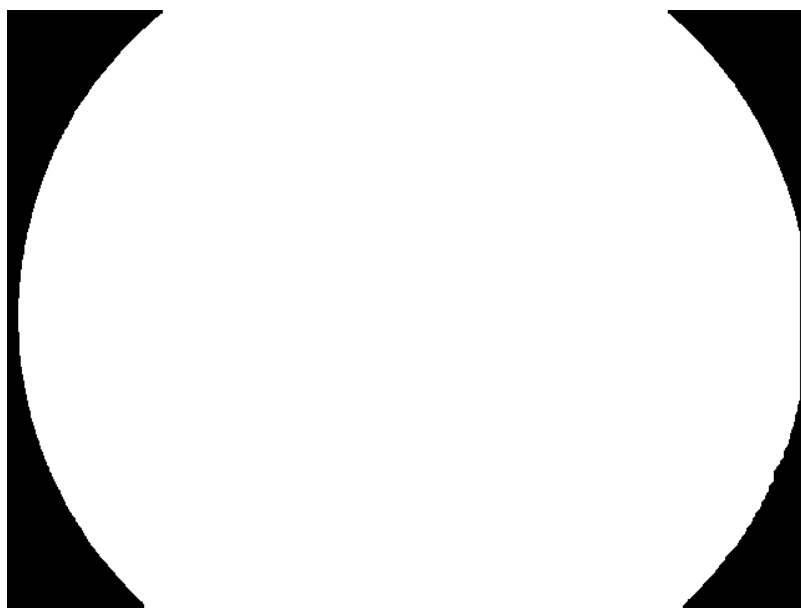

Figure 12 The mask

All these first passages have enabled us to reduce the damages created by the different illumination of the zone on the image, due to the use of RetCam. If we haven't done this procedure, all the further filtering wouldn't be useful.

In order to reduce the same effects we have also tried various frequency filtering, as for example the homomorphus filter or filters that cut various frequency, but no one of these methods brought the same benefits of the morphological operation.

Now we are going to describe the various kind of filter that we have tried.

3.1.1 Ricci-Perfetti filter

This filter is based on Ricci Perfetti's algorithm[6], and give us two different features.

It consists on the creation of numerous rectangular masks that are formed by a line that could have various inclinations. We consider 12 angles (15 degrees of angular resolution) as we can see in Figure13.
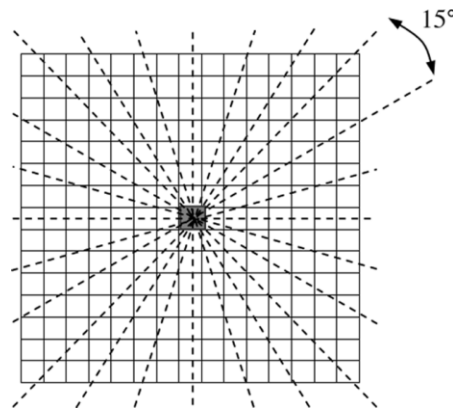

Figure 13 The various inclination

The average grey level is evaluated along these lines passing through the target pixel (x,y). The line with the largest average grey level is found and its value is denoted with L(x,y). If we consider the average grey level in the square window, N(x,y), we can find the line strength, defined as S(x,y)=L(x,y)-N(x,y).

The line strength is large if the winning line is aligned within a vessel; otherwise, we have a partial overlap and the line strength is lower than before, as we see in Figure14.a. This difference allows us to discriminate vessel pixels from non-vessel pixels, so for each pixel we kept the best line strength as the value in that pixel for the first features.

In order to overcome some problems, for example the fact that in the proximity of a large and bright vessel the line strength of a pixel can be comparable with that of the pixels inside a darker and thinner vessel, a line of three pixels is considered, centered on the midpoint of the main line and orthogonal to it, as we can see in Figure14.b.
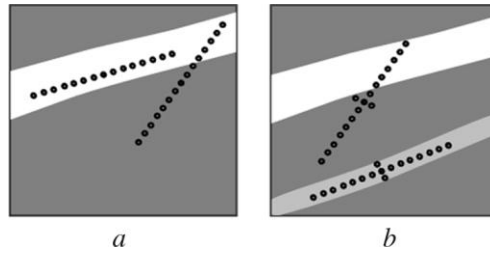
Figure 14 The two kind of lines

We calculate also its average grey level, noted with L0(x,y), and we derived is strength as S0(x,y)=L0(x,y)-N(x,y).

The strength of the shorter line will be negligible for out-of-the-vessel pixels, while it is comparable to the main line strength for inside pixels. This additional information helps to discriminate between inside and outside pixels, decreasing the number of false positives, so the second features of this methods is given by the value of S0 for each pixel.

S, Figure15, and S0, Figure16, will be features that we never use separated, because their information works in pair.

We have tried to change the dimension of the mask, because the original report sets it at 15 pixels, and in our case, due to the lower dimension of the vessels, we have found that the best result is at 7 pixels.
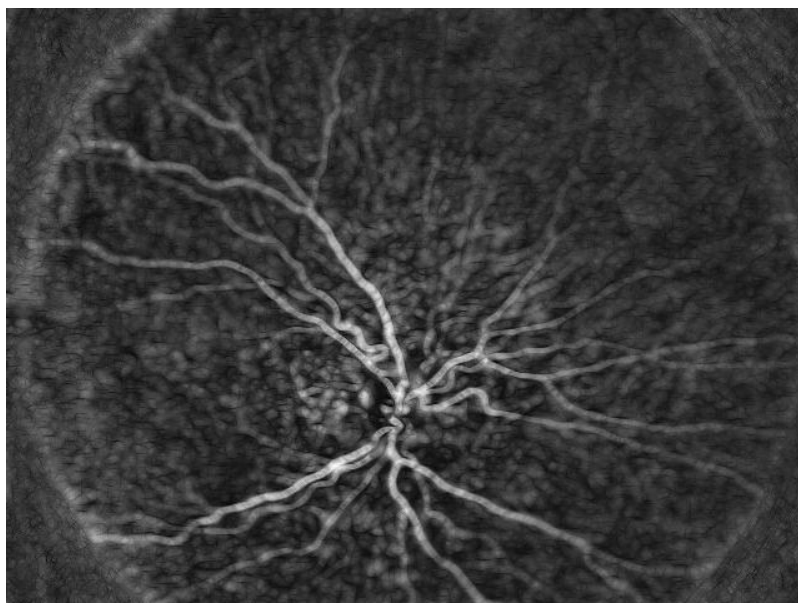

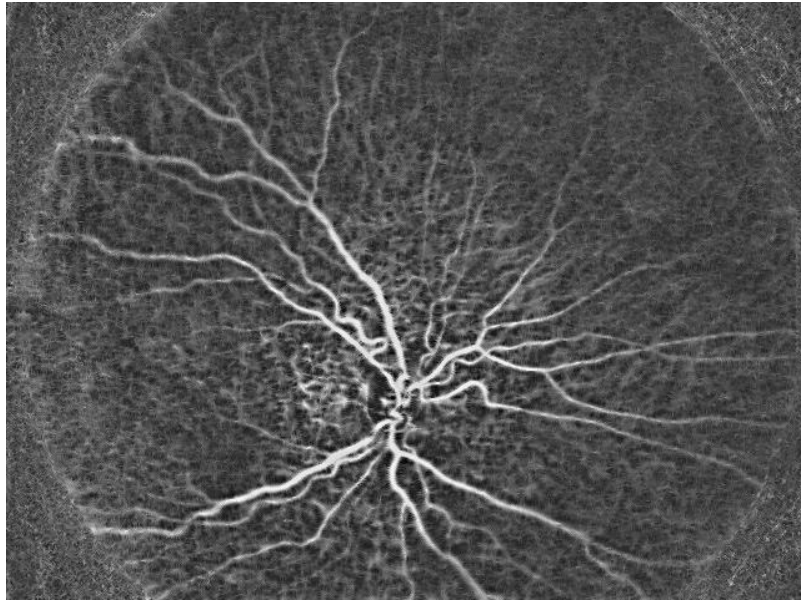**Figure 15 Ricci Perfetti's S feature**

Figure 16 Ricci Perfetti's S0 feature

3.1.2 Hessian filter

We have calculated the hessian matrix of the picture, which has the propriety to stand out the structure in the image with a clearly defined directionality. We have done this because this kind of structures are most vessels.

Hessian is able to find this propriety on an image because it finds in the second derivate of the imagine (the hessian matrix) the maximum eigenvalue and the corresponding eigenvector, which indicates the present of a structure with directionality.

The result of the operation has been cleaned with a morphologic opening that was necessary to delete the little blur due to little structure in the original image. We see the result in Figure17.
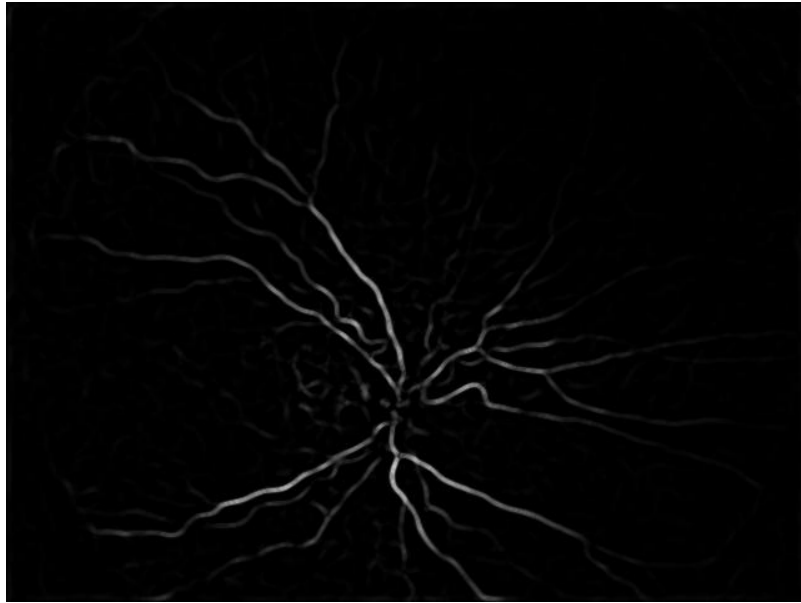
Figure 17 Hessian filter of the image

3.1.3 Matched filter

This filter is based on a research of the vessels based on their shape[7].

