

Vol. 23, No. 1/2, 2014

Machine
GRAPHICS & VISION

International Journal

Published by
The Faculty of Applied Informatics and Mathematics
Warsaw University of Life Sciences – SGGW

Nowoursynowska 159, 02-776 Warsaw, Poland

in cooperation with

The Association for Image Processing, Poland – TPO

M a c h i n e G R A P H I C S & V I S I O N

Editor-in-Chief: Leszek J. Chmielewski, The Faculty of Applied Informatics
and Mathematics – WZIM, Warsaw University of Life Sciences – SGGW
Past Editor and Founder: Wojciech S. Mokrzycki
Honorary Editor: Juliusz L. Kulikowski

Executive Editors:

Ryszard Kozera, Warsaw University of Life Sciences – SGGW, Poland
Arkadiusz Orłowski, Warsaw University of Life Sciences – SGGW, Poland
Artur Przelaskowski, Warsaw University of Technology, Poland
Artur Wiliński, Warsaw University of Life Sciences – SGGW, Poland

Advisory Board:

Bengtsson E. , Uppsala University (SE)	Hlaváč V. , Czech Technical University, Prague, (CZ)	Palenichka R. , University of Quebec in Outaouais, Gatineau (CA)	Stepień C. , Warsaw University of Technology (PL)
Bhattacharya P. , Concordia University, Montreal (CA)	Kiciak P. , Warsaw University (PL)	Palus H. , Silesian University of Technology, Gliwice (PL)	Szyszkowicz M. , Ottawa (CA)
Chetverikov D. , Computer and Automation Research Institute HAS, Budapest (HU)	Klette R. , The Auckland University (NZ)	Pamuła J. , Academy of Fine Arts, Cracow (PL)	Śluzek A. , Abu Dhabi, Khalifa University (AE)
Choraś R. , University of Technology & Agriculture, Bydgoszcz (PL)	Kolingerova I. , University of West Bohemia, Plzeň (CZ)	Pieczynski W. , Institut National des Télécommunications, Evry (FR)	Tadeusiewicz R. , AGH University of Science and Technology, Cracow (PL)
Czachórski T. , Institute of Theoretical and Applied Informatics, PAS, Gliwice (PL)	Krasnoproshin V. , Byelorussian State University, Minsk, (BY)	Postaire J.-G. , Université des Sciences et Technologies, Villeneuve d'Ascq (FR)	Titov V.S. , Technical University of Kursk (RU)
Flasiński M. , Jagiellonian University, Cracow (PL)	Kurzyński M. , Wrocław Technical University (PL)	Punys V. , Kaunas University of Technology, Kaunas (LT)	Tombre K. , INRIA Lorraine, Villers-les-Nancy (FR)
Gagalowicz A. , INRIA, Le Chesnay (FR)	Leonardis A. , University of Ljubljana (SI)	Sharma A. , Panjab University, Chandigarh (INDIA)	Truphanov M. , Kursk State Technical University (RU)
Gillies D. , Imperial College, London (GB)	Malina W. , Gdańsk University of Technology (PL)	Skala V. , University of West Bohemia, Plzen (CZ)	Valev V. , University of North Florida, Jacksonville (US)
Grabska E. , Jagiellonian University, Cracow (PL)	Materka A. , Technical University of Łódź (PL)	Stapor K. , Silesian University of Technology, Gliwice (PL)	Velichova D. , Slovak Technical University, Bratislava (SK)
Hillion A. , IT - TELECOM Bretagne, Brest (FR)	Myszkowski K. , Max-Planck-Institute for Computer Science (DE)		Wojciechowski K. , Silesian University of Technology, Gliwice (PL)
	Nagahashi H. , Tokyo Institute of Technology (JP)		Zamojski W. , Wrocław Technical University (PL)
	Nieniewski M. , University of Łódź (PL)		

Editorial Office Secretary: Luiza Ochnio, Warsaw University of Life Sciences – SGGW

TeXsetting by Urszula Kruś

Language verification by Beata Konikowska

Vol. 23, No. 1/2, 2014

Machine
GRAPHICS & VISION

International Journal

Published by
The Faculty of Applied Informatics and Mathematics
Warsaw University of Life Sciences – SGGW

Nowoursynowska 159, 02-776 Warsaw, Poland

in cooperation with

The Association for Image Processing, Poland – TPO

The goal of **Machine GRAPHICS & VISION** is to provide a medium for exchanging information on image processing and image communication between the computer and its environment.

The journal concentrates on computational methods of synthesis and analysis of the models underlying computer generated, analyzed or processed imagery. In particular, the emphasis is on the methods of:

- image synthesis,
- computational geometry,
- three-dimensional imaging,
- computer graphics art,
- diagrammatic representation and reasoning,
- graphical data presentation and knowledge-oriented visualization,
- animation,
- image models and transforms,
- mathematical approaches to image analysis and processing
 - analytical geometry
 - statistical methods
 - functional analysis and spectral methods
 - logical methods
 - formal linguistic methods
 - fractals
- visual pattern matching and recognition
- three-dimensional scene mapping and modeling
 - statistical methods
 - spectral methods
 - structural methods
- motion capture
- composite image understanding
- monocular and stereo vision
- modeling of human visual perception

The journal encourages papers on research and applications to processing of scientific, medical and technological images, engineering design, graphical arts, environment protection, etc. State-of-the-art, expository as well as tutorial papers in the above-mentioned areas are welcome.

