



Proceedings of the 16th Danish Conference on Pattern Recognition and Image Analysis

Eds. Søren I. Olsen

Technical Report no. 08-10

ISSN: 0107-8283

Dept. of Computer Science

University of Copenhagen • Universitetsparken 1

DK-2100 Copenhagen • Denmark

Proceedings

of

the 16th Danish Conference on Pattern Recognition and Image Analysis

organised by

Danish Society for Automatic Pattern Recognition
(Dansk Selskab for Automatisk Genkendelse af Mønstre)
Department of Computer Science, University of Copenhagen

Thursday 21 August 2008

Editor Søren I. Olsen

Table of Contents

L. Sørensen, M. de Bruijne	4
<i>Pattern recognition based emphysema for pattern recognition</i>	
P. Johansen	6
<i>A random monotone and differentiable function – or how to parameterize a curve</i>	
P. Lo, J. Sporning, H. Ashraf, J. J. H. Pedersen, M. de Bruijne	8
<i>Voxel classification based vessel-guided airway segmentation</i>	
L. L. Larsen, M. Ganz, M. Nielsen	10
<i>Growth analysis of atherosclerotic plaques</i>	
N. P. Tiilikainen, A. Bartoli, S. I. Olsen	12
<i>Registering and Retexturing Cartoon-Like Videos using Contours</i>	
V. K. Perfanov	14
<i>Investigation of point triangulation methods for optimality and performance in Structure from Motion systems</i>	
V. Andersen, H. Aanæs, A. Bærentzen, M. Nielsen	16
<i>Markov random fields on 3D polygonal meshes</i>	
D. Gustavsson, K. S. Pedersen, M. Nielsen	18
<i>Multi-scale natural images: A database and some statistics</i>	
C. B. Madsen	20
<i>The CVMT/AAU sphere gantry</i>	
C. B. Madsen	23
<i>Estimating Radiances of the Sun and the Sky from a Single Image Containing Shadows</i>	
K. Nasrollahi, T.B. Moeslund	25
<i>Face Image Quality and its Improvement in a Face Detection System</i>	

Pattern recognition based emphysema quantification

Lauge Sørensen^a and Marleen de Bruijne^{a,b}

^a Department of Computer Science, University of Copenhagen, Denmark;

^b Biomedical Imaging Group Rotterdam, Erasmus MC, The Netherlands

1. INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a major cause of death and a growing health problem worldwide. It is characterized by limitation of airflow in the airway and it comprises two components: Chronic bronchitis, which is an inflammation of the small airways, and emphysema, which is characterized by gradual loss of lung tissue. We focus on the assessment of emphysema in computed tomography (CT) images of the lung. Emphysema is thought to be the main cause of shortness of breath and disability in COPD, and the two emphysema classes considered here are: Centrilobular emphysema (CLE) defined as multiple, small, spotty lucencies, that may have thin walls and paraseptal emphysema (PSE) defined as multiple, lucencies in a single layer along the border of the lung, commonly with thin walls visible. Normal lung tissue (NT) is added as a third class, thus we have the classes $\omega_i \in \{NT, CLE, PSE\}$.

One way to objectively analyze the properties of the disease patterns of the various subtypes is to use texture analysis.¹ Previously we explored the use of local binary patterns (LBP), originally proposed by Ojala *et al.*,² as features for discriminating regions of interest (ROI) based on texture in CT slices of lungs, which showed promising results.³ In this abstract we propose a quantitative measure for emphysema, based on a pattern classification approach that utilizes local texture information.

2. MATERIALS AND METHODS

The quantitative measure is evaluated on a set of 117 CT slices from a population of 39 patients. All patients also underwent lung function tests (LFT), the standard non-CT measure for emphysema diagnosis, prior to the CT scanning of the lungs. We use the LFT forced vital capacity percent predicted (FEV₁%pred), which is the amount of air in liters that you can forcibly blow out in one second, adjusted for age, sex, and height, for evaluating our measure. CT is the most sensitive method of determining emphysema subtype and assessing severity,⁴ still CT based quantitative emphysema measures should correlate with LFT.

An ROI is represented by the joint intensity and LBP histogram. The rotation invariant LBP formulation is used with a radius of 1 pixel and 8 samples on the circle.^{2,3} A k NN classifier⁵ is build using a set of 216 31×31 pixel ROIs annotated from some of the CT slices, by computing feature histograms and using these as prototypes in the k NN. These feature histograms represent the three classes: NT (107 observations, of which 48 were near the lung border or hilum area), CLE (50 observations), and PSE (59 observations).

Full lung classification is performed by computing the posterior class probability for each pixel position $\mathbf{x} = [x, y]^T$ in the lung, based on the local 31×31 pixel neighborhood around the pixel, using the trained classifier. A quantitative measure for emphysema is derived from the posterior by computing the average posterior probability of each class ω_i in the lung. Thus, we obtain the mean class posterior (MCP)

$$\text{MCP}_{\omega_i} = \frac{1}{N} \sum_j^N P(\omega_i | \mathbf{x}_j), \quad (1)$$

where N is the number of pixels in the lung and the posterior probability is given by

$$P(\omega_i | \mathbf{x}) = \frac{\min_m L_{hist}(\mathbf{x}, \omega_i, m)}{\sum_{j=1}^3 \min_m L_{hist}(\mathbf{x}, \omega_j, m)},$$

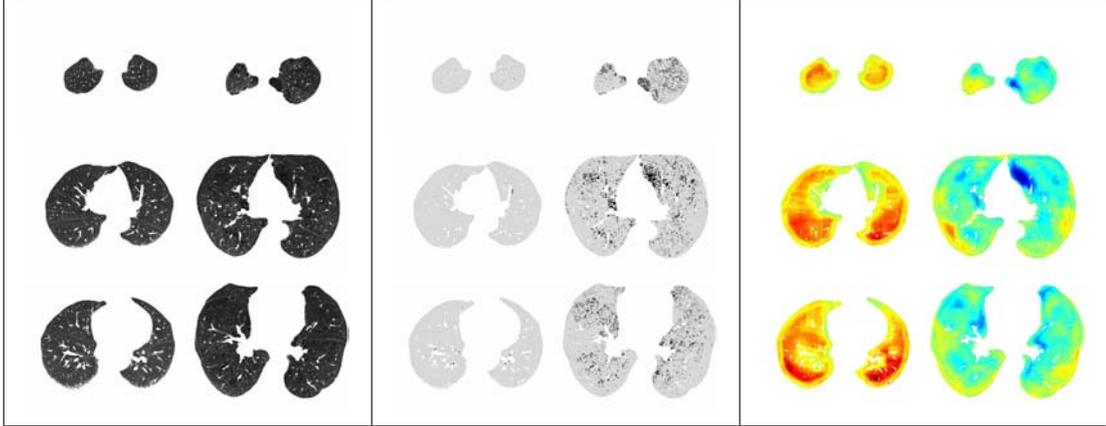


Figure 1. Example images from two patients. Each column of each part shows three CT slices from the same patient, a healthy patient to the left and a COPD patient to the right (according to LFT). **Left:** The segmented images used in the classification. **Middle:** A threshold of -910 Hounsfield units applied to the segmented images. The areas below the threshold are indicated in black and the lung segmentations are indicated in light gray. **Right:** Mean NT posterior. Dark red means that a NT prototype histogram is very similar to the given pixel's histogram in feature space. Dark blue means that all NT prototype histograms are dissimilar. Refer to the electronic version for colors.

where $L_{hist}(\mathbf{x}, \omega_i, m)$ is the histogram intersection between the feature histogram of the current pixel and the feature histogram $H^{\omega_i, m}$ of prototype m from class ω_i

$$L_{hist}(\mathbf{x}, \omega_i, m) = 1 - \sum_{b=1}^B \min(f(\mathbf{x})_b, H_b^{\omega_i, m}),$$

where $f(\mathbf{x})$ is some function that extracts the local neighborhood around pixel \mathbf{x} and computes the feature histogram, B is the number of feature histogram bins, and b denotes the bin index.