As we see in the Figure18, the grey profile of different vessels has always a similar shape that consists in a sort of valley, which becomes deeper the more the pixels are in the centre of the vessel.
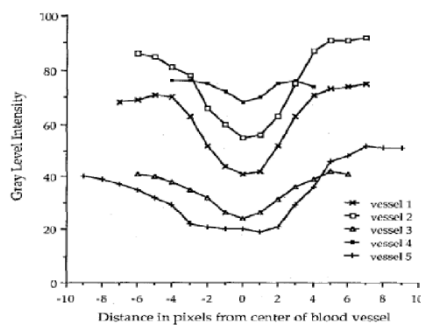

Figure 18 Vessels profile

This shape can be approximated with a Gaussian function or something similar like a Laplacian or a Laplacian of Gaussian. In this way, for every shape we have created a set of masks with the function that we chose before, the one with different angle, as in Ricci Perfetti's filter case. We have convolved the various set of masks with the

28

negative of the images (for maximization reasons) and we have taken the best result get from the different inclination of the function from each pixel.

At the end, we have chosen the function that gives the best exaltation of the vessels in the images and that was the Laplacian of Gaussian. The result is shown in Figure19.
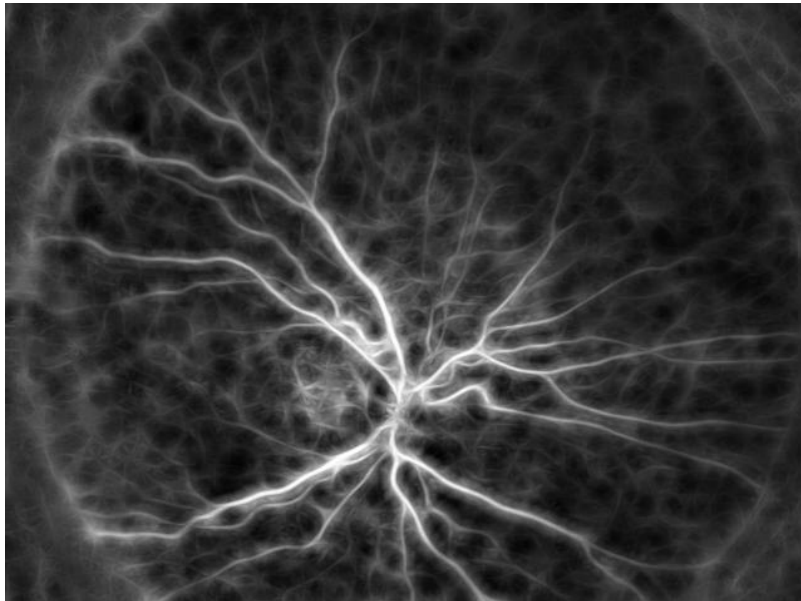

Figure 19 Matched filter result

3.1.4 Filter combination

We have tried a combination of filters as suggest in [4]. The idea was to apply a filter on the result of a different kind of filter, which can improve the characteristic of the vessel better than only one filter.

First of all for every test we have filtered the image with a bilateral filter, in order to reduce the noise without damaging the edge-information in the images.

3.1.4.1 Matched of Hessian

The first test was based on the application of a matched filter on the result of a hessian filter. This trying is a little raw, but gave some useful results.

3.1.4.2 Hessian of matched

We also tried the reverse of the precedent step, but it have given us various problems, because the hessian filter after the matched filter exalts lot of structures that don't exist in the original image, but have been created by the hessian.

3.1.4.3 Big combination

Starting from 3.1.4.1 we have tried to apply another time the hessian filter and the result was good, because it exalts the vessels and reduces the other little structure. So we have repeated the passages and the result was acceptable, but it started to show some non-realistic structure and also start to reduce the visibility of the bifurcation of the vessels. We have also tried the third test, but it was worst.

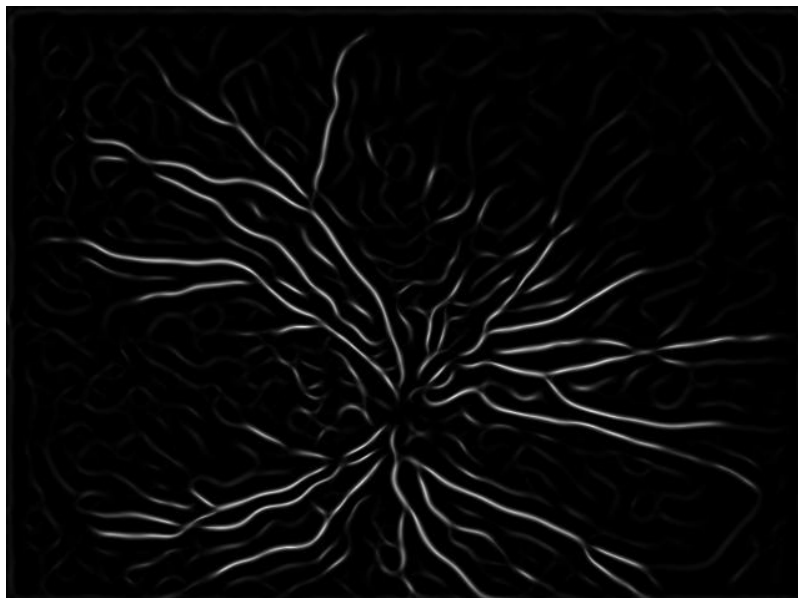After that, we kept the second step as the best, as we see in Figure20.

Figure 20 Combination of filter

3.1.4.4 Laplacian of Gaussian

We have also tried to used the Laplacian of Gaussian of the images as features, but it was useless.

The main reason that caused its failure is that this kind of filter is used for the exaltation of the edge of the image, and this penalizes it in our program, because edge comes in evidence but the inside of the vessel, especially of the big ones, takes the same value of the background, so the classifier won't be able to recognize it.

We have also tried to sum the result of this filter to the image, but it didn't make anything useful.

3.1.4.5 Gradient

We have tried also to calculated the gradient of the images, but it had the same drawbacks of the 3.1.5, so we have discard also this.

3.2 Textures

What is a texture? It doesn't exist an uniform definition of texture, but when we see it, we can recognize it. We have collected some definitions in order to clarify the context:

- We may regard texture as what constitutes a macroscopic region. Its structure is simply attributed to the repetitive patterns in which elements or primitives are arranged according to a placement rule[8].

- Region in an image has a constant texture if a set of local statistics or other local properties of the picture function are constant, slowly varying, or approximately periodic[9]

- The image texture we consider is nonfigurative and cellular... An image texture is described by the number and types of its (tonal) primitives and the spatial organization or layout of its (tonal) primitives... A fundamental characteristic of texture: it cannot be analyzed without a frame of reference of tonal primitive being stated or implied. For any smooth gray-tone surface, there exists a scale such that when the surface is examined, it has no texture.

Then as resolution increases, it takes on a fine texture and then a coarse texture[10]

- Texture is defined for our purposes as an attribute of a field having no components that appear enumerable. The phase relations between the components are thus not apparent. Nor should the field contain an obvious gradient. The intent of this definition is to direct attention of the observer to the global properties of the display — i.e., its overall "coarseness," "bumpiness," or "fineness." Physically, nonenumerable (aperiodic) patterns are generated by stochastic as opposed to deterministic processes. Perceptually, however, the set of all patterns without obvious enumerable components will include many deterministic (and even periodic) textures[11]

- Texture is an apparently paradoxical notion. On the one hand, it is commonly used in the early processing of visual information, especially for practical classification purposes. On the other hand, no one has succeeded in producing a commonly accepted definition of texture. The resolution of this paradox, we feel, will depend on a richer, more developed model for early visual information processing, a central aspect of which will be representational systems at many different levels of abstraction. These levels will most probably include actual intensities at the bottom and will progress through edge and orientation descriptors to surface, and perhaps volumetric descriptors. Given these multi-level structures, it seems clear that they should be included in the definition of, and in the computation of, texture descriptors [12].

This collection of descriptions demonstrates that the "definition" of texture is formulated by different people depending on the particular application and that

there is no generally agreed upon. Some are perceptually motivated, and others are driven completely by the application in which the definition will be used.

Textures in our contest would make us able to keep from image some characteristics that other kind of filters are not able to extract. In fact at the contrary of the usual filter, the extraction of textures doesn't manipulate the images, but extract only information. For example a zone with a vessel would have different texture from a zone without vessel.

For this reason our first attempt was to evaluate the textures on the green and on the blue channel of the original image. We have used these channels because they would have the best information about the vessel.

Our hope was that these results or their combinations would have been sufficient for our classifier to see the vessels, but soon we have noticed that the textures have some problems.