Copyright 2014 Faculty of Applied informatics and Mathematics (WZIM), SGGW

It is the condition of publication that manuscript submitted to this journal has not been published and will not be simultaneously submitted or published elsewhere. By submitting the manuscript, the authors agree that the copyright for their article is transferred to the publisher if and when the article is accepted for publication. The copyright covers the exclusive rights to reproduce and distribute the paper, including reprints, photographic reproductions, microforms, diskettes, Internet, or any other reproductions of similar nature, and translations. No part of this publication may be reproduced, stored in a data-base system or transmitted in any form or by any means, whether electronic, electrostatic, mechanical, on magnetic tape, by photocopying, recording or otherwise, without prior permission in writing from the copyright holder.

The copyright owner's consent does not extend to copying for the purposes of: general distribution, promotion, creation of new works, or resale. Specific written permission must be obtained from the publisher for copying. Please contact the **Editorial Office: Faculty of Applied Informatics and Mathematics (WZIM) of the Warsaw University of Life Sciences – SGGW, Nowoursynowska 159, 02-776 Warsaw, Poland, e-mail: MGV@sggw.pl, <http://mgv.wzim.sggw.pl>**

Published Quarterly

Subscription. **MG&V** (ISSN 1230-0535) can be ordered directly through Foreign Trade Enterprise ARS POLONA Joint Stock Company, Obrońców 25, 03-933 Warszawa, Poland
Fax: +48 22/509 86 48, Tel. +48 22/509 86 63, e-mail: arspolona@arspolona.com.pl

IMPROVED ILLUMINATION CORRECTION THAT PRESERVES MEDIUM-SIZED OBJECTS

Anders Hast¹ and Andrea Marchetti²

¹*Department of Information Technology, Uppsala University, Uppsala, Sweden*

²*Consiglio Nazionale delle Ricerche, Institute of Informatics and Telematics, Pisa, Italy*

Abstract. Illumination correction is a method used for removing the influence of light coming from the environment and of other distorting factors in the image capturing process. An algorithm based on the luminance mapping is proposed that can be used to remove low frequency variations in the intensity, and to increase the contrast in low contrast areas when necessary. Moreover, the algorithm can be employed to preserve the intensity of medium-sized objects with different intensity or colour than their surroundings, which otherwise would tend to be washed out. Furthermore, examples are given showing how the method can be used for both greyscale images and colour photos.

Key words: illumination correction, luminance mapping, image stitching, image mosaicing, vignetting.

1. Introduction

Varying illumination in captured images is a common problem in many different areas, such as Photography, Microscopy and Aerial imaging. Hence different names are used for both the problem and the remedies across different disciplines. Vignetting (Yu, 2004 [19]; Zheng et al, 2006 [24]) occurs due to different mechanisms (Goldman and Chen, 2005 [20]) that cause brightness falloff away from the image centre, and is prevalent in photography. The mechanisms are primarily: *mechanical-, optical- and natural vignetting* (Sidney 2002 [12]). The first one is due to the incoming light being blocked by external objects, such as filters or lens hoods. Optical vignetting, on the other hand, is due to the dimensions of the lens, and can be remedied by closing down the aperture. The full aperture will yield gradual darkening towards the corners, and thus a brighter spot in the middle. Natural vignetting, or natural illumination falloff, is proportional to the fourth power of the incoming light angle cosine. The remedy is to use a gradual grey filter or post-processing, like the method proposed herein, which we will refer to as *Illumination Correction*.

Non-uniform illumination, or even dirty lenses and dust (Young, 2000 [9]), cause problems in Microscopy (Leong et al, 2003 [14]), and the remedying procedure is often called illumination or shading correction. In Magnetic Resonance Imaging (MRI), varying shade is known as RF-inhomogeneity or bias (Agus et al, 2007 [25]; Ardizzone et al, 2007 [26]). In Face Recognition, varying illumination is a challenging problem (Guo

et al, 2005 [23]; Zhu et al, 2003 [17]) and the remedying method is, besides illumination correction, also called illumination normalization. The illumination itself is often referred to as background light. In Mammography, contrast-limited adaptive histogram equalization (CLAHE) has been used (Puff et al, 1994 [5]). In aerial and satellite photos, physical lighting also affects relief presentation (Rocchini et al, 2005 [22]; Tzelepis et al, 2003 [16]). However, it is also a problem for image stitching and mosaicing of panoramas, where the focus often is put on making the transition from one image to the other as smooth as possible (Burt and Adelson, 1983 [2]; Levin et al, 2004 [18]; Pérez et al 2003 [15]).