3. RESULTS AND CONCLUSION

The most standard approach to computerized emphysema quantification in CT is applying a threshold and measuring the amount of pixels below the threshold relative to the total amount of pixels, sometimes referred to as density mask⁴ or emphysema index. In the middle part of Figure 1 we show the thresholded versions of the segmented CT slices shown in the left part. For comparison, MCP_{NT} obtained by applying (1) with $\omega_i = NT$ to all pixels in the lung segmentation are shown in Figure 1 right.

The correlation between $FEV_1\%pred$ and MCP_{NT} , MCP_{CLE} , and MCP_{PSE} is 0.73, -0.58, and -0.73 respectively, and all correlation coefficients are significant ($p < 0.001$). Thus, our proposed quantitative measure agrees well with the de facto non-CT emphysema measure.

REFERENCES

1. Tuceryan, M., Jain, A.K.: Texture analysis. In: The Handbook of Pattern Recognition and Computer Vision (2nd Edition). World Scientific Publishing (1998) 207–248
2. Ojala et al.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell **24**(7) (2002) 971–987
3. Sørensen, L., Shaker, S.B., de Bruijne, M.: Texture classification in lung ct using local binary patterns. MICCAI (2008) (Accepted).
4. Müller, N.L., Staples, C.A., Miller, R.R., Abboud, R.T.: "density mask". an objective method to quantitate emphysema using computed tomography. Chest **94**(4) (Oct 1988) 782–787
5. Jain, A., Duin, R., Mao, J.: Statistical pattern recognition: a review. IEEE Trans Pattern Anal Mach Intell **22**(1) (Jan. 2000) 4–37

A random monotone and differentiable function - or how to parameterize a curve.
Working document

Peter Johansen, DIKU, e-mail: peterjo@diku.dk

Abstract

A random monotone function can be defined as follows. Consider a function $f(x)$ with first derivative $1 + \epsilon_1$ and higher order derivatives ϵ_i , where ϵ_i are stochastic i.i.d variables. We can construct an n times iterated function $f^{o_n} = f(f(\dots))$. The variance of the derivatives will have increased. In order to avoid the increase we instead iterate n times the function f_n with stochastic variables ϵ/n . We will define a random monotone function as the limit of f_n as n tends to infinity. One can see without too much difficulty that the first derivative will follow a lognormal distribution: the function is monotone. This note presents the distribution of a random monotone and differentiable function and discusses ways to obtain the distribution of the derivatives of the random function.

Introduction

In our field - computer vision and pattern recognition - statistical methods play an increasing role. "We see what we are looking for". In statistics we model what we are looking for by a prior distribution of objects or patterns, and we weigh the evidence given by the data by our expectations given by the prior. It is important to choose a proper prior, and this note presents a proposal for an uncommitted prior for a scalar monotone and differentiable function of a real variable. This is not meant to imply that an uncommitted prior should always be chosen, but it is important to be able to choose a prior that does not bias our analysis. Assumptions in a specific situation should be modelled by an additional specific prior. The uncommitted prior is the "rock bottom" when we know nothing, or when we do not want to assume anything specific that may bias our analysis.

If one tries to set up a model for a differentiable and monotone function in which the derivatives were random numbers one would violate monotonicity. However, if one could use a linear function and impose infinitesimally small derivatives of any order, one would get a monotone function. We shall approximate this situation to any order of the function's Taylor series.

A random monotone function can be defined as follows.

Definition Consider a function $f(x)$ with first derivative $1 + \epsilon_1$ and higher order derivatives ϵ_i , where ϵ_i are stochastic i.i.d variables. We can construct an n times iterated function $f^{o_n} = f(f(\dots))$. The variance of the derivatives will have increased. In order to avoid the increase we instead iterate n times the function f_n with stochastic variables ϵ_i/n . We define a random monotone function as the limit of f_n as n tends to infinity.

End definition

First we consider the situation in which the derivatives are real numbers and not distributions, or in other words, the distributions are delta functions. In that case the analysis can be performed exactly.

First recurrence equations are set up to find the Taylor coefficients of an n times iterated function. Subsequently letting $n \rightarrow \infty$ one obtains the infinitely often iterated function if such one exists. For simplicity initially only the two first coefficients are considered.

Recurrence equations

We set up a recurrence equation for the derivatives of the iterated function f_∞ .

$$f = f_x x + f_{xx} x^2$$

$$\begin{aligned}
f_n &= a_n x + b_n x^2 \\
f_n(f(x)) &= a_n(f_x x + f_{xx} x^2) + b_n(f_x x + f_{xx} x^2)^2 \\
&= a_n f_x x + (a_n f_{xx} + b_n f_x^2) x^2 \quad (\text{modulo } x^3)
\end{aligned}$$

$$\begin{aligned}
a_{n+1} &= f_x a_n \\
b_{n+1} &= f_{xx} a_n + f_x^2 b_n
\end{aligned}$$

We first solve the equations in the eigensystem of the matrix M for real values f_x and f_{xx} . We want to find the limit for n tending to infinity of M^n . In the eigensystem, the eigenvectors are multiplied by the eigenvalues to the power n . We are to represent the polynomial as a linear combination of the eigenvectors. The effect of the matrix is to multiply the coefficient of an eigenvector by its eigenvalue. This way the limit is found.

Stochastic variables

If we let the variables be stochastic variables, the argument does not carry over. The reason is that at each iteration the values change according to the distribution of the variables and accordingly the eigenvectors change. In the present case one eigenvector is unchanged, but it is not obvious how to take advantage of this property. When the numbers are stochastic variables the solution for the first derivative a_∞ is the lognormal distribution. This is seen by taking the logarithm before going to the limit. The factors of the first derivative a_n are i.i.d. and the infinite sum of their logarithms is the normal distribution by the central limit theorem. Notice that we only need to assume bounded variance but no specific probability distribution function of the stochastic variables. The question is how one generalizes this argument from scalars to matrices, which is what we need in this case.

The plan is to use the method from Jackson et al [1]. They found the distribution of an infinite random matrix product by first deriving a partial differential equation for the distribution, and solving that equation to find an infinite sum for the distribution. Nielsen et al [2] used their result to compute an uncommitted warp from an image onto another that transfers a set of landmarks to their designated position.

References

- [1] Andrew D. Jackson, Benny Lautrup, Peter Johansen, and Mads Nielsen. Products of random matrices. *Phys. Rev. E*, 66(6):5, December 2002. Article 66124.
- [2] Mads Nielsen, Peter Johansen, Andrew D Jackson, Benny Lautrup, and Søren Hauberg. Brownian warps for non-rigid registration. *Journal of Mathematical Imaging and Vision*, 31:221–231, 2008.

Voxel classification based vessel-guided airway segmentation

Pechin Lo^a, Jon Sporning^a, Haseem Ashraf^b, Jesper J.H. Pedersen^c and Marleen de Bruijne^{a,d}

^aImage Group, Department of Computer Science, University of Copenhagen, Denmark;

^bDepartment of Respiratory Medicine, Gentofte University Hospital, Denmark;

^cDepartment of Thoracic Surgery, Gentofte University Hospital, Denmark;

^dBiomedical Imaging Group, Departments of Radiology and Medical Informatics, Erasmus MC
– University Medical Center Rotterdam, The Netherlands

1. INTRODUCTION

It has been shown in various studies that analysis of airways in CT is very important for the analysis of various lung diseases. Airway tree segmentation plays a critical role in these studies, offering a starting point for conducting measurements on the airways. Nevertheless, current available airway segmentation methods are still far from perfect, limiting the amount of obtainable measurements.