The first textures were affected by the errors due to the non-homogeneous illumination in the images; for this reason we tried to calculated them on the green channel filtered with the morphological opening and with the bilateral filter. We have used these results as features in the classifier.

We now describe every single texture that we have extract; each of them was evaluated in a square window around the pixel that could have dimension of 3x3 pixels, 5x5 or 7x7.

3.2.1 Mean

We have done the mean in the square windows around the pixel. This reduces the noise but it damages the edge and the information on the small vessels. We show it in Figure21.
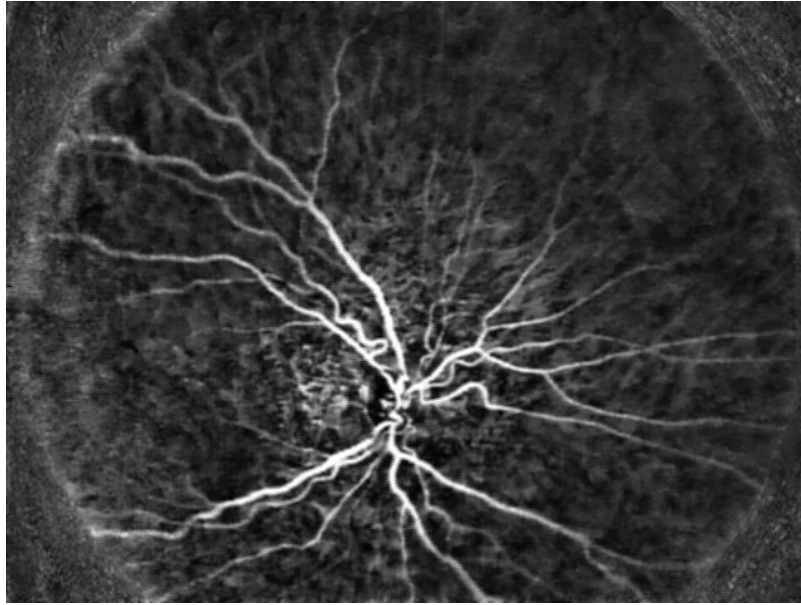
Figure 21 Mean texture

3.2.2 Standard deviation

We have calculated the standard deviation on the rectangular windows. This emphasizes the zone near the edge because into the window there are pixels of a lot of grey levels that pull up the standard deviation. Unfortunately, if the window is too little and falls all inside the vessel, it gives the same value of the background as we can see in Figure22.
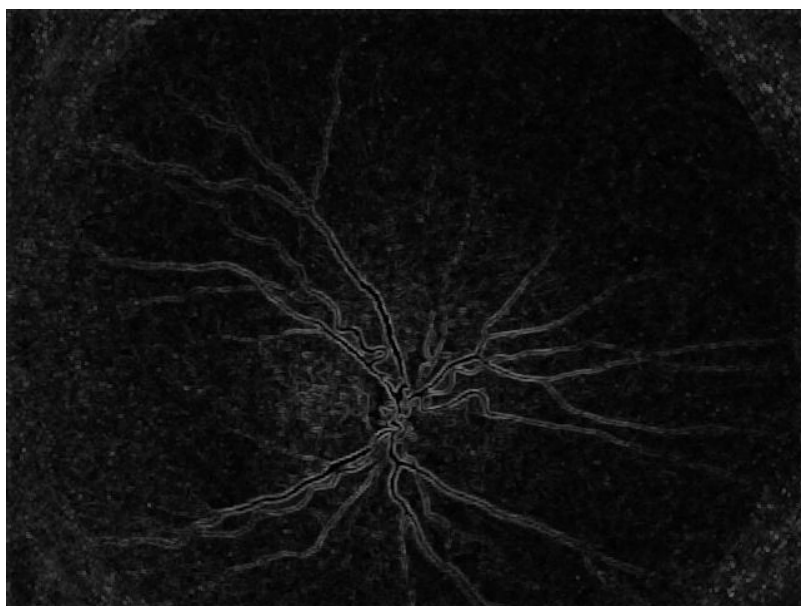

Figure 22 Standard deviation

3.2.3 Kurtosis

Kurtosis is a measure of the "peakedness" of the probability distribution of a real-valued random variable, that in our situation is the grey level of the pixels. Kurtosis is a descriptor of the shape of a probability distribution.

Even for this parameter we have seen that we had different values near the vessels because the presence in the same window of vessel pixels and background pixels cause a not-normal distribution.

We can see it in Figure23.

Figure 23 Kurtosis

3.2.4 Skewness

Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable, that is always the grey level of the pixels. The skewness value can be positive or negative, or even undefined. Qualitatively, a negative skew indicates that the tail on the left side of the probability density function is longer than the right side and the bulk of the values lie to the right of the mean. A positive skew indicates that the tail on the right side is longer than the left side and the bulk of the values lie to the left of the mean. A zero value indicates that the

values are relatively evenly distributed on both sides of the mean, typically but not necessarily implying a symmetric distribution.

As in the previous cases we see the values of skewness have changes near the vessels, where the distribution is not gaussian.
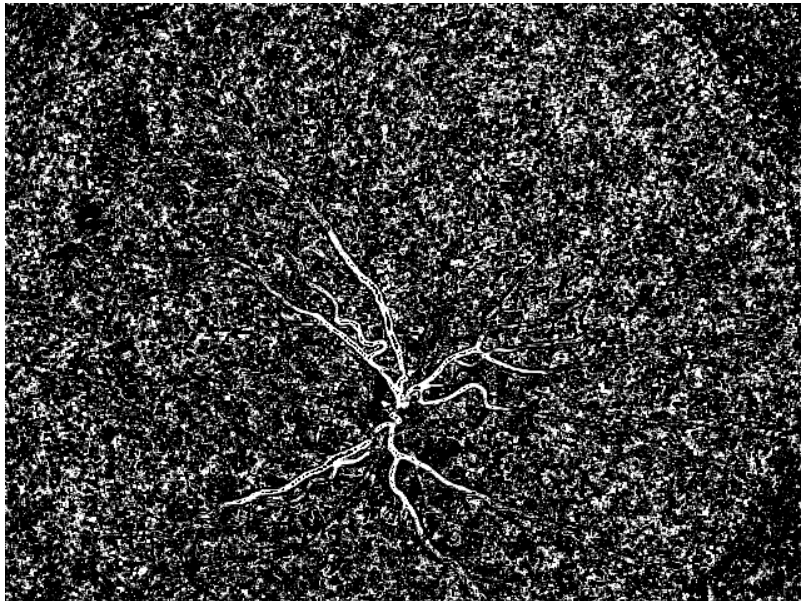
We can see it in Figure24.

Figure 24 Skewness

3.2.5 Entropy

Entropy is a statistical measure of the average information content one is missing when one does not know the value of the random variable. In our contest entropy is derived from the histogram of the image, because our variable is the grey value of the pixels.

It gives us a measure about the disorder in the image, so it would change near the vessels, as we can see in Figure25.
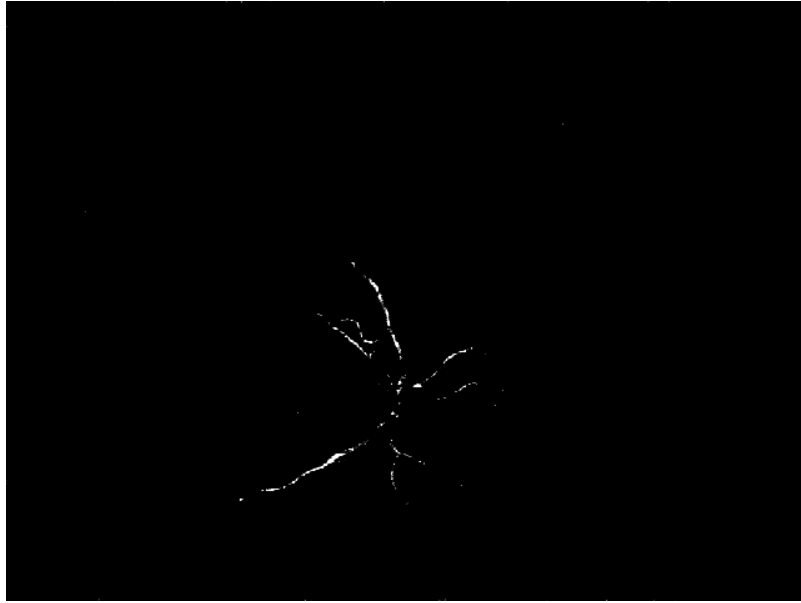
Figure 25 Entropy

3.2.6 Textures based on co-occurence matrix.

A co-occurrence matrix is a matrix that is defined over an image to be the distribution of co-occurring values at a given offset. It gives a tabulation of how often different combinations of pixel brightness values, the grey levels, occur in an image[13].

We have used it for the calculation of the second order texture; those textures consider the relationship between groups of two pixels in the image.

Between the various kind of second order texture we have used contrast, homogeneity, correlation and energy.

3.2.6.1 Contrast

Contrast is a linear combination of two or more factor level averages whose coefficients add up to zero.

In the co-occurrence matrix we calculated it as:

$$\sum_{i,j=0}^{N-1} P_{i,j}\,(i-j)^2$$

where *i* and *j* are the index in the matrix.

When *i* and *j* are equal, the cell is on the diagonal and (*i-j*)=0. These values represent pixels entirely similar to their neighbor, so they are given a weight of 0.