In this paper we will focus on a method that was previously developed for *retrospective illumination correction of greyscale historical aerial photos* (Hast and Marchetti, 2011 [30]). The method can be used both for minimizing the change in intensity over the image and for increasing the contrast in low contrast areas. This is important, as the contrast is generally lower in high and low intensity areas. We also present an important enhancement of the method, aimed at preserving the original intensity of medium-sized objects having intensity different than that of their surroundings. Here by *medium-sized objects* we mean objects that are much smaller than uneven illumination varying across the image, but still much larger than small details in the image. Preservation of their intensity would otherwise be impossible, as the goal of any illumination correction method is to minimize the changes in intensity across the image, which will affect such objects. Furthermore, we show how the new method can also be used for colour images, and that it will help preserve colours better.

2. Previous Research

There are many methods for illumination correction, each of them tailored to the specific field where it is used, as mentioned in the introduction. There are many different ways to handle the problem of varying illumination. One of the first approaches published was to use a homomorphic filter (Oppenheim et al, 1968 [1]), which takes advantage of the fact that the intensity in the image can be divided into two components: the illumination of an object and the amount of light reflected by that object (Pajarez et al, 2005 [21]). A homomorphic filter uses Fourier transform to remove from the image illumination, which is characterized by low spatial variations, and to keep reflectance, which is characterized by high frequency changes. Hence multiplication needs to be converted to addition, which can be done by taking the logarithm of the functions, as Fourier transform only can be used when the noise is an additive term. The main disadvantage of this approach is that it requires the use of image size padding when the image size is not a power of 2, which can lead to distortions in the boundary regions (ERDAS, 2010 [29]).

Another way to handle the problem (Zhu et al, 2003 [17]) is to model the lighting change as a local affine transformation (AT) of the pixel value, where the low spatial

variations caused can be removed using a multi-resolution low pass filter that estimates the background lighting (Zhu et al, 2003 [17]). A similar approach can be adopted using a Gaussian filter (Leong et al, 2004 [14]). Others handle the background light using a reference picture acquired in a controlled environment with known lighting conditions and a white object (Yu, 2004 [19]). This is, of course, not always possible, especially in retrospect. Yucong et al (Yucung et al 2005 [23]) use histogram equalization combined with ICR (Illumination Compensation based on Multiple Regression Model) (Ko et al, 2002 [11]), which aims at finding the plane that best fits the intensity distribution in the image using a multiple regression model. Then this plane is used to remove illumination of face images.

Local Range Modification (LRM) (Fahnestock and Schowengerdt, 1983 [3]) finds interpolated minimum and maximum pixel values in a neighbourhood (contextual region), and stretches them to the desired range. Similarly, the contrast-limited adaptive histogram equalization method (CLAHE) (Pizer et al, 1987 [4]) stretches the histogram in its contextual region. This improves the image in such way that the contrast is enhanced. In microscopy (Yuong et al, 2000 [9]) make use of both homomorphic filters and morphological filtering, which yields an estimate of the background illumination employed to correct images. Others have developed methods for finding the *vignetting function* from a single image (Zheng et al, 2006 [24]) or from multiple images, with or without the response curve (Goldman and Chen, 2005 [20]). Wavelet-based approaches for image and contrast enhancement are also popular (Reeves and Jernigan, 1997 [8]), and can be used to enhance contrast in digital mammography (Laine et al, 1995 [7]). Yet others make use of a combination of methods, such as wavelets and the homomorphic filter (Ashiba et al, 2008 [27]; Yoon and Ro, 2002 [13]).

3. The Proposed Method

A part of the proposed method was recently presented as a poster (Hast and Marchetti, 2011 [30]). Here we present the novel idea of preserving medium-sized areas with different intensity, and show how the new method can be used for colour images.

3.1. The Main Idea

The main idea of the previous paper was to use the luminance mapping (Hertzmann et al, 2001 [10]), which is usually employed in colour correction (Vrhel and Trussel, 1994 [6]). Here a linear map is applied that matches the means and deviations of luminance distributions, i.e. the intensity. If $A(x, y)$ is the luminance of a pixel in image A , then it is remapped using the distribution of image B :

$$A(x, y) = \frac{\sigma_B}{\sigma_A} (A(x, y) - \mu_A) + \mu_b . \quad (1)$$

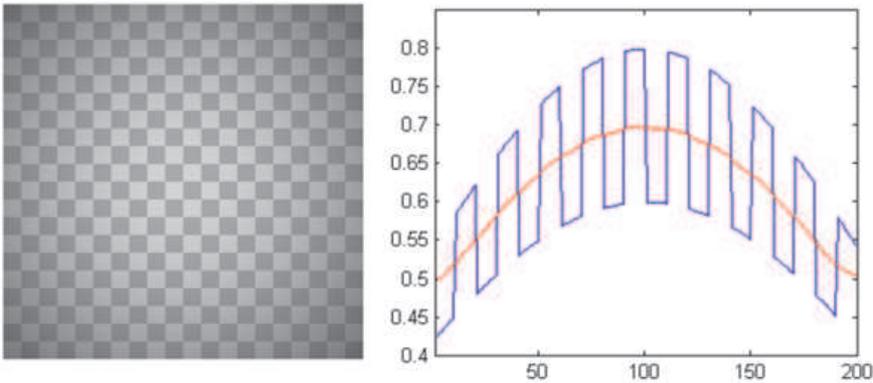


Fig. 1. The image on the left contains an obvious shading artefact, visible as intensity falloff away from the centre. The diagram on the right shows a cut section taken in the middle depicted in blue, and the local mean – in red.

where μ_A and μ_B are the mean luminances, and σ_A and σ_B are the deviations of the luminances, both taken with respect to the luminance distributions in A and B , respectively. The result of such remapping, if applied to all three colour channels, is that image A will have colours looking similar to those in image B .