Previously, we have proposed a method for airway segmentation based on voxel classification and region growing.¹ In this paper, we propose to incorporate airway and vessel orientation information to further improve the voxel classification based method. This is done by using an orientation similarity measure that is computed from the orientation of the local structure of a candidate airway voxel and the orientation of a neighboring vessel. The orientation similarity measure is then used as an additional criterion in the region growing.

The motivation for our work lies in the fact that there is an artery following every airway branches. This characteristic has been used previously for arteries and veins separation from segmented vessels.² To our knowledge, no work has been done on explicitly using information on the vessels for airway tree segmentation.

2. METHODOLOGY

Airway Appearance Model We based this appearance model on a work done previously,¹ which uses a KNN classifier. Given a set of features \mathbf{x} computed at a particular position in the image, the airway probability of \mathbf{x} from the KNN classifier is defined as $P(A|\mathbf{x}) = K_A(\mathbf{x})/K$ where A is the airway class, $K_A(\mathbf{x})$ is the number of nearest neighbors around \mathbf{x} belonging to the airway class, obtained from a total of K nearest neighbors.

Vessel Orientation Extraction of vessel orientation involves three steps: vessel tree segmentation, centerline extraction and orientation computation. The vessel tree is segmented from the CT image by first using an intensity based threshold followed by a threshold process that uses eigenvalues of the Hessian matrix. This second threshold process is performed on the following two measurements

$$m_1 = \frac{|\lambda_1| - |\lambda_2|}{|\lambda_1| + |\lambda_2|} \quad \text{and} \quad m_2 = \frac{|\lambda_1| - |\lambda_3|}{|\lambda_1| + |\lambda_3|}$$

where $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3|$ are the three eigenvalues of the Hessian matrix. By retaining voxels with $m_1 \leq t_{m_1}$ and $m_2 \geq t_{m_2}$, bright non-vessel structures can be removed. The centerlines are extracted from the segmented vessel tree using a 3D thinning algorithm.³ Subsequently, the vessel orientation at the centerline voxels is measured as the eigenvector corresponding to λ_3 .

Orientation Similarity Measure Extraction of orientation from airways is performed on the airway probability image, where the airway resembles a solid bright tube. We define the orientation similarity measure $s = |\cos(\theta)|$, where θ is the angle between the orientation of an airway candidate voxel and the orientation of the centerline of the vessel nearest to it.

Segmentation Framework The airway tree is extracted using a region growing algorithm, with a decision function defined as

$$D(P(A|\mathbf{x}), s) = \begin{cases} 1, & P(A|\mathbf{x}) \geq T_u, \\ 1, & T_u > P(A|\mathbf{x}) \geq T_l \text{ and } s \geq T_s, \\ 0, & \text{otherwise.} \end{cases}$$

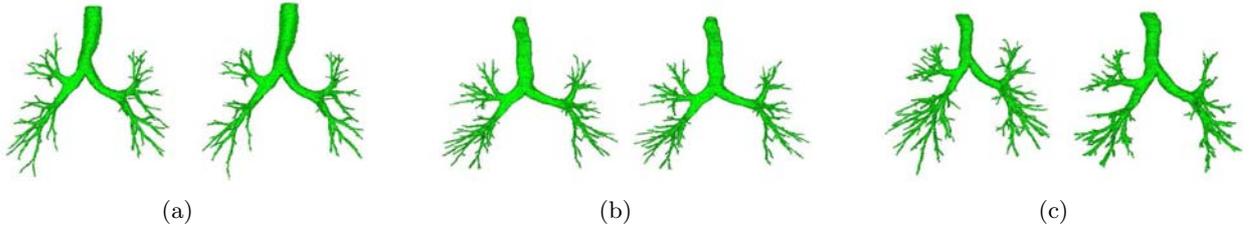


Figure 1. Surface renderings of the segmentation results using airway probability with (left) and without (right) orientation similarity measure.)

where $P(A|\mathbf{x})$ is the airway probability and s is the orientation similarity measure of a particular voxel. A voxel is labelled as an airway if $D(P(A|\mathbf{x}), s) = 1$. Suitable values for the thresholds T_u , T_l and T_s can be found for instance using cross validation.

3. RESULTS

Experiments were conducted on 20 low-dose CT images from different subjects enrolled in the Danish Lung Cancer Screening Trial (DLCST). A two-fold cross validation experiment was conducted.

Fine tuning of T_u , T_l and T_s was done using total branch length and explosion detection based on a modified fast marching algorithm.⁴ The threshold combination selected was the one that had the highest total branch length without any explosion detected for all training cases. For comparison, an airway probability based threshold and an image intensity based threshold were also obtained using the same fine tuning approach.

Results from our experiments showed that the proposed method gives better result as compared to region growing using only airway appearance model, in terms of more and longer branches detected ($p = 0.013$). Examples of our results are given in Fig. 1. On the other hand, region growing based on image intensity performed significantly worse than the appearance model based methods. Not only were the branches detected were less and shorter in general, a serious explosion also occurred in one of the test cases.

4. CONCLUSIONS

An airway tree segmentation method that uses a classification based airway appearance model combined with vessel orientation information is presented. The use of the airway probability image makes it possible to determine the orientation of an airway candidate voxel using Hessian matrix eigen analysis. The airway orientation of the candidate voxel is then compared with the orientation from a vessel nearest to it to form an orientation similarity measure. This orientation similarity measure is used to lower the threshold for airway probability during the region growing process, resulting in a more complete segmentation with longer airway branches. Results from our experiments showed that augmenting airway appearance model with our orientation similarity measure gives better segmentation than with only the airway appearance model.

REFERENCES

1. Lo, P., de Bruijne, M.: Voxel classification based airway tree segmentation. Volume 6914., SPIE (2008)
2. Bülow, T., Wiemker, R., Blaffert, T., Lorenz, C., Renisch, S.: Automatic extraction of the pulmonary artery tree from multi-slice CT data. Volume 5746., SPIE (2005) 730–740
3. Wang, T., Basu, A.: A note on ‘A fully parallel 3D thinning algorithm and its applications’. Pattern Recognition Letters **28**(4) (March 2007) 501–506
4. Schlathöler, T., Lorenz, C., Carlsen, I.C., Renisch, S., Deschamps, T.: Simultaneous segmentation and tree reconstruction of the airways for virtual bronchoscopy. Volume 4684., SPIE (2002) 103–113

Growth Analysis of Atherosclerotic Plaques

Lene Lillemark Larsen¹, Melanie Ganz^{1,2}, and Mads Nielsen^{1,2}

Department for Computer Science, Copenhagen University, Denmark
lene@diku.dk
Nordic Bioscience, Herlev, Denmark

Abstract. We present a novel method to analyze the growth of abdominal atherosclerotic plaques based on x-ray projections. The growth analysis can aid progression monitoring in clinical trials and in population screening programs. Our results are based on a longitudinal study over 8.5 years. The annotations of the calcifications are matched longitudinally using thin plate spline registration and area overlap calculations. The growth of the calcifications is measured by the distribution of the geometry statistics of the calcifications. The method was evaluated on 135 subjects with a total number of 611 calcifications. Our results show, for instance longitudinal growth of calcifications with a mean of 2.53 mm (± 1.95) in the blood flow direction and correlations with pathologically related biomarkers.

1 Introduction

Atherosclerosis is a primary cardiovascular disease (CVD), that is the main cause of the disease burden (illness and death) in Europe. It is a chronic inflammatory process that builds up plaque in the intimal wall of the arteries. Atherosclerosis starts in the childhood and progresses during adolescence as development of fatty streaks, that evolves to fibrous caps and lastly formation of calcifications, which could result in strokes and heart attacks.