If *i* and *j* differ by 1, there is a small contrast, and the weight is 1.If *i* and *j* differ by 2, contrast is increasing and the weight is 4. The weights continue to increase exponentially as (*i-j*) increases. We can see it in Figure26.
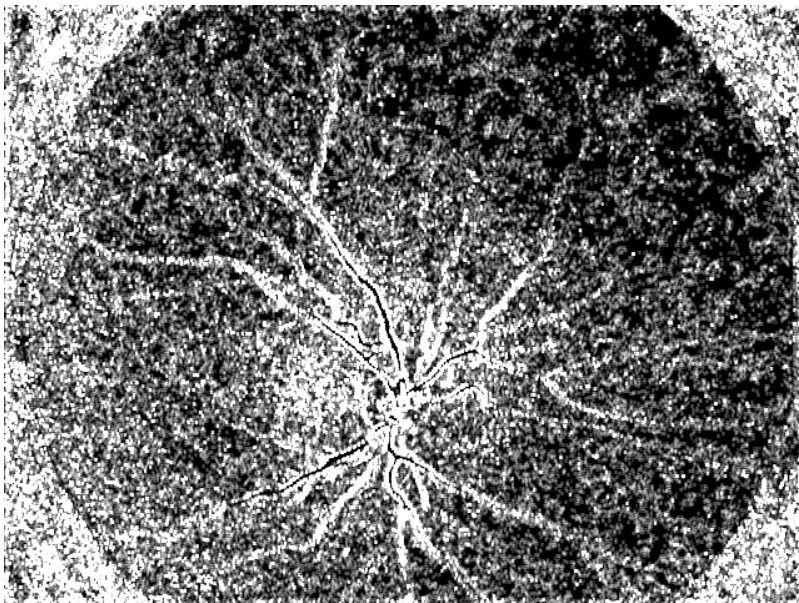

Figure 26 Contrast

3.2.6.2 Homogeneity

Homogeneity examines how much the variability of data-values changes throughout a dataset. In our case, it measures the variability of the grey level of the pixel.

In the co-occurrence matrix we calculated it as:

$$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2}$$

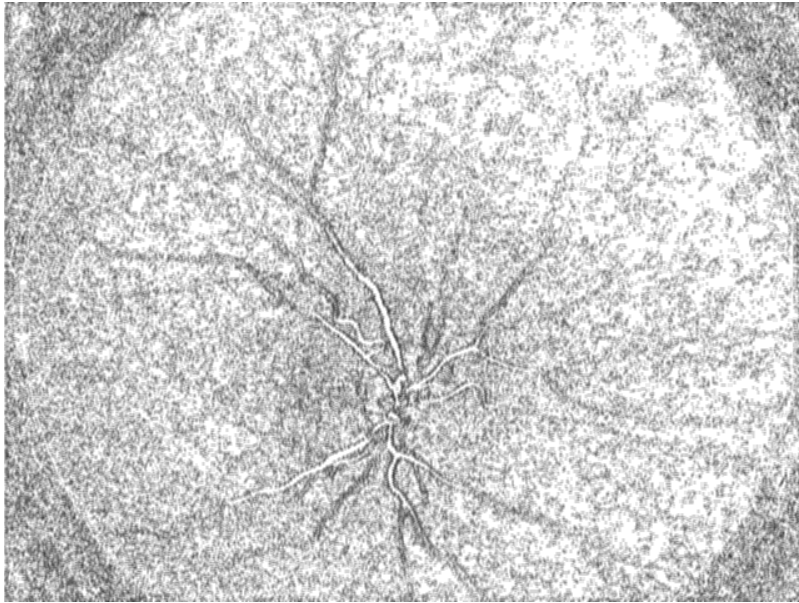We notice that homogeneity weights values by the inverse of the contrast weight. We can see it in Figure27.

Figure 27 Homogeneity

3.2.6.3 Correlation

Correlation measures the linear dependency of grey levels on those of neighboring pixels. If it is 0 the pixels are uncorrelated, if it is 1 they are perfectly correlated.

In the co-occurrence matrix we calculated it as:

$$\sum_{i,j=0}^{N-1} P_{i,j} \frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}}$$

where μ_i, μ_j are the mean in *i* and in *j*, and σ_i, σ_j are the respective standard deviation.
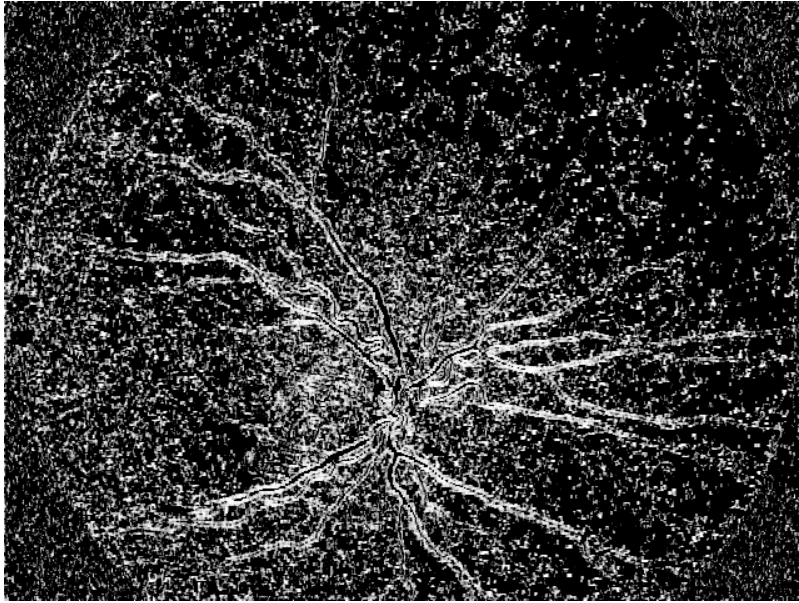
We can see it in Figure28.

Figure 28 Correlation

3.2.6.4 Energy

Energy measures the "power" of the grey level.

In the co-occurrence matrix we calculated it as:

$$\sum_{i,j=0}^{N-1} P(i,j)^2$$
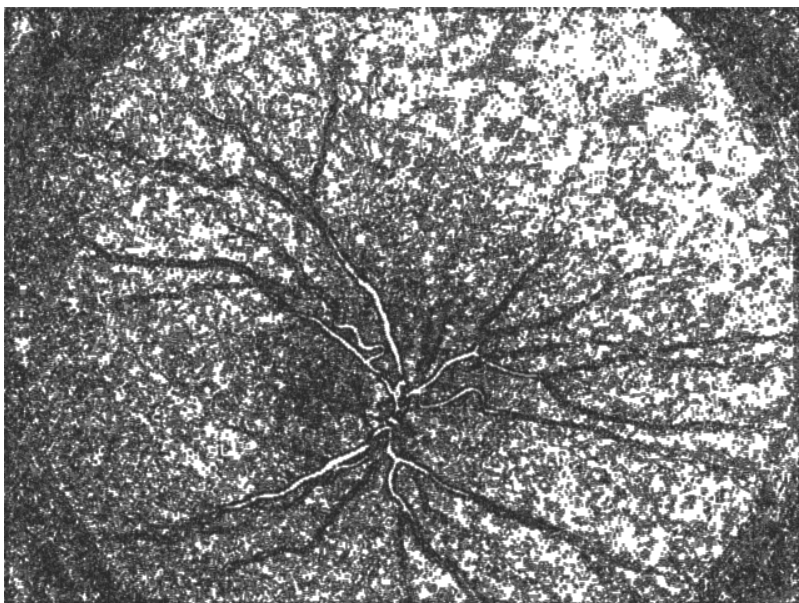
We can see it in Figure29.

Figure 29 Energy

3.3 Conclusions

All the features we have extracted from the images have the same target: emphasize the vessels. It is important to have numerous features, because they are evaluated with different methods, and this fact would help the classifier to recognize the vessels, because every single method would compensate the errors of the others.

Now the problem is the choice of the best set of features, because not all are useful for the classifier.

In the next chapter we will describe how the classifier works, and afterwards we will see how the choice of the features has been made.

Chapter 4

The classifier

We have extracted all the features from the images and in this way we collected a lot of different information.

Our purpose was to use this information in the best possible way in order to distinguish which pixels in the images are vessels and which are background.

In this situation the best method to make a decision is to use a classifier.

The classifier is a program that is able to analyse and group different elements in different classes. In our situation the elements that have to be assign are the pixels and the classes are two: vessel and background.

4.1 SVM Classifier

In our pixel classification we used a classifier called support vector machine (SVM) classifier [14]. It is free, and it is implementable on every computing program such as Matlab, Python, C++, and others.

SVM is a supervised classifier and for this reason it needs a ground truth to do the test. This is an hand-made image, in which we mark the pixels that belong to the vessels and the pixels that belong to the background. We show a ground truth in Figure30 and the original corresponding image in Figure31.