Let us take as an example the image of the checkerboard on the left of Fig. 1, which has an obvious intensity variation. Looking at the horizontal cut section taken from the centre and depicted on the right, we can see the intensity variation over the cut. The *local mean* is the red curve that runs along the centre of the curve. If the *local mean* is translated to the *global mean*, i.e. transforming the curve to a line, then a major portion of the varying light is removed.

A *low-pass filter* is used to remove an estimate of the illumination that contains the lowest frequencies in the image, and thus an approximation of the *local mean* μ_A . This is done by using a Gaussian filter (3) modified in such a way that it also compensates for the borders. The trick of the trade is to divide the result by a likewise smoothed image that before the smoothing contained the intensity 1.0 everywhere. The smoothed image will hence contain 1.0 everywhere except at the borders, where the values will be lower – and this will compensate for the corresponding effect in the input image. Nevertheless, the mask must be quite large, up to the input image size.

Since the illumination is computed using very large Gaussian masks, this approach becomes impractical even with modern computers for very large photos. In the case of historical aerial photos, we use images of sizes up to 5000×6000 pixels, and the mask can be just as large. Nonetheless, it is still possible to compute the illumination using a downscaled version \mathbf{h} of the captured image H , which is therefore blurred much faster,

and then is resized to its original size. Since the result will be a heavily blurred image containing smooth and slow variations, an interpolation of such image will affect the result very little. The local mean is hence computed as a Gaussian filtered image using the convolution:

$$\mu_a(x, y) = G * \mathbf{h}(x, y) \quad (2)$$

where

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} . \quad (3)$$

The width of the distribution used to delimit the low frequencies is determined by σ . The larger the value of σ , the greater the smoothing effect. The resulting blurred image will be very close to what the full scale Gaussian filter would produce, and will not be detrimental to the final quality, as long as the downscaling is reasonable. However, it should be mentioned that in Figs. 1 and 2 the scale factor of 1 was used, i.e. there was no downscaling at all.

In the next step, the local deviation σ_A is computed as the Gaussian blurred absolute value of the difference between the downscaled image \mathbf{h} and μ_a :

$$\sigma_a(x, y) = G * |\mathbf{h}(x, y) - \mu_a(x, y)| . \quad (4)$$

The local deviation σ_A is computed by upscaling σ_a bilinearly to the full size. The illumination-corrected image \tilde{H} is computed as:

$$\tilde{H}(x, y) = \frac{\sigma}{\sigma_A(\mathbf{x}, \mathbf{y})} (H(x, y) - \mu_A(x, y)) + \mu . \quad (5)$$

Here σ is computed as the mean of σ_a , i.e. the mean variance in the whole downscaled image, and μ is the mean of μ_a , i.e. the overall mean intensity of that image. This equation is used in three steps with the same size of the Gaussian kernel for computing Eqs. (4) and (5). The result is shown in Fig. 2. The proposed contrast enhancement using the luminance mapping yields a straightened curve, where the amplitude is almost the same along the whole curve and the shading artefact is removed from the image.

The proposed method yields satisfactory results for images containing textures with small variations, even if they are quite drastic, as the checkerboard in Figs. 1 and 2. However, if the image contains medium-sized objects with different intensity than their surroundings as in Fig. 3 (left), then the result obtained using the method as explained so far will be quite disappointing, as shown to the right. The black and white squares (medium-sized objects) tend to get the same grey intensity as their surroundings, and give the impression of being washed out.

Fig. 4 shows on the left a cut section of the same image as that on the left in Fig. 3. A Gaussian filter is used to compute the mean, depicted in red, using different σ 's in the top and bottom rows. The corresponding results of the compensation are shown

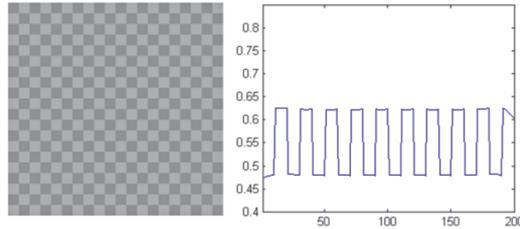


Fig. 2. The image from Fig. 1 and the cut section after applying the proposed algorithm. The low frequency has been removed, and the contrast – enhanced.

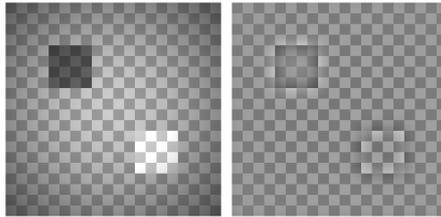


Fig. 3. The image on the left contains, except for vignetting, also one black and one white object. As seen to the right, these will be washed out, since with the original version of the proposed algorithm they tend to get the same intensity as the rest of the image.

on the right. Obviously, the result shown in the top row has lower variation along the whole curve. However, in the lower row the black object is preserved much better, even if the curve is not so straight elsewhere as in the top row, i.e. some illumination remains. Hence a solution is needed that gives us the best of both cases.