X-ray imaging is an attractive image modality for calcifications quantification that can aid in large scale clinical trials and screening programs due to the low cost, fast examination and non-invasiveness. Other image modalities such as intravascular ultrasound, computed tomography and magnetic resonance imaging have also been used for studies of the atherosclerotic growth, but to a smaller extent due to the cost and the patient discomfort. Available x-ray data already exist from clinical trials and routine screening for osteoporosis, which can be used for analyzing the growth of the plaques.

Our hypothesis is that the growth patterns are good estimators of the progression of atherosclerosis. To our knowledge this is the first study that characterizes the growth patterns of atherosclerotic calcifications from x-ray projections. Our region of interest is the lumbar region denoted by L1-L4. The amount of calcified deposit in L1-L4 can be an indicator of the risk of future cardiovascular events.

2 The Measuring of the Growth

The annotations of the calcifications are registered by thin plate spline registration and then a suitable match is found using area overlap which can be seen in Fig. 1. Now we

extract information of the individual calcifications about the growth patterns. The growth of the calcifications can be measured in three different biologically meaningful directions; longitudinal (in the blood flow direction), circumferential (around the aorta) and radial (the direction into the aorta). The longitudinal growth is measured in the change in the height of the calcifications. The radial growth is measured as the change in the width of the calcifications. The area of the calcifications could give information about the overall growth of the calcifications. The overall growth direction we have measured as the difference in the center of mass of the matched calcifications.

3 Results and Discussion

We have been able to match 32.4% of the baseline calcifications to one corresponding follow-up calcification. The registration and matching process is necessary to describe the growth patterns for the calcifications and especially the growth direction, measured by change in the center of mass.

The results show that the matched calcifications are growing longitudinally with a mean growth of 2.53 mm (± 4.33). This is an increase of 50.4 %. The mean growth in the radial direction is 0.29 mm (± 1.70), corresponding to 13.7 %. The center of mass is moved 2.02 (± 1.95) mm in the longitudinal direction. The calcifications grows downward the aorta, which could be caused by the turbulence in the direction of the blood flow, when the aorta wall becomes non-smooth, due to the existing calcifications. We have correlated our growth patterns with known biological risk factors. Our results show that high cholesterol and triglyceride levels correlates with growth in width of the calcifications. This could correspond to intimal calcifications that likely are related to elevated cholesterol and lipids that induce atherosclerosis. A high glucose level correlate mainly to growth in heights possibly corresponding to the elongated medial calcifications observed in diabetes patients. Glucose is a main factor of growth of all types of calcifications, which can also be seen in the correlation.

At this point we can describe clinically meaningful growth patterns of the atherosclerotic calcifications. More work will be based on modifying the growth patterns and get correlations with the different characteristic biomarkers. We will also try to further distinguish the growth pattern of medial and intimal calcifications in our future growth patterns.

We have shown that registration and plaque matching based on x-ray images can give a good description of the growth patterns, which indicates the ability for a simple and low cost method to measure the longitudinal progression of atherosclerosis.

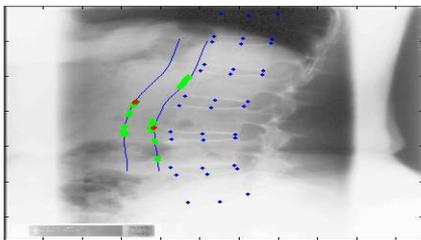


Fig. 1. An example of the matched calcifications. Blue shows the annotations of the vertebrae and the aorta wall. Red are the baseline calcifications and green are the follow-up calcifications. Note that the area overlap is a good indicator of the matching calcifications.

Registering and Retexturing Cartoon-Like Videos using Contours

N. P. Tiilikainen^{1,2}

A. Bartoli^{2,1}

S. I. Olsen¹

¹DIKU, Copenhagen, Denmark

²LASMEA, Clermont-Ferrand, France

June 23, 2008

Extended Abstract

Realistic retexturing of non-rigidly deforming surfaces is an important problem in computer vision as it has a wide variety of real world applications, especially in the movie and entertainment industry. A key step in producing visually pleasing retexturing results is registration. For non-rigid surfaces in general the registration problem is difficult, since the shape of the scene changes between the frames. Furthermore, if the observed surface is predominantly void of texture, or if the texture is poorly distributed, it is not possible to capture sufficient feature point correspondences for correctly modeling the deformations.

The main goal of this work is therefore to perform registration and retexturing of challenging videos of non-rigid surfaces where traditional methods would fail. Such cases are often found in cartoon videos, where there is a high number of smooth contours and only little or spurious texture.

We propose a novel method for registering and retexturing cartoon-like videos by means of joint contour detection and point to point curve matching. The main idea is to fit a parametric 3D active surface model in the spatiotemporal space. A regularization term is used to limit the change in curvature over time. We also show that with cross-validation it is possible to automatically estimate a suitable value for the regularization parameter, controlling the tradeoff between the regularization and the data term.

The two main contributions of this work are:

- First, we present a method for registering videos relying solely on the presence of strong contours. The contour model is based on a 3D parametric active surface model. The standard active contour model was introduced by Kass *et al.* and later extended to 3D. In our case the 3D active surface is a collection of closed curves, each lying in a separate image in the video, creating a tubular shaped surface spanning the spatiotemporal space. We introduce a novel regularization term to the energy functional of the 3D active surface that enables joint contour detection and point to point matching of the curves in the 3D active surface. We match the curves based on their spatial curvature as this is the only visually meaningful cue that is available. We assume that the region of interest is not subject to occlusions and deforms smoothly.
- Second, we address one of the inherent problems with the active contour model: automatically choosing the value of the regularization parameter. The regularization parameter controls the tradeoff between the data and the regularization term. Often the value of this parameter is chosen based on empirical observations. Instead we propose to automatically compute a suitable value for the regularization parameter by maximizing the predictivity of the 3D active surface using k -fold cross-validation.

In order to perform the final retexturing we retrieve a set of $\mathbb{R}^2 \mapsto \mathbb{R}^2$ Thin-Plate Spline (TPS) warps between a selected reference frame and the remaining frames. The data points for computing the TPS warps are extracted from the curves in our 3D active surface by uniformly sampling each curve at the same interval.

We report convincing registration and retexturing results on cartoon videos.

Investigation of point triangulation methods for optimality and performance in Structure from Motion systems

Vesselin Perfanov
vp@imm.dtu.dk
DTU Informatics

Structure from Motion (SFM) systems are composed of cameras and structure in the form of 3D points and other features. It is most often that the structure components outnumber the cameras by a great margin. It is not uncommon to have a configuration with 3 cameras observing more than 500 3D points. Therefore fast 3D point triangulation algorithms are required in the different stages of a SFM system: at the initialization of the reconstruction during the primary estimation of the camera geometry, when new points from new cameras are being added to the system, etc. This presentation will give an overview of existing triangulation methods with emphasis on performance versus optimality, and will suggest a fast triangulation algorithm based on linear constraints.

The structure and camera motion estimation in a SFM system is based on the minimization of some norm of the reprojection error between the 3D points and their images in the cameras. Most classical methods are based on minimizing the sum of squared errors, the L_2 norm, after initializing the structure by an algebraic method ([2]). It has been shown (in [4] amongst others) that first, the algebraic method can produce initial estimates that are erroneous due to its minimization of a meaningless error measure and second, the L_2 norm based function is not convex, and often gets stuck in local minima.

L_∞ norm minimization. The optimization frameworks presented in [4] and [5] build upon the theory of convex optimization. It has been shown that the individual components of the reprojection error are quasiconvex functions $f_i(x)$ that retain single minima in the region in front of the cameras in the Euclidean case. Therefore the triangulation problem can be reformulated as a constraint *minimax* problem over the L_∞ norm of the error and solved as a feasibility problem with quadratic constraints via Second Order Cone Programming (SOCP). This is effectively the same as projecting quadratic cones through the cameras centers and the measured image points into space and asking the question if there exist an intersection of all cones in 3D space for a given cone radius γ .