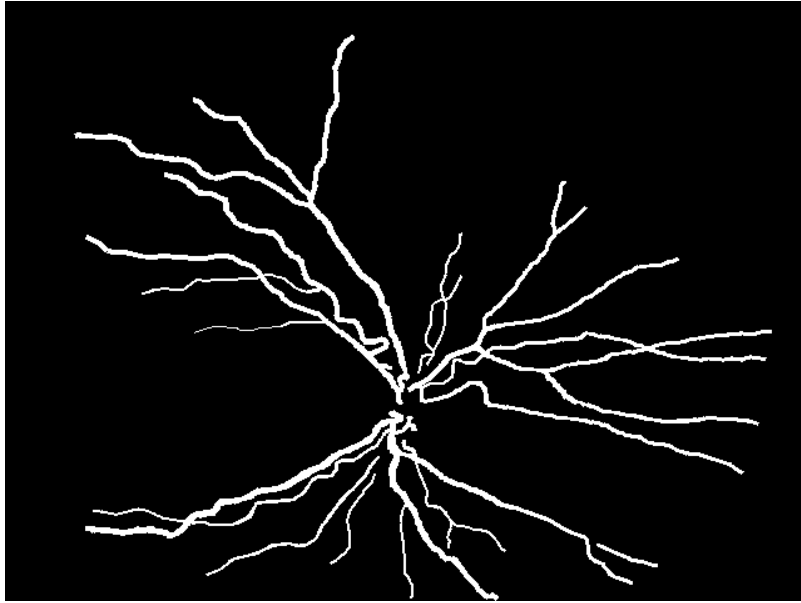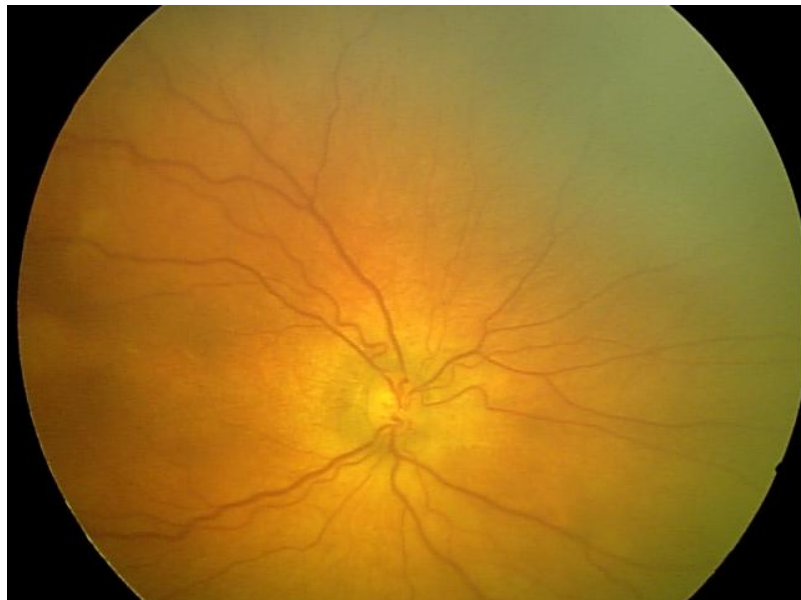**Figure 30 Ground truth**

Figure 31 Original Image

From the combination of the ground truth and the training set of features, this classifier learns the difference between the characteristic of the vessels and the characteristic of the background.

It resumes all its knowledge in the creation of a model, which the classifier will use while it is making the test on the features of the other images.

One of the advantages of SVM is that it maps the training vectors into a higher (maybe infinite) dimensional space, and always finds a linear separating hyperplane with the maximal margin in this higher dimensional space. That is the reason why SVM is a linear classifier.

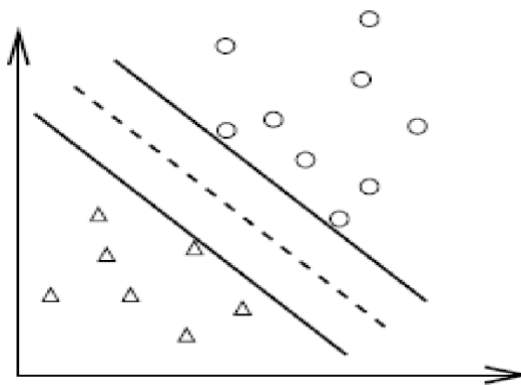In Figure32 we can see the separation of the features in two different classes in their space.

Figure 32 Separation of the features

4.2 Validation

In the previous section we have shown that the classifier needs a training set in order to build the model that enables it to distinguish the element.

If we want to achieve the best precision, we have to make the training on the biggest number of images that is possible, but if we do this process on the whole set of images, we will not have any image to test the results of the classifier, and we can't validate these results.

In order to overcome this problem we used a method called n-fold validation. This process consists in making the division in n parts of our set of images. Every part is a fold and it contains 20 divided by n images. After this operation we train the classifier on the first fold and test on the others in order to have the results; afterwards, we train on the second fold and test on the others, in this way we can

have the second part of the results, and so on. At the end, we calculate the mean of the various results and it gives us the performance of our classifier.

In our case we have used a 2-fold validation, which means that we have divided our images in two folds, the first one with the first ten images, and the second one with the others.

4.3 Setting of parameters

The classifier, such as every kind of program, bases his job on some parameters that can influence its actions.

In the SVM the parameters concern a lot of things, as for example the choice of kernel, the tolerance of the errors due to the position of the hyperplane, the weight of the classes, the criterion of halting the program, the step for the various iterations, and others.

The best way to set them is to try every possible combination, but this is obviously impossible for computational reasons. We have tried to set them separately, we found the best results for every single parameter, and then we used them all together.

The most important parameters were the two which manipulated the kernel of the classifier. In order to have the best set of these coefficients, we have tested their combinations using only few pixels in the images. Thanks to this method we have improved a lot our results.

4.4 Conclusions

In this chapter we have shown the principal characteristic of the classifier that we have chosen and how it works. The following step is to use it on the features.

In the next chapter we will see how fundamental is the best choice of the features, because the performances of the classifier depend most of it.

Chapter 5

Features selection

Features selection has been certainly the finest passage of our work, because the quality of the result depends most on it.

We had not only to choose the best features, but also to decide their number, because the use of too many features would cause damages as well as the use of not many features.

5.1 Prerequisites

The first step has been the scaling of the features, as the creators of the classifier suggest[14]. Without this passage the classifier loses a lot of time and the results are wrong.

Once we have done this passage, we have to elaborated every feature so the vessels of each matrix will be enhanced in the same way. Our purpose is to render all the vessel brighter (or lower) than the other parts of the feature, because if some features have the vessels bright and other have the vessel dark the classifier doesn't work at best.

After these operations, we have tidied up the features and we could start with a first visual selection. During this process we cut out those features that clearly couldn't help the classifier, because they were not able to enhance the vessels. In Figure33 we can see an example.
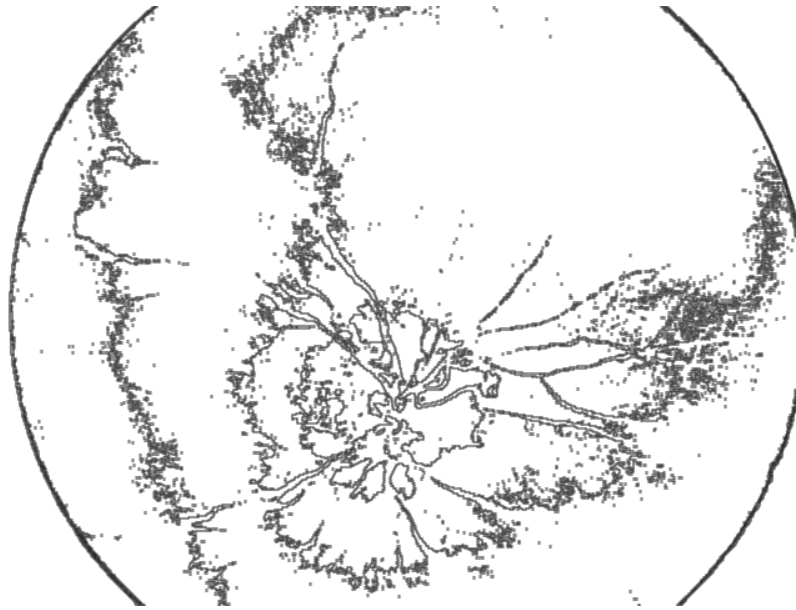
Figure 33 Useless features

With this passage we have eliminated the worst features, and thanks to it we could start the real selection.

5.2 Selection methods

Our first idea was to use the correlation between the features as parameter for the selection, but we have soon discarded it. We did this because the correlation has not given us sure information, as we have already expected from[15], and because in our situation all the features had an high correlation between them, so it has not been discriminant.

Furthermore we have also discarded the idea of an exhaustive research of the best combination of the features, because the high number of features in our posses would have taken the program to a huge number of trying (in fact exhaustive research is an NP-hard problem).

In this situation we decided to use a greedy algorithm that has given us a sub-optimal solution. We could chose between two kind of algorithms that are supported by the same idea of a progressive research of the best combination of features. These two methods are the forward research (called wrap) and the

backward research (called unwrap). Both of them consider the learning machine a perfect black box and they work only on the results that the classifier gives.

The wrap methodology proceeds in this way: we choose one feature and we use only it for the train and for the test of the classifier; we do this with all the features. Then we keep only the feature with the best result: this will be our start for the next round. Indeed, once a time, we add to the selected feature all the other features, so we obtain couples of features and we use them for the train and the test of the classifier. We compare the results again and we keep only the couple with the best result. Then we proceed with set of three features in the same way and so on.

We stop only when we see that the new set of features doesn't improve the previous results, or it improves only less than a threshold that we have chosen.

When we stop we have the sub-optimal combination of features that will be used in the classification.

The unwrap methodology is very similar to the wrap, but it proceed in the opposite way. If we would have used it, we would have start from the whole set of features and discard once feature a time and see the various results. In this way we would have taken the best set and proceed with the discarding until we could stop because the results don't improve.

We have chosen the wrap method, because it is faster than the unwrap.

Using this method, we probably don't have the best features set, but anyway it is a good method to discard the worst and useless features, and we also bypass the correlation method because the features are used together and they give us the certainty of their usefulness in group.