The new idea is to incorporate a *band pass filter* in the original approach, which temporarily diminishes the objects having different intensity by simply removing them from the original image. Hence, what is done is actually band rejection, following which a low pass filter is used to remove the varying intensity in the image. Finally, the objects are reinserted by simply adding the band pass-filtered image. This process is also repeated using different mask sizes for the low pass filter, like in the original approach. The band pass filter will therefore be chosen depending on the size of the objects that are to be preserved.

Fig. 5 shows how the image in Fig. 3 (left) is band pass-filtered (top left), and the original image is suppressed (top right) by the band pass image. The lower left picture shows the result after one iteration, where the low pass-filtered image is enhanced by the band pass image. Clearly, it has lower intensity variation, and the black and white squares are better preserved. Nevertheless, as there is some small variation across the

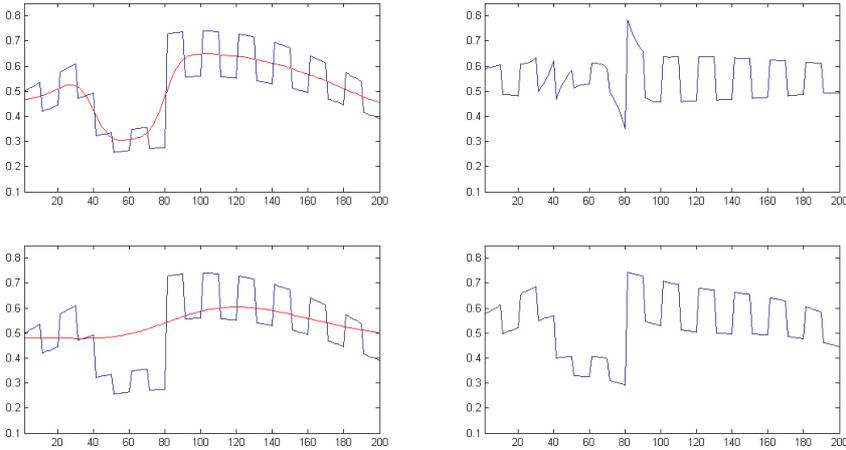


Fig. 4. On the left there is a cut section of the image that cuts through the black square. The average is depicted in red, using Gaussian filters with different values of σ . The corrected curves are shown on the right. The top curve has a lower variation over its whole length. The bottom curve preserves the black square, but is more bent elsewhere compared to the top one.

band pass-filtered image, the whole process needs to be repeated, just like in the original method. The result after three iterations is shown on the upper right. The varying illumination has been removed, and the objects are preserved.

In our tests we used a band pass filter known as a difference of a Gaussians filter (DOG) (Russ 2011 [31]). It is constructed as the difference of two Gaussian low pass filters with different values of σ . Hence it yields an image that contains medium-size objects, but also removes a major part of low variations in the image, as well as high frequency details. Together with increasing the contrast in low-contrast regions, the algorithm will also decrease the contrast in the areas where it is already high. Hence it is often necessary to avoid contrast enhancement in those areas. This can be easily done by not allowing σ_A to exceed some predefined percentage of σ_B . In this manner, we can adjust the image to obtain the desired result.

3.2. Repeating and Halving

It was found earlier that repeating the algorithm a couple of times using different σ 's gave a more pleasing result, and this is still true for our new approach. At first, such a method was applied only because it helped preserve medium-sized objects of different intensities, but it turned out to have a rather small impact on such objects. What is more important is that it helps remove low frequency variations in a better way, especially

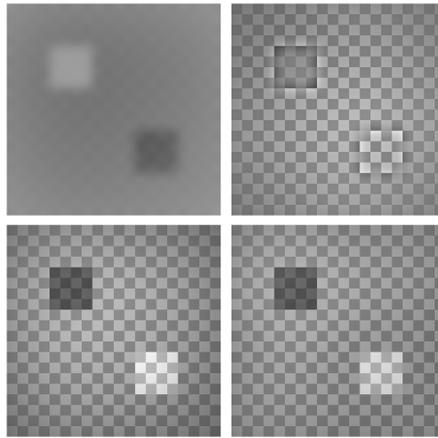


Fig. 5. The result of the band pass filter is shown on the upper left (scaled to the $[0..1]$ range), and the result of its adding to the input image – on the upper right. The lower left picture shows the result of one step of the previously explained illumination correction, involving the removal of the band pass-filtered image. Clearly, the fields are not suppressed and the shading is corrected; however, more cycles are needed to give the desired result, which is shown on the lower right.

when increasing the contrast, than using a mask of the same size one or several times, as shown in Fig. 6. There is nothing that prevents us from repeating the method fewer or more times than three. However, empirical tests have established that three times are often good enough, and that more iterations do not improve the result much. In any case, this approach needs to be used when preserving medium-size objects, as the band pass-filtered image contains some small variations. Repeating the process gradually eliminates these variations. It has been noted that halving of the mask gives both less distortion and fewer problems at the image edges. Moreover, it has been found that doubling of the mask size from size a to b gives an almost identical result as halving the size from b to a . In the method presented here we have chosen to halve the mask size in each step.