However, even though SOCP problems are solved in polynomial time, this method requires the solution to multiple SOCP problems via bisection for γ that increases the computational time. Furthermore the handling of projective space increases the computational load because multiple minima have to be found and handled separately.

Minimizing the largest error vector component (LP method). One possibility for improving the computational efficiency while sacrificing some accuracy is to minimize the maximum of the x and y components of the error vector. This is equivalent to fitting a regular tetrahedron through the cameras and the image points and minimizing its side at the image plane while an intersection between the tetrahedrons from all cameras exist. The minimization can be set up as a direct LP *minimax* problem of a linear function

subject to linear constraints in most convex optimization packages. Alternatively it can be solved as a feasibility problem by a fast heuristic method involving $16 \times m \times (m - 1)$ 3D point re-projections and plane intersections where m is the number of cameras.

Direct solution to the L_∞ norm problem. Instead of solving multiple feasibility problems via SOCP it is possible to reformulate the convex envelope of all quasiconvex functions $f_i(x)$ into a continuous piece wise quadratic function that can be minimized directly, together with its constraints, by an interior point method [1]. If we assume that the function has the same linear constraints as in the previously mentioned strategy, the derivatives of the barrier function will be simpler, and furthermore we can use the LP method to find a starting point and assure that a feasible set of points exist in the search region. This method converges in few Newton method steps and is fast due to its low dimensionality and strong convexity. Furthermore this method minimizes a more meaningful error measure than the LP method and has the possibility for incorporating covariance matrices of the uncertainty of the image points' locations.

Once thoroughly tested, implementations of these triangulation algorithms will find their way into the SFM system which implementation is a part of my PHD. They will be used particularly in the initialization of the cameras and structure and thereafter in the camera resectioning. Classical global optimization as bundle adjustment [5] will still be used to optimize the L_2 norm. By employing robust convex optimization techniques for initialization we hope to avoid local minima. With the continuous development of the convex optimization frameworks as presented in [4] and [5] the classical bundle adjustment may be completely reformulated as a large convex problem with a single minimum. However the demand for real-time, online, dense SFM systems will put heavy pressure on the computation time of those algorithms.

References:

- [1] Boyd S., Vandenberghe L.: **Convex Optimization**. Cambridge University Press (2004) , ISBN: ISBN-10: 0521833787
- [2] Hartley, R.I., Zisserman, A.: **Multiple View Geometry in Computer Vision**. Cambridge University Press (2004) Second Edition , ISBN: 0521 54051 8
- [3] Stewénius, H., Schaffalitzky, F., Nistér, D.: **How hard is three-view triangulation really?** In: Int. Conf. Computer Vision, Beijing, China (2005) 686–693
- [4] Kahl F., Hartley R., **Multiple View Geometry Under the L-infinity Norm**, In IEEE Transactions on Pattern Recognition and Machine Intelligence 2008
- [5] Kahl F., Agarwal S., Chandraker M., Kriegman D. and Belongie S., **Practical Global Optimization for Multiview Geometry**, In: International Journal of Computer Vision, 2007
- [5] Triggs B., McLauchlan P. and Hartley H., Fitzgibbon A., **Bundle Adjustment — A Modern Synthesis**, Vision Algorithms: Theory and Practice 153-177, Springer Verlag 2000

Markov Random Fields on 3D Polygonal Meshes

Vedrana Andersen*, Henrik Aanæs*,
Andreas Bærentzen* and Mads Nielsen†

June 30, 2008

Abstract

Markov Random Fields (MRF) have been used extensively for solving Image Analysis problems at all levels. The local property of MRF makes it convenient for modeling dependencies of image pixels, and the MRFs-Gibbs equivalence theorem provides a joint probability in a simple form, making MRF theory useful for statistical Image Analysis. The majority of these MRF applications are built on a regular pixel grid, even though MRF theory, in its general form, does not require regularity of the sites.

Despite the simplicity and the flexibility of the formulation, MRF has not yet found its application in the field of 3D mesh modeling. Still, many concepts used in Image Analysis have over time been generalized and applied to mesh surfaces. This presentation investigates the use of MRF for formulating priors on 3D surfaces represented as triangle meshes.

Defining MRF on surface meshes is straightforward if one uses mesh connectivity to define the MRF neighborhood. The MRF site can be assigned to any of the basic entities of a triangular mesh — vertices, edges or faces. All three possibilities are discussed here. The idea is addressed by focusing on mesh smoothing, which is of great interest in many applications of geometry processing, e.g., computer vision and reverse engineering.

Firstly, a mesh-smoothing vertex process is described. It is a process that combines a smoothness prior described through MRF with the simple observation model into MAP-MRF framework. Smoothing

¹DTU Informatics

²DIKU

is done by iterative vertex replacement, using Metropolis sampling and simulated annealing scheme.

An edge process for detecting features (ridges) is shown next, where the feature-detecting function allows specifying the sharpness of the ridge. Edge process can be coupled with the vertex process in a feature-preserving mesh-smoothing method.

Lastly, a face process for normal filtering is presented. Filtering face normals is a first step in a different mesh smoothing procedure, and can also be coupled with the edge process.

The biggest challenge in formulating MRF on triangle meshes is dealing with its irregularity. A vertex can have a different number of adjacent vertices, and an edge can have a different number of adjacent edges. As a result, for example, one can not directly compare the potentials corresponding to two vertices.

Preliminary but promising experimental results are presented, proving the feasibility and demonstrating the use of MRF on triangular meshes. Developed priors and the optimization methods are discussed, and some possible improvements suggested.

MULTI-SCALE NATURAL IMAGES: A DATABASE AND SOME STATISTICS

David Gustavsson, Kim Steenstrup Pedersen and Mads Nielsen

DIKU, University of Copenhagen
{davidg, kimstp, madsn}@diku.dk

1. INTRODUCTION

Images contain different types of information, from highly stochastic textures, such as grass and fur, to highly geometric structures, such as houses and cars. Furthermore, most images contain a mix of geometric structures and stochastic textures.

It is well known from scale space theory that the image contents does not only depends on the objects in an image but also on the scale that the image has been captured ([8, 5]). At a coarse scale finer details are suppressed while the coarse scale structure are brought out. At a finer scale the coarse scale geometric structures are suppressed while the finer scale details are brought out.

Different image processing tools are suitable for different type of image contents. A tool suitable for a type of image content maybe useless for another type of image content. Most tools in image processing are very image content dependent. Segmentation of an image containing geometric structures calls for edge-based methods, while segmentation of an image containing texture calls for texture based segmentation methods (or pre-processing that transform the image textures to geometric structures).

A database containing an ensemble of image sequences containing the same scene captured at different scales is presented.

The main purpose (and applications) for collecting the database are:

Geometric Structure and Texture

The image contents depends on the scale that it has been captured at. By capture the same scene at different scales the image contents will differ - geometric structures will be transformed into texture and texture will be transformed into geometric structure. How does the image contents change over scales? How can the image contents be characterized in terms of geometric structure and texture? How can an image complexity measure be constructed that capture the image contents in terms of geometric structure and texture?

On what scale does a brick wall decompose in to a set of bricks, on what scale does a scrub decompose in to a set of twigs? How can this transitions be measured using an image complexity measure.

Zoom-In and Zoom-Out

The zoom-in and zoom-out problems have received a lot of research interest in the recent years ([2, 4]). A high resolution image should be shown on a small display in a mobile phone or camera, and a low resolution image should be shown on a large display or using a projector. Zoom-out (or sub-sampling) is also a common pre-processing step, motivated solely by computation time, in many image processing application.

Zoom-in: creating an image with higher spatial resolution from an image with lower spatial resolution.

Zoom-out: creating an image with lower spatial resolution from an image with higher spatial resolution.

Zoom-in is related to image interpolation, inpainting and texture synthesis. Common method used for zoom-out is low-pass filtering followed by sub-sampling, and block average.