We can notice that this method against the exhaustive research pulls down the number of trying from $\sum_m \binom{n}{m}$ to $n+(n\text{-}1)+(n\text{-}2)+..+(n\text{-}m+1)$ where n is the total number of features and m is the number of features we use at the end of the

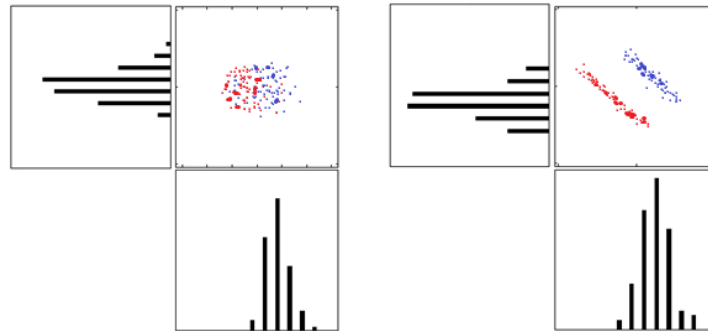research. In Figure34 we show a set of two useless features against a set of two useful features.

Figure 34 Useless vs useful features

5.3 ROC

Another method for the evaluation of the usefulness of a feature is to calculate its receiver operating characteristic (ROC).

ROC is the curve in which is represented the true positive rate against the false positive rate for a binary classifier system as its discrimination threshold is varied.

In our contest true positive pixels are those one that belong to the vessels and that the classifier sets as pixels of the background; false positive pixels are those one that belong to the background and that the classifier sets as pixels of the vessels.

The discrimination threshold is the grey level that we chose for the conversion of the image from an RGB representation to a black and white representation. All the pixels that have grey level higher than the threshold are set as vessels, the other as background.

We have started the threshold at zeros, so all the pixels were set as vessels and we have increased it until is equal to one, where all the pixels were set as background. At every step we have considered the true positive rate and the false positive rate. We could use zero and one as lower and upper bound values, because all the features has been normalized.

If we plot one against the other the two rates, we obtain a graphic as we show in Figure35.

Here the red line represents the ROC of one feature and the green line represents the ROC of the original image.
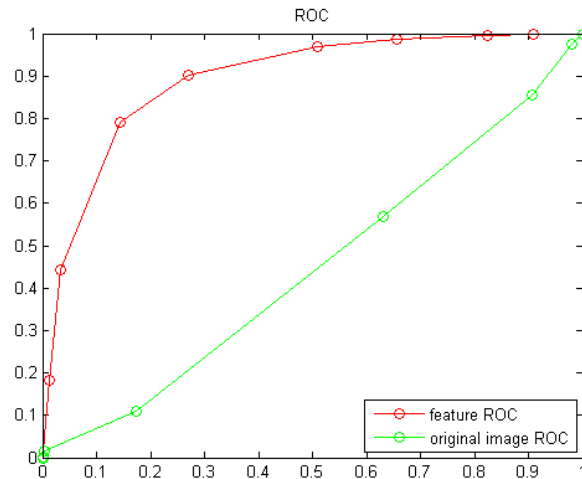

Figure 35 ROC

ROC helps us to make another idea of the usefulness of the single feature because if we compare two ROCs, the one that grows earlier indicates that its feature is better than the other.

5.4 Conclusions

Finally we have used our classification on the set of features that we have extract. For timing reasons we have used the department cluster for making the training and the testing, because both of them, especially the training, are very heavy computational operations that a normal computer is not able to do in reasonable time.

In the next chapter we will show the results of our work.

Chapter 6

Results

In the previous chapter we have shown how we choose the best set of features for the classification. Once we have done it, we could use the classifier for making the training on the two folds. This was the fundamental operation for the SVM, which at the end returned a model where all the parameters, that the test needs, are saved. Afterwards, we made the testing for each image, using always the best set of features. At the end of this operation the classifier returned two matrices, that we could use to extract our results.

6.1 Classifier output

At the end of its job, the classifier gives back two kind of output.

The first one is a black and white image, which has the same dimensions of the original picture and which represents the final prevision of the classifier; the white pixels are those that the SVM has recognized as vessels pixels, while the black ones are those that it has recognized as background pixels.

We can use this matrix as the final result, but it gives us only the final step of the classifier. We show an example in Figure36.
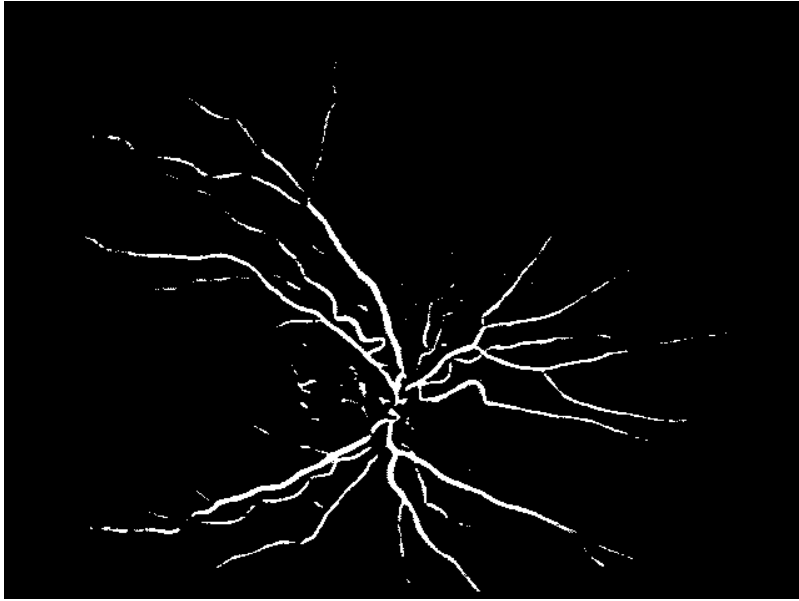
Figure 36 SVM prediction

The second result returned by the SVM is another matrix with the same dimension
of the image, but it shows the representation of the probability level that every
single pixel has to belong to one of the class. This can be very useful because we can
work on it and extract the results in a different way from the classifier. We can see it
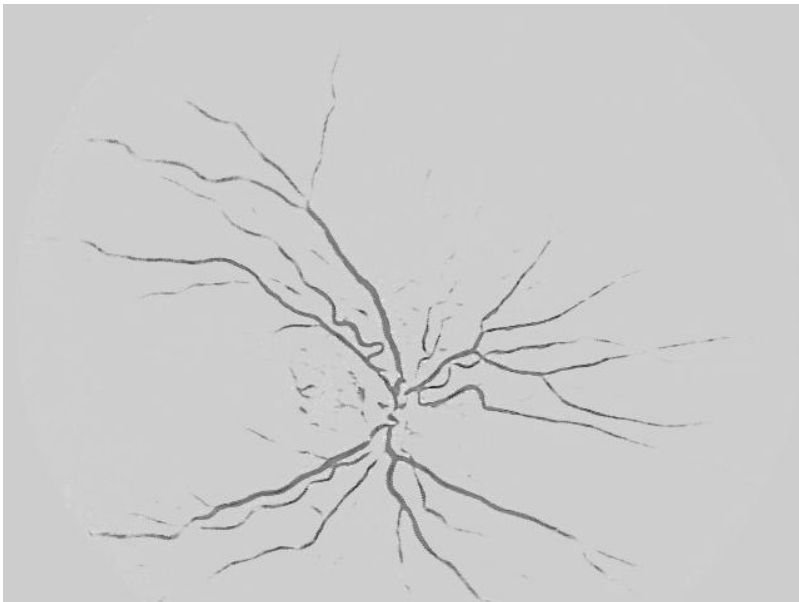in Figure37.

Figure 37 Probability level

In order to evaluate the performance using this kind of image, we based on the maximum average accuracy (MAA).

To determine the MAA, the result images produced by the algorithm were thresholded with variable thresholds. Using these thresholded images the accuracy per image could be determined by taking the total number of pixels of the vessels that were correctly classified and the background pixels for each image and then dividing this sum by the total number of pixels in the mask we calculated before. To finally find out the MAA we take the maximum average value of the accuracy for all images.

6.2 Evaluation indexes

We have already mentioned the accuracy as an index to evaluate the performances of the algorithm, but it is not the only one. In fact, we can use also the sensitivity and the false detection; those parameters, as well as the accuracy, give us quantitative results. Otherwise, we can use the ROC that gives us a qualitative idea of the results.

6.2.1 Sensitivity

Sensitivity is the percentage of vessels pixels in the image that the classifier recognizes as correct.

We can not use this parameter on his own, because it doesn't show anything about the error that we have made. Indeed, if we set all the pixels in the image as parts of the vessel class, we have 100% of sensibility, but we have also wrongly set all the background pixels and this parameter is not able to point out the error. This is the reason why sensitivity is always paired up with false detection.

6.2.2 False detection

False detection is the percentage, in the whole group of pixels that the classifier sets as vessels, that instead are pixels of the background.

This parameter gives us an idea about the size of the classifier error.

6.2.3 Accuracy

Accuracy is an index that doesn't need other evaluations to show how a classifier has worked. It gives us the percentage of the pixels that have been correctly classified, without making distinction between vessels pixels or background pixels.

If we want to use this parameter we have to pay attention to the fact that accuracy has a sort of "lower bound", and if our results in percentage are high, but too near to this bound, they are not so good as they seem.

To illustrate this situation we take our situation in example: in our images the background represents on the average the 95% of the whole pixel of the mask; so if we assign all the pixels to the background without making a classification, we have an accuracy of the 95%. This is our lower bound, any performance lower than this doesn't work.