3.3. Scaling

The original approach proposed that both the local mean and the deviation be obtained by a Gaussian filter, where both can be computed from downscaled versions of the original image, and then upscaled by bilinear (or bicubic) interpolation. This will decrease computation time substantially, as Gaussian convolution is computationally expensive; however, it will not affect the result for moderate downscales.

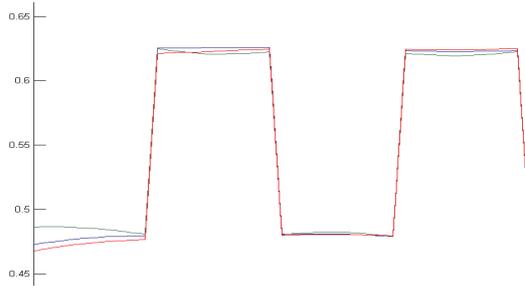


Fig. 6. A close-up of the curves obtained using different techniques for removing the low frequencies and increasing the contrast. The blue curve is the proposed method, which includes repeating the process three times while halving the mask size. The green curve is the result of using the same mask size three times. It is generally more bent at the tops and bottoms, where it should be straight. The red curve is the result of one iteration only, and it is straighter at the tops and bottoms, but decays too rapidly at the image edges.

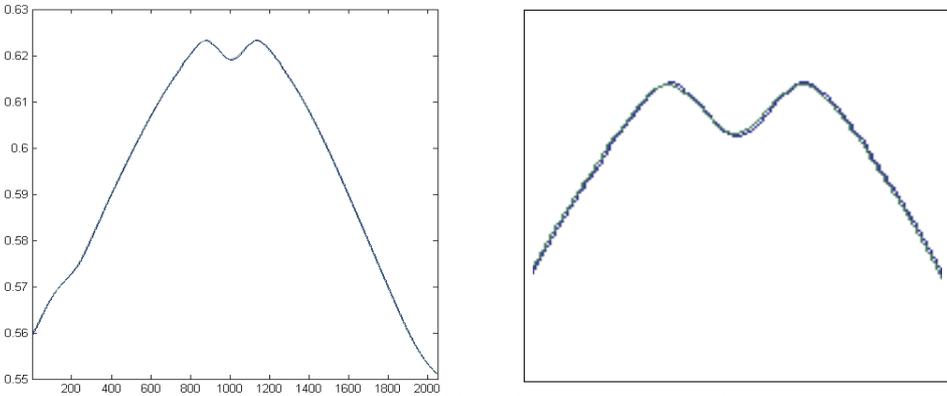


Fig. 7. A comparison of a cut section in the blurred image, using a scaled image (200×200 , green curve) for Gaussian convolution and the full image (2048×204 , blue curve). Usually, the difference is negligible for reasonable scales. Left: full graph; right: close-up of its central part.

As shown in Fig. 7, the difference between blurring a smaller image using Gaussian convolution and then interpolating it to full scale compared to blurring the full scale image directly is negligible. The reason is that the blurred image becomes smooth, and interpolation of a smooth curve still yields a smooth curve as long as the scale is not too large. Nonetheless, Gaussian convolution is time-consuming and, for instance, the use of the full size image took ten times more time to compute in our implementation of the algorithm than scaling each side by four (13.0 s compared to 1.3 s for a 800×800

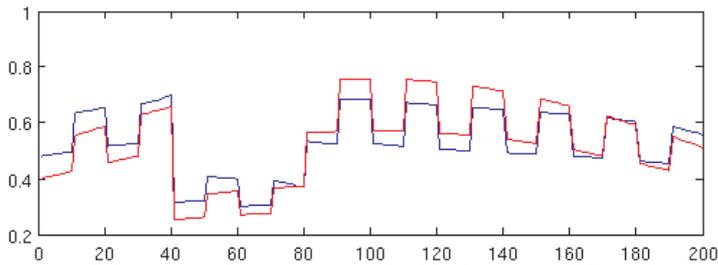


Fig. 8. A cut section of the image that intersects the black square. The red curve is the original image with a clearly visible illumination artefact. The blue is the illumination-corrected image, where the curve is much more straight.

image). Hence the proposed method of working with downscaled images is an important improvement when it comes to computational cost, while still maintaining the quality.

The result of using the proposed algorithm for the artificial image in Fig. 3 (left) is shown in Fig. 8. The red curve is a cut plane of the image in Fig. 3 (left), and the blue curve is a cut plane of the resulting image (Fig. 5, bottom right). The blue curve is straighter, which means that illumination has been removed, while the object is still preserved.

4. Results

Let us examine some results and possibilities of using the proposed method for illumination correction. Fig. 9 shows a historical photo of the river Arno and the west part of Pisa taken during WW-II.

In the top row there are some cropped original photos with different degrees of uneven illumination. The photos in the middle row have been processed with the original method. Both the large, bright field in the leftmost image and the dark field at the top of the rightmost image, as well as the river, have become greyish and are washed out. One can also note a ripple effect on a thin line outside the borders of the field and the river, which is a bit darker outside the field and a bit lighter outside the river. In the bottom row, the new approach is used. As a result, both the field and the river now have more accurate intensity, and the ripple effect is no longer visible.