What is the objective for zoom-in and zoom-out? Should the zoomed image be similar to the scene captured at the corresponding scale or should it just be visual appealing?

Segmentation - Cue Integration

Images contain a lot of edges, some of the edges are object boundaries will other edges are part of a texture. When is an edge an object boundary that can be used directly in the segmentation and the is it part of a texture (that can only be used indirectly in the segmentation)? An image complexity measure that characterize the image contents with respect to geometric structures and textures is informative for deciding if an edge is a boundary or port of a texture.

2. A MULTI-SCALE GEOMETRIC STRUCTURES AND TEXTURE DATABASE

The database contains images of the same scene captured at different scale. The camera that has been used is a Nikon D40X and three different objectives: 18-55 mm, 55-200 mm and 70-300 mm. The camera has been placed on a tripod stand facing the scene. A region of interest in the scene of such a size that it is present at all scales has been selected. The scene, with the region of interest approximately in the center, is captured at different scale by adjusting/changing the objective. The scene is captured at 15 different scale, the focal length is from 18 mm to 300 mm - roughly 4 octaves and 16

times magnification. A 1×1 regions in the least zoomed image corresponds to a 16×16 region in the most zoomed image. The image resolution is 2592×3872 .

The scenes selected for the database are mostly natural images containing both man-made environments - mostly buildings - and natural environments - trees, tree trunks and bushes. In many cases the same type of scenes has been captured but with different distance between the camera and the scene, which are change the image contents captured. By varying the distance between the camera and the scene, each set of images captured using a fixed focal length will be an ensemble of natural images.

The region present in all images in a sequence has been extracted, resulting in sequences of regions containing the same part of the scene captured at different scales.

3. NATURAL IMAGE STATISTICS

To verify the soundness of the database content, a number of well known statistical properties, with some extensions, of natural images is verified on the database. The soundness of the image database is verified on the ensemble of images in the database (i.e. using all images in the database), and on the ensemble of images captured using the same focal length (i.e. on sets containing one image from each sequence).

One of the earliest result in the area of characterization of natural images is the scaling property ([6, 7, 3]). The scaling property was first formulated as power spectra of a large ensemble of natural images follow a power law

$$S(\omega) = \frac{A}{|\omega|^{2-\eta}} \quad (1)$$

where ω is the spatial frequency, and A is a constant that depends on the overall contrast in the image. η is usually a small value and values close to 0.2 has been reported ([7, 3]). It should also be noted that η depends on the type of images ([9]) and that small image databases with specific contents - for example beaches and blue skies - may have η far from 0.2. The scale invariant property of natural images can also be expressed in the spatial domain using the correlation function.

It has been reported, [3], that the distribution of the partial derivatives of an ensemble of natural images can be modeled by an Generalized Laplacian Distribution

$$p(x) = \frac{1}{Z} e^{-|\frac{x}{s}|^\alpha} \quad (2)$$

where α and s are parameters estimated from the ensemble of natural images. The parameters s and α are related to the variance and kurtosis.

Compared with the Gaussian distribution, the Generalized Laplacian distribution (usually) has a sharper peak at zero and 'heavy tails'. Most natural images contain homogenous regions, objects under similar illumination, with similar or smoothly varying intensities which corresponds to the sharp peak at zero, at

the object boundary the intensities change rapidly which corresponds to the 'heavy tails'.

It is natural to consider how the size of homogenous regions in natural images are distributed. Alvarez et. al. ([1]) analyze the size distribution of homogenous regions in natural images, in terms of area and perimeter, and they show that the size distribution of homogenous regions in natural images follow a power law

$$f(s) = \frac{A}{s^\alpha} \quad (3)$$

where s is the area, A and α are image dependent parameters. The parameters α and A can be estimated by log-regression. For ensembles of natural images $\alpha \approx 2$, for individual images the α varies. For image containing larger geometric structures α is often smaller around 1.5, while for image containing small scale texture α is often larger around 3.0.

The statistics computed on the database and on the sequences are consistent with the result previous reported, but not identical.

4. REFERENCES

- [1] L. Alvarez, Y. Gousseau, and J.-M. Morel. Scales in natural images and a consequence on their bounded variation norm. pages 247–258, 1999.
- [2] Tony F Chan and Jianhong Shen. *Image Processing and Analysis - variational, PDE, wavelet, and stochastic methods*. SIAM, 2006.
- [3] Jinggang Huang and David Mumford. Statistics of natural images and models. *cvpr*, 01:1541, 1999.
- [4] Sune Hogild Keller. *Video Upscaling Using Variational Methods*. PhD thesis, University of Copenhagen, 2007.
- [5] Jan J Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.
- [6] D. L. Ruderman and W. Bialek. Statistics of natural images: Scaling in the woods. *Physical Review Letters*, 73(6):814–817, August 1994.
- [7] Daniel L. Ruderman. Origins of scaling in natural images. *Vision Research*, 37(23):3385–3398, 1997.
- [8] Bart M. ter Haar Romeny. *Front-End Vision and Multi-Scale Image Analysis: Multi-Scale Computer Vision Theory and Applications, written in Mathematica*, volume 27 of *Computational Imaging and Vision*. Kluwer Academic Publishers, 2003.
- [9] Antonio Torralba and Aude Oliva. Statistics of natural image categories. *Network: Computation in Neural Systems*, 14(3):391 – 412, August 2003.

The CVMT/AAU Sphere Gantry

Claus B. Madsen

Laboratory of Computer Vision and Media Technology, Aalborg University, Denmark
cbm@cvmt.aau.dk

1. Introduction

The reflectance characteristics of real world surfaces/materials are very complex and hard to model analytically. Image based techniques are rapidly gaining popularity in computer graphics, in particular for special effects.

With image based techniques the computer graphics rendering is somehow based on real images (hence the name). There is work on image-based modeling, rendering and lighting.

The computer vision and media technology lab at Aalborg University (CVMT/AAU) has designed and constructed a “Sphere Gantry”, essentially a rig with two rotational degrees of freedom, allowing for placing a camera or a light source at any position on a sphere around a test object.

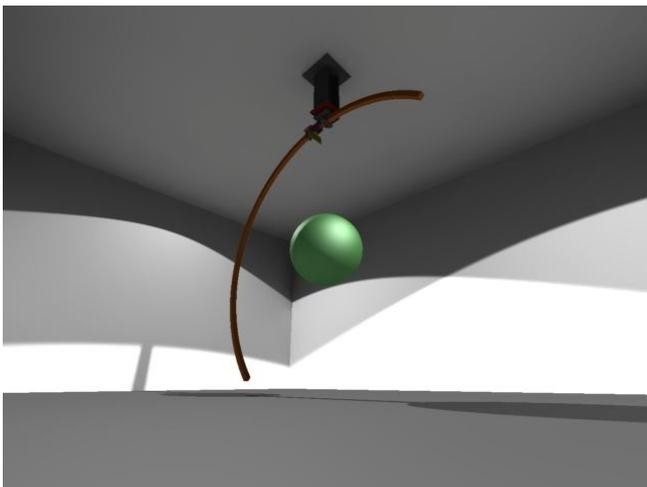


Figure 1: 3D rendering of the CVMT/AAU sphere gantry

With the sphere gantry it is possible to acquire images of test objects in a very systematic and controlled manner. We are using it for two different types of acquisition: 1) static illumination, moving camera, and 2) static camera, moving illumination.



Figure 2: The actual sphere gantry on the lab. Unfortunately the rig is not motorized so it is manually moved during acquisition.

2. Static illumination, moving camera

By mounting a high quality digital SLR camera on the application tray of the sphere gantry, placing an object in the gantry’s center, and making sure the illumination is held constant, we can acquire a large number of images of the test object from every possible angle.

We have used such data sets with around 300 images for modeling the appearance of a surface with spherical harmonics and made a real-time graphics rendering system to recreate the view-dependent appearance of the surfaces in real-time.