In our situation the lower bound is very high because the two classes are not proportionate; the background includes almost the totality of the elements.

Usually, in the images taken with fundus camera, this bound is about at 87%[6] and this fact shows again how difficult is the classification in Retcam images.

6.2.4 ROC

As we did with the features, we can calculate the ROC also for the probability matrix that the classifier returned. This method doesn't give us a quantitative evaluation, but just a qualitative vision of the classifier job.

6.3 Our results

The first attempts, using all the original features, gave us a sensibility of more or less 20% and a false detection of 18%. These results were useless and that was the reason why we started to manipulate the features.

The first attempts of features selection improved the results up to 35% of sensibility combined with 20% of false detection, but they wasn't good enough, so we decided to filter the images.

We have soon noticed an improvement of the quality of the results using these new features and, after the application of the wrap selection, we decided to use only seven features because this set gave us the best results.

The features that we used are: the original image processed with bilateral filter, the Ricci Perfetti strength line, the Ricci Perfetti orthogonal strength line, the hessian of the matched, the combination we called "Big filter", the mean calculated with a window of three pixels, and the mean calculated with a window of five pixels.

Thanks to this set of features we obtained the best results, which give us an average sensitivity of 60,35% with an average false detection lower than 25%.

The average accuracy, that is the most important parameter, is equal to 97,31%.

We have divided the results into two tables, and each table represents the results of the fold that we have used for the classification.

Each table shows for each image the total number of white pixel of the ground truth, the confusion matrix, the sensitivity, the false detection and the accuracy. All the parameters are evaluated using the threshold that the MAA method returns.

In Table1 we refer to the first ten images, in Table2 we refer to the second ten images.

In the last Figures we show some examples of images with the respective ground truths and previsions.

IMAGE	TOTAL OF WHITE PIXEL	CONFUSION MATRIX		SENSITIVITY	FALSE DETECTION	ACCURACY
1	16153	253866	7633	52,75	11,07	96,79
		1061	8520			
2	10067	259157	3676	63,48	20,53	98,03
		1651	6391			
3	7833	251602	2637	66,33	44,62	97,41
		4187	5196			
4	13091	254833	5503	57,96	23,89	97,08
		2382	7588			
5	9622	246988	3648	62,09	25,83	97,79
		2081	5974			
6	15750	255279	6863	56,43	13,92	96,95
		1437	8887			
7	12540	256282	4977	60,31	15,03	97,66
		1338	7563			
8	10916	256026	4399	59,70	15,45	97,92
		1191	6517			
9	9823	258965	3651	62,83	33,70	97,50
		3137	6172			
10	15082	254461	6129	59,36	17,99	97,02
		1964	8953			
AVERAGE				60,12	22,20	97,42

Table1 Results on the first ten images

IMAGE	TOTAL OF VESSEL PIXELS	CONFUSION MATRIX		SENSITIVITY	FALSE DETECTION	ACCURACY
11	7188	258126	2086	70,98	53,60	97,06
		5893	5102			
12	13466	253409	6211	53,88	18,36	97,08
		1632	7255			
13	9719	251742	2907	70,09	36,57	97,42
		3928	6812			
14	15132	250231	6957	54,02	12,97	96,93
		1218	8175			
15	9306	256516	3243	65,15	35,43	97,56
		3327	6063			
16	14173	255736	5220	63,17	16,75	97,42
		1802	8953			
17	12600	253662	5391	57,21	28,62	96,92
		2890	7209			
18	13111	254423	6554	50,01	31,69	96,45
		3042	6557			
19	8741	258625	2497	71,43	30,03	98,08
		2680	6244			
20	13915	253669	6967	49,93	12,54	97,04
		996	6948			
AVERAGE				60,59	27,66	97,20

Table2 Results on the other images
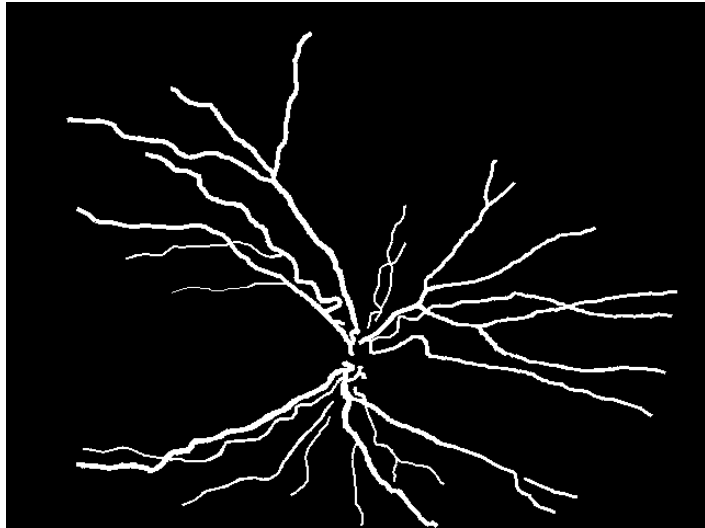
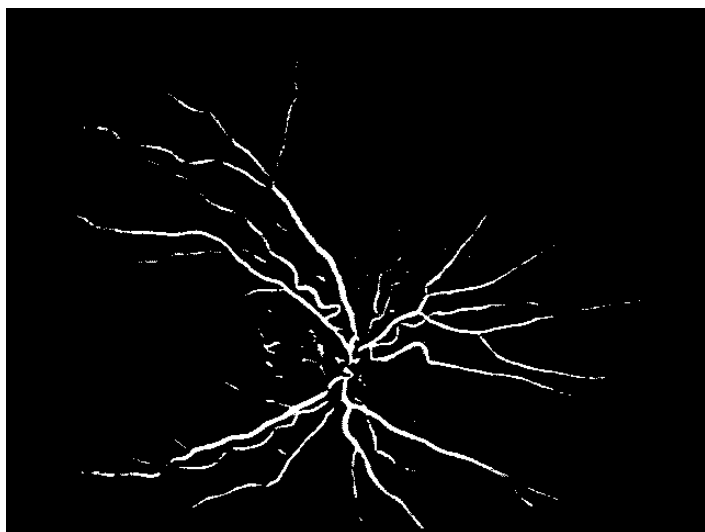**Figure 38 Image 1**

Figure 39 Image 1 ground truth

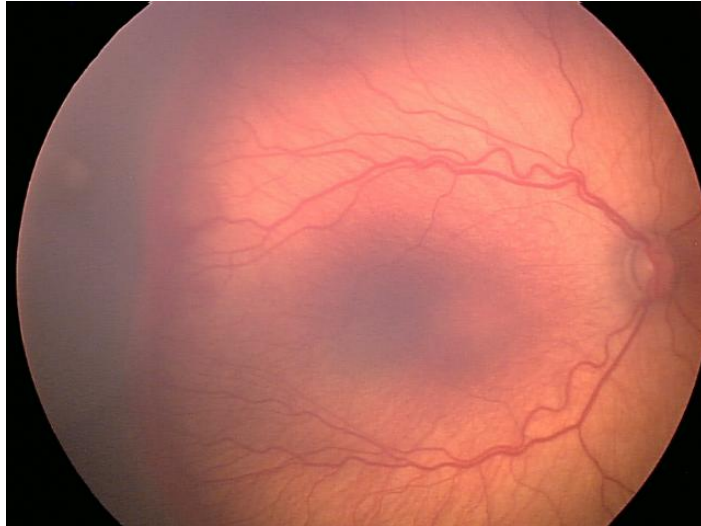**Figure 40 Image 1 Prediction**

Figure 41 Image 4

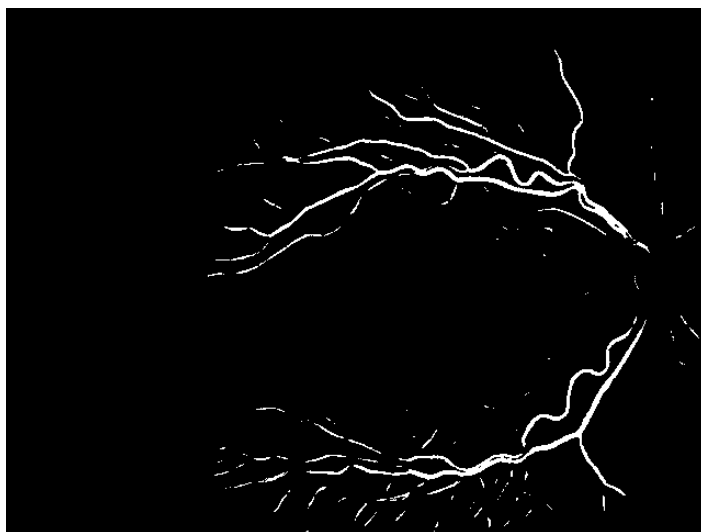

Figure 42 Image 4 ground truth

Figure 43 Image 4 prediction
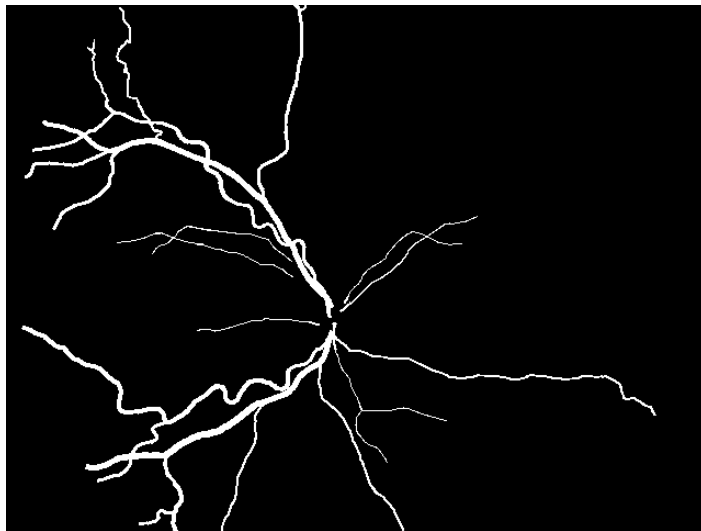
Figura 44 Image 8