4.1. Colour Images

Figure 10 shows some results of applying the proposed method to colour images. The leftmost image in the top row shows a wall, and the proposed method was applied to the V component of the image transformed to the HSV space (Sonka, 2008 [28]). This allowed for working solely on the image intensity, regardless of the colour. The said

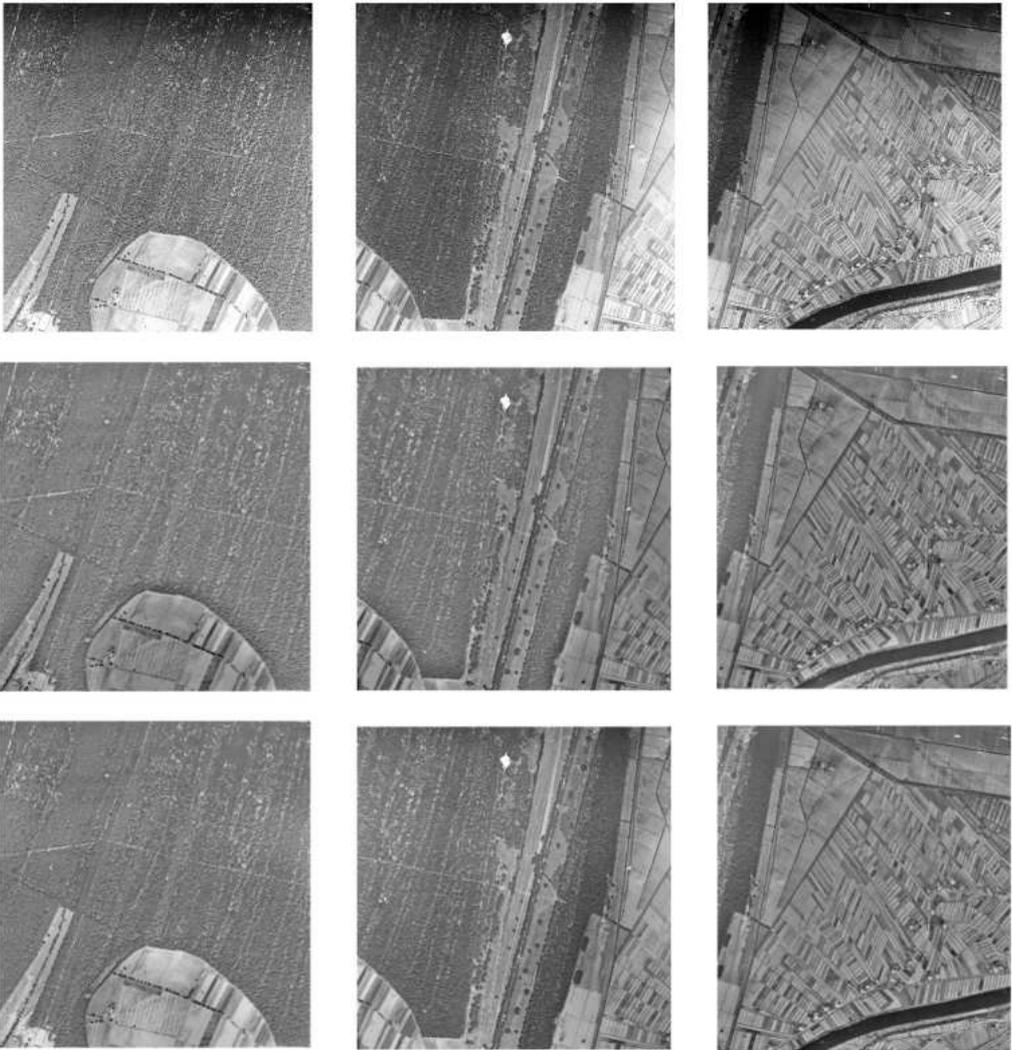


Fig. 9. ©MiBAC-ICCD, Aerofototeca Nazionale, fondo RAF. A series of historical photos of Arno close to Pisa taken during WW-II in the top row, showing different degrees of uneven illumination. In the middle row, the original method is used to correct the lighting. Both the large fields and the river are washed out, and their borders also show a ripple effect. In the bottom row, the new approach is used. Both the fields and the river have more accurate intensities, and the ripple effect is no longer visible.