We have demonstrated that it is possible to very convincingly visualize complex glossy surfaces with highlights, and minute creases. The changes in glossiness caused by fingerprint left by humans while handling the material are also easily visible, see Figure 3.



Figure 3: View dependent appearance of glossy surface recreated in real-time based on 300 images acquired with the sphere gantry.



Figure 5: Image number 154 of the same sequence showing a completely different illumination direction.

3. Static camera, moving illumination

If, conversely, the camera is mounted on a tripod and the *light source* is placed on the sphere gantry application tray then it is possible to acquire a completely different data set for a set of objects. We basically take an image for every different illumination angle, and this sequence can subsequently be used for synthesizing the appearance of the scene under arbitrary illumination conditions.



Figure 4: Image 15 in an 348 image sequence with a scene illuminated from all positions on a sphere around the scene. In this image the light source is roughly vertically above the scene.



Figure 6: Average image of the entire sequence corresponding to the appearance of the scene had it been illuminated by a spherical light source placed around the scene (a “sky dome”).



Figure 7: Average of 3 randomly selected light source positions.

4. Conclusions

We have designed and constructed a rig which allows to make systematic acquisitions where either the camera or the light source is placed on all possible locations on a sphere around a test object.

Current experimentation involves using acquired datasets to synthesize view-direction-dependent appearance and illumination-direction-dependent appearance.

Acknowledgments

This research is funded by the CoSPE project (grant number 26-04-0171) under the Danish Research Agency. This support is gratefully acknowledged.

References

- [1] C.B. Madsen, B. Mortensen, and J. Andersen. Real-Time View-Dependent Visualization of Real World Glossy Surfaces. In *Proceedings:International Conference on Graphics Theory and Applications(GRAPP), Funchal, Madeira – Portugal, page 231 - 240, January 2008.*

Estimating Radiances of the Sun and the Sky

From a Single Image Containing Shadows

Claus B. Madsen

Laboratory of Computer Vision and Media Technology, Aalborg University, Denmark
cbm@cvm.t.aau.dk

1. Introduction

Augmented Reality is about inserting computer graphics objects into images of reality. One of the main difficulties in achieving a convincing illusion is to obtain a model of the illumination conditions in the scene so as to be able to render the virtual objects with shading and shadows that are consistent with the real scene.

Most work in Augmented Reality is based on recording and storing the scene illumination in an image, a so-called environment map or a light probe image. An environment map is a 360 by 180 degree (full omni-directional) image where each pixel corresponds to a direction in space, and the value of the pixel is the incident radiance from that particular direction.

Light probe images are used extensively for special effects in movie production and a tried and tested approach delivering very high illumination precision and convincing results. The main problem is that the approach assumes the illumination is static in the scene. Once acquired it is not allowed to change.

We have developed a new technique for estimating the illumination conditions in outdoor scenes from a single image. This technique does not require acquisition of a light probe image. Rather it assumes that some additional information is available for the image: 1) date and time for the acquisition of the image, 2) the longitude and latitude for the position on Earth where the image is taken, and 3) knowledge of the direction of gravity relative to the camera. These pieces of information can easily be made available building a clock, and GPS received and an inertial sensor into the camera.

With this information available it is straight forward to compute the position of the Sun relative to the camera coordinate system. Using this knowledge we have developed a technique which, based on the shadows present in the image, can estimate the radiances from the Sun and the Sky, respectively.

Using these radiances we can render augmented objects into the images with very realistic illumination regarding both shading and shadowing.

2. Approach

The approach is quite simple, and rests on a few assumptions: 1) there are predominantly diffuse surfaces in the scene, and 2) the Sun is shining. The steps are as follows:

1. Acquire the image, noting the date, time and position on Earth
2. Compute the Sun's position relative to the camera
3. Detect areas in the image which are in shadow which, with the technique we are using, as a bi-product yields the ratio between Sun + Sky irradiance to Sky irradiance
4. Use Sun's position and estimated irradiance ratio to compute the radiance of the Sky and of the Sun (up to a scaling factor)
5. Render virtual objects into the image



Figure 1: Image acquired on a sunny day



Figure 2: Image with shadows removed



Figure 3: Image with virtual objects rendered with estimated illumination

Figure 3 shows how we can convincingly render virtual objects into an image with the developed technique for estimating the scene illumination conditions based only on shadow information in the scene.



Figure 4: Example image involving multiple surfaces and textures



Figure 5: The checkered surface is used for calibrating the camera to the ground plane since we do not at present have an inertial sensor to provide the direction of gravity.

Conclusions

Our approach is based on an assumption that there are shadows in the image. We are presently working on developing techniques that will very robustly detect shadows if they exist, and we are also combine this work with another technique which can estimate the illumination in overcast weather conditions.

Acknowledgments

This research is funded by the CoSPE project (grant number 26-04-0171) under the Danish Research Agency. This support is gratefully acknowledged.

References

- [1] C.B. Madsen and M. Nielsen. Towards Probeless Augmented Reality – a Position paper. In *Proceedings:International Conference on Graphics Theory and Applications(GRAPP), Funchal, Madeira – Portugal, page 255 – 261, January 2008.*

Face Image Quality and its Improvement in a Face Detection System

Kamal Nasrollahi, Thomas B. Moeslund

Laboratory of Computer Vision and Media Technology, Aalborg University
Aalborg, Denmark, (e-mail: {kn, tbm}@cvmt.dk)

Abstract: When a person passes by a surveillance camera a sequence of images is obtained. Most of these images are redundant and usually keeping some of them which have better quality is sufficient. So before performing any analysis on the face of a person, the face at the first step needs to be detected. In the second step the quality of the different face images needs to be evaluated. Finally, after choosing the best image(s) based on this quality assessment, in the third step, if this image(s) is not satisfying a predefined set of measures for good quality images, its quality should be improved. In this work we are trying to develop a system to deal with the video sequences in these 3 steps.

1. INTRODUCTION

In order to have a fully automated vision based application related to human face processing like Face Recognition, Facial Feature Extraction, Video Conferencing and Human-Machine Interaction, the size and position of the face should be determined before, i.e. Face Detection.

Improving in Face Detection in any of the aforementioned systems causes the results of them to be more reliable. Face Detection is challenging due to the problems in (Yang *et al.*, 2002):

- **Pose:** The images of a face vary due to the relative camera-face pose (frontal, 45 degree, profile, upside down), and some facial features such as an eye or the nose may become partially or wholly occluded.
- **Presence or absence of structural components:** Facial features such as beards, mustaches, and glasses may or may not be present and there is a great deal of variability among these components including shape, color, and size.
- **Facial expression:** The appearance of faces is directly affected by a person's facial expression.
- **Occlusion:** Faces may be partially occluded by other objects. In an image with a group of people, some faces may partially occlude other faces.
- **Image orientation:** Face images directly vary for different rotations about the camera's optical axis.
- **Imaging conditions:** When the image is formed, factors such as lighting (spectra, source distribution and intensity) and camera characteristics (sensor response, lenses) affect the appearance of a face.

If the Face Detection is used in an application like surveillance camera, while a person is walking in front of the camera, a lot of face images will be extracted from that

person. Depending on the application, most of these images are useless and only keeping the best high quality one(s) will be sufficient for each person. So we need a mechanism to assess the quality of the face image, i.e. Face Quality Assessment.

If the quality of this best chosen face image(s) is not satisfying a set of predefined measures for good quality images we need another mechanism to improve its quality, i.e. Quality Improvement.