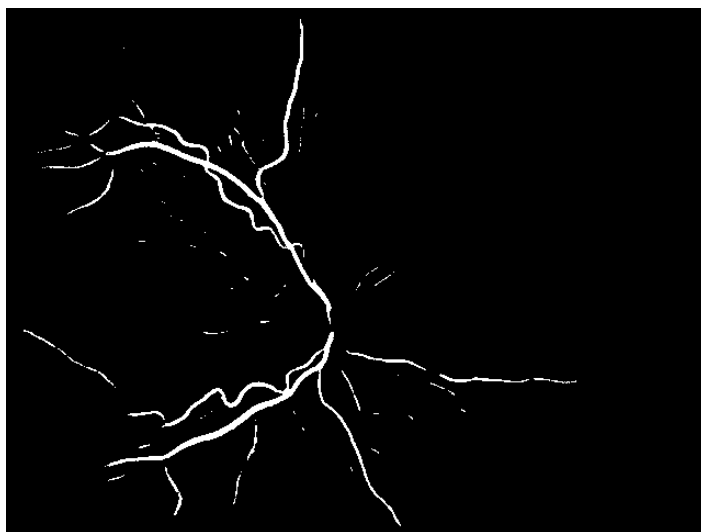**Figure 45 Image 8 ground truth**

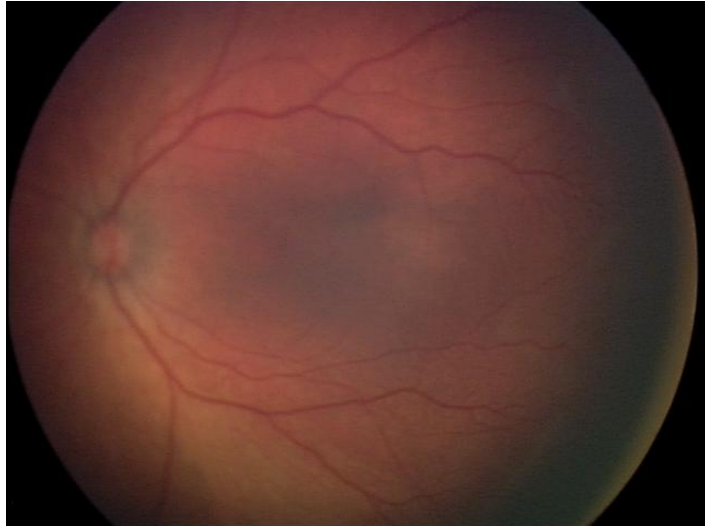**Figura 46 Image 8 prediction**
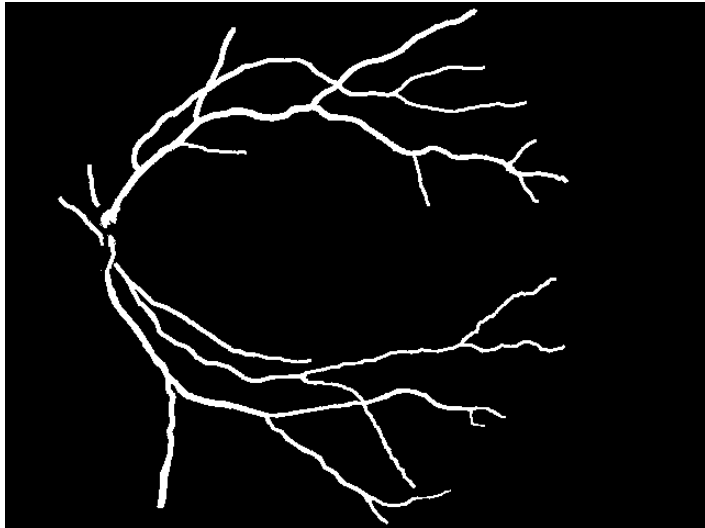
Figure 47 Image16

Figure 48 Image16 ground truth

Figure 49 Image16 prediction

Chapter 7

Discussion and conclusions

Our first purpose was the creation of a program that could be able to track the vessels network in a RetCam image, in order to make the first step in a wider program, which aim is to make the automatic diagnosis of the retinopathy of prematurity.

Unfortunately the results that we have achieved permit only a first qualitative diagnosis.

In fact this kind of prediction enables us to make an evaluation only on the extent of the vessels growth on the retina. However even if this result is partial, it could be very useful because vessel extension is one of the hallmarks of the ROP, and this knowledge can enable the specialists to follow up the evolution of the vessels growth.

For what concerns the other parameters needed by a correct diagnosis, as vessels tortuosity or dilatation, this kind of images are not sufficiently detailed to give secure information. However, we can try to extract these parameters from some of the various features that we used.

For example, we can see how the "big filter", as we called it, exalted the tortuosity of the vessel, and in this way it can be possible that the images, as the one in Figure3, can give us information about tortuosity or other important parameters.
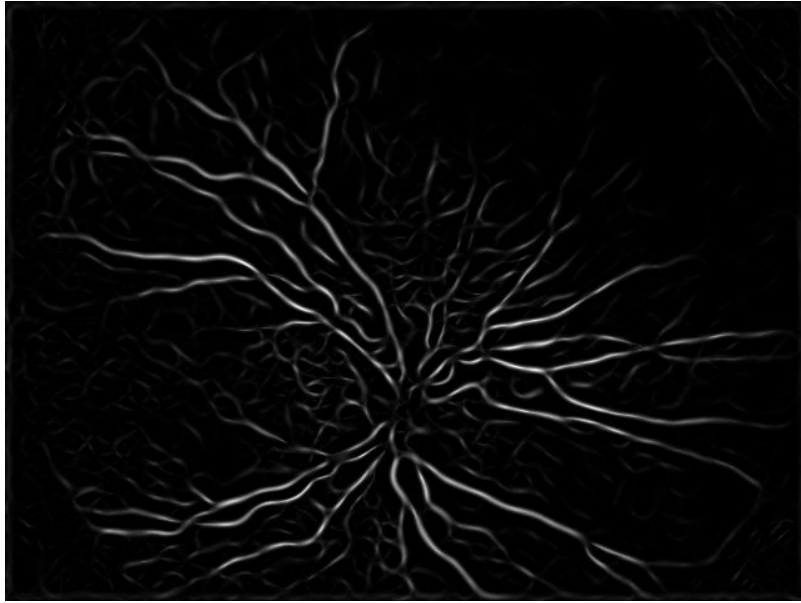
Figure 50 Useful features

7.1 Comparison with other methods

Finally we try to compare the results of our job with other results we have used to take the cue from [6]. As we have already said, our job is a total innovation and for this reason we must be careful in the comparison, because other methods use always images taken from fundus camera or angiography, so the quality of the images is much better than ours.

Other studies start from a lower bound of the accuracy that is about 87% and this fact means that there are more vessels in these images than in ours, where this bound is around 95%.

Some of these methods are surely worse than our job, for example [7] or [8], because their attempts use only geometrical or morphological proprieties and they don't cross 88% of accuracy [7] and 92%[8].

The results of the accuracy of the other studies are placed between 93% and 95% of accuracy, and we suppose that also our program can achieve these kind of results, if it is applied on these kind of images.

7.2 Further aims

We think that we have achieved the best possible results from this kind of images. In further attempts probably the use of other features, derived from other filters, can improve the performances, but not in a significant way.

We think that probably the only element that could improve significantly the performance is the use of another kind of classifier, but this is not a foregone conclusion.

References

[1] *The international classification of retinopathy of prematurity revisited*

[2] http://www.nei.nih.gov/health/rop

[3] C. Wu, R.A. Petersen, D.K.Vanderveen*, RetCam Imaging for Retinopathy of Prematurity Screening*

[4] M. Niemeijer, J. Staal, B.van Ginneken, M. Loog, M.D. Abramoff, *Comparative Study of Retinical Vessel Segmentation Methods on a New Publicly Available Database*

[5] F. Zana, J.C. Klein, *Segmentation of Vessel-Like Patterns Using Mathematical Morphology and Curvature Evaluation*

[6] E. Ricci, R. Perfetti, *Retinal Blood Vessel Segmentation Using Line Operators and Support Vector Classification*

[7] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, M. Goldbaum, *Detection of Blood Vessels in Retinal Images Using Two-Dimensional Matched Filters*

[8] H. Tamura, S. Mori, Y. Yamawaki, *Textural Features Corresponding to Visual Perception*

[9] J. Sklansky, *Image Segmentation and Feature Extraction*

[10] R.M. Haralick, *Statistical and Structural Approaches to Texture*

[11] W. Richards, A. Polit, *Texture matching*

[12] S. W. Zucker, K. Kant, *Multiple-level Representations for Texture Discrimination*

[13] http://www.fp.ucalgary.ca/mhallbey/what_is_texture.htm

[14] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

[15] I. Guyon, A. Elisseeff, *An Introduction to Variable and Features Selection*