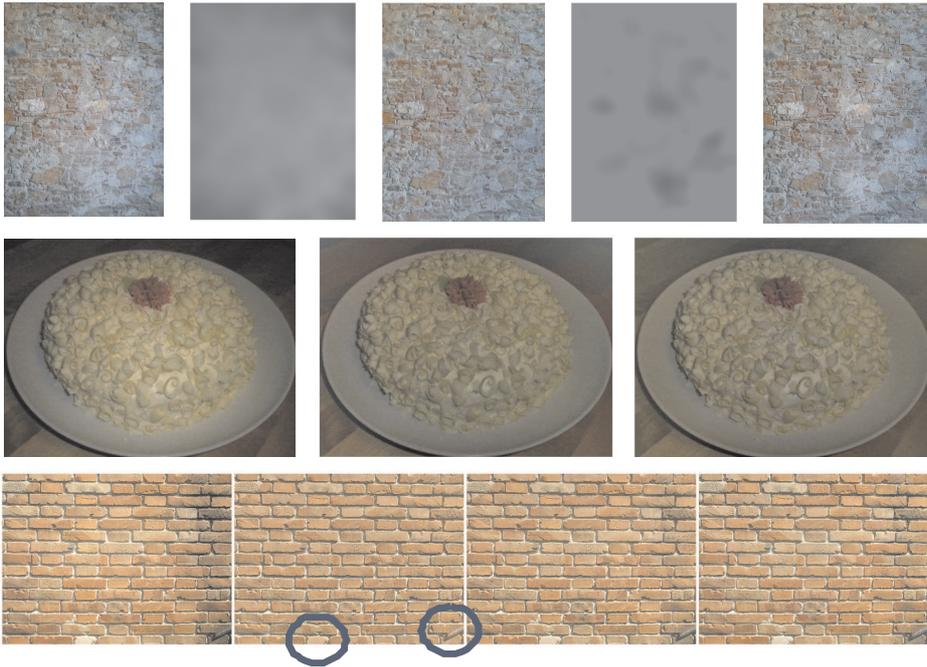


Fig. 10. ©Anders Hast. Top row, from left to right: Wall in Pisa, its uneven illumination, wall corrected using the original method (HSV, V only), band pass image, corrected using the new method (HSV, V only). Middle row, from left to right: Cake, corrected using the new method (HSV, V only), corrected using the new method (RGB). Bottom row, from left to right: wall with small intensity variation, corrected with the original method (RGB), with the new method (RGB), and again with the new method (HSV, V only).

image contains very little variation in illumination, as seen in the next image to the right. The subsequent image to the right shows the result of using the original method, and we can see that the whiter parts of the stone texture have become grey and dull. With the new method, these whiter parts become separated (the second image to the right). They are used for compensation according to the proposed method, and the result is the rightmost image. The middle row shows a cake in a rather large variation of illumination. Since the upper part is close to black, the conversion between RGB and HSV yields an undetermined colour, and the upper corners in the image are not represented correctly. The image to the right is processed on all three colour channels (RGB) directly. Now the upper corners of the image are represented correctly, and the result is better.

In the bottom row there is a wall with low intensity variations (left). When working on RGB with the old method, we obtain incorrect colours, as can be seen in the next



Fig. 11. ©Anders Hast. Left: Sunset on the beach. Right: image corrected using the new method (RGB). The beach is now more visible, while the sky has become darker.

image to the right. The dark iron bar, as well as the white part of a brick, both marked by blue circles, tend to become reddish. In the first image to the right the new method has been used for RGB, while in the second image to the right the new method has been employed to process the V part of HSV. The results for the latter two are very similar, and the colours in the mentioned critical parts are obviously better represented. The choice of the method is determined by the amount of colour information present in the image. If there are parts which are close to either black or white depending on uneven illumination, then working on all three colour channels (RGB) is necessary; otherwise the alternative yielding the best results can choose.

The new method can also be used to attenuate parts of an image that are very dark, and where the details are barely visible. Hence the proposed method need not be used just for normalizing the illumination in the whole picture, and can be employed for processing images to obtain other effects. Figure 11 shows (on the left) a photo of a sunset, where the chairs on the beach are barely visible. To the right there is the result of processing each colour channel separately. Now the chairs are clearly visible, and the method allows for enhancing the contrast so that the image seems to have been taken in much stronger light than there actually was. The drawback is that the sky becomes darker, as the algorithm tries to equalize the differences in illumination. This can easily be handled by interpolating between the images to obtain a desirable result.

5. Discussion

When using the proposed method in practice, the user must provide some parameters. Scaling was briefly discussed in 2.3, and can make the whole process a lot faster. Nonetheless, the scale factor cannot be too large. For example, the scale factor of 10 is almost 60 times faster (2.4 s in matlab) in our implementation than the scale factor

of 1 (151 s in matlab) for the rightmost photo in Fig. 9. However, the mean difference between the images is less than 0.08%, and is not noticeable in their visual inspection. Nevertheless, we propose that the scale parameter be set as low as possible while still giving acceptable total computational time.

The number of iterations was discussed in Section 2.2, and the value of 3 was used for the images we have tested, even if the difference sometimes was very small when using just one or two iterations. Nonetheless, the tests indicate that four or more iterations do not give any visible improvement, and therefore we have used the default value of 3 in our tests.

Other parameters are the σ 's for the low and band pass filters discussed in Section 2.1. The first must be set depending on how much of the image is occupied by uneven light. In our experience, this is quite similar for most of the images in a set of aerial images taken during one flight, so a default parameter can be chosen for the first image, and then reused for the others. The second σ depends on the size of the medium-sized objects to be preserved. As this can vary a lot from image to image, the user must keep adjusting this parameter until the output image looks good.

Moreover, we have added a parameter that increases the negative amplitudes of the band pass filtered image, and another one that increases the positive amplitudes. In this way, we can choose to what extent the white and black objects shall be preserved, respectively. These two parameters should also be gradually changed by the user until a satisfactory image is obtained. If there are no medium-sized objects, these parameters