1.1 State-of-the-Art in this field

The proposed Face Detection systems in the literature can be categorized as follows (Yang *et al.*, 2002):

1. **Knowledge-based methods:** These rule-based methods encode human knowledge of what constitutes a typical face. Usually, the rules capture the relationships between facial features.
2. **Feature invariant approaches:** These algorithms aim to find structural features that exist even when the pose, viewpoint, or lighting conditions vary, and then use these to locate faces.
3. **Template matching methods:** Several standard patterns of a face are stored to describe the face as a whole or the facial features separately. The correlations between an input image and the stored patterns are computed for detection.
4. **Appearance-based methods:** In contrast to template matching, the models (or templates) are learned from a set of training images which should capture the representative variability of facial appearance. These learned models are then used for detection.

There are two important issues in Face Detection applications: Speed and Detection Rate. These two factors are

working against each other and a typical Face Detection application should to find a tradeoff between them.

There are a lot of well known examples of systems which have made acceptable tradeoffs between speed and detection rate. (Viola and Jones 2001) introduce AdaBoost with a cascade scheme and apply an integral image concept for face detection. They propose a two-class AdaBoost learning algorithm for training efficient classifiers and a cascade structure for rejecting none face images. (Cheng *et al.*, 2005) after compensating the colors of the input images and deskewing tilted faces locate mouth corners and determine a discriminate function for positioning eyes. (Zhong *et al.*, 2007) use a luminance-conditional distribution for modeling the skin color information and then by morphological operations extract the skin-region rectangles. Finally, they use template matching based on a linear transformation to detect a face in each skin-region rectangle. (Mansour *et al.*, 2005) propose a face detection algorithm based on light control, skin detection and color segmentation techniques. Their method detects the face rectangles that contain eyes and mouth and then constructs expected regions resulted from skin detection and color segmentation stages. Next they search inside them for any possible face features (eyes, and mouth) and pass these expected mouth and eyes rectangle to a neural network to confirm the presence of a face.

Assessing the quality of face images and trying to improve it or extracting a good quality image from poor quality one(s) are new fields in face analysis. In different works related to face quality assessment, different features of the face have been used including: Sharpness, illumination, head rotation, face size, presence of skin pixels, openness of eyes and red eyes. (Xiufeng *et al.*, 2007) have tried to standardize the quality of face images by facial symmetry based methods. (Adam and Robert 2007) have extracted 6 features for each face and after assigning a score to each feature, combines them into a general score. (Subasic *et al.*, 2005) consider more features and interpret the scores related to each feature as a fuzzy value. (Fronthaler *et al.*, 2006) have studied orientation tensor with a set of symmetry descriptors to assess the quality of face images. (Nasrollahi and Moeslund, 2008) have used a locally scoring technique to assign suitable scores to the involved features in the quality assessment. They have tested their system on video sequences.

The rest of this paper is organized as follows: in Section 2 the current study is described, Section 3 gives a brief discussion and Section 4 concludes the paper.

2. THE CURRENT STUDY

According to the block diagram of the proposed system (see Fig. 1) this study deals with the images taken from a surveillance camera in three steps: Face Detection, Face Quality Assessment and Improving the Quality of the Face Images which are described in the following subsections, respectively.

2.1 Face Detection

Our face detector (Nasrollahi *et al.*, 2008) uses the information of the skin color to separate skin or skin-like regions from non-skin ones. Then, for each of the skin or skin-like regions, some of the most important features of the face like size, number of holes inside the region, cross correlation between the regions and a face template, etc. are extracted. Based on these feature a cascaded classifier based on a neural network decides which of these regions can be a face. The accepted regions as a face go through the next steps of the system for further processing.

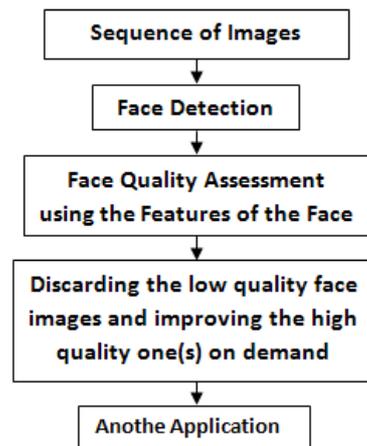


Fig. 1. The Block Diagram of the Proposed System

2.2 Face Quality Assessment

In (Nasrollahi and Moeslund, 2008) we have analyzed different features and found that 4 features are efficient for face image quality assessment. These features are out-of-plan-rotation, sharpness, brightness, and face size. In order to compare the different images of each person, we have used a locally scoring technique, which assigns a score to each of the aforementioned features. Then these scores are combined and normalized to obtain one score for each face.

2.3 Face Quality Improvement

If the best chosen images in the previous step are not satisfying a set of predefined measures for good quality images we would try to extract a good quality image from the poor quality ones. For each special purpose there is a special set of predefined measures, usually. For example for machine readable travel documents the International Civil Aviation Organization (ICAO) has defined a set of required conditions for the faces to be considered as high quality faces. Based on the important features in quality, there are a lot of techniques to improve the quality of the face. One of these important features which we are trying to improve is the pose of the face. We would like to use a 3D model of the face to produce a frontal face image from a non-frontal one. This part of the system is still under investigation.

3. DISCUSSION

The reliability of the system is highly dependent on the Face Detector. We use a Face Detector to be able to detect faces in image in real time and at the same time find an acceptable tradeoff between the number of false positives and false negatives. We use a set of features in Quality Assessment which are stable, reliable and easy to extract. We use the locally scoring technique to compare the values of same features in different images of a person.

Finally, in Improving the Quality of Face Image, for each of the involved features in the quality assessment we would try to develop reliable methods to improve the quality on demand.

4. CONCLUSION

The expected result for this study is a face detection application which has the auto ability of quality assessment of face images and tries to extract the good quality face images from the poor ones if need be. The three different part of the project will be tested using videos from surveillance cameras.

REFERENCES

- A. Fournery and R. Laganieri (2007). Constructing Face Image Logs that are Both Complete and Concise. **In:** *4th Canadian Conference on Computer Vision and Robot Vision*, IEEE Press, Canada.
- C. Chiang and C. Huang (2005). A robust method for detecting arbitrarily tilted human faces in color images. **In:** *Pattern Recognition*, vol. 26, no. 16, pp. 2518--2536.
- G. Xiufeng, Z. Stan, L. Rong and P. Zhang (2007). Standardization of Face Image Sample Quality. **In:** *ICB*, pp. 242--251, Springer-Verlag, Berlin.
- H. Fronthaler, K. Kollreider and J. Bigun (2006). Automatic Image Quality Assessment with Application in Biometrics. **In:** *International Conference on Computer Vision and Pattern Recognition*, IEEE Press.
- I.M. Mansour and S.R. Weshah (2005). A Robust Algorithm for Face Detection in Color Images. **In:** *Visualization, Imaging, and Image Processing*. Int. Conf., Spain.
- K. Nasrollahi and T.B. Moeslund (2008). Face Quality Assessment System in Video Sequences. **In:** *The First COST 2101 Workshop on Biometrics and Identity Management (BIOID 2008)*, Springer, Denmark.
- K. Nasrollahi, M. Rahmati and T.B. Moeslund (2008). A Neural Network Based Cascaded Classifier for Face Detection in Color Images with Complex Background. **In:** *Image Analysis and Recognition*. Int. Conf. Springer, Portugal.
- M. Subasic, S. Loncaric, T. Petkovic and H. Bogunvoic (2005). Face Image Validation System. **In:** *4th International Symposium on Image and Signal Processing and Analysis*, pp. 30--36.
- M. Yang, Kriegman D. and N. Ahuja (2002). Detecting Faces in Images: A Survey. **In:** *IEEE Trans Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34--58.
- P. Viola and M. Jones (2001). Rapid Object Detection Using a Boosted Cascade of Simple Features. **In:** *IEEE Computer Vision and Pattern Recognition*, pp. 511--518.
- Z. Jin, Z. Lou, J. Yang and Q. Sun (2007). Face detection using template matching and skin-color information. **In:** *Neurocomputing*, vol. 70, no. 4, pp. 794--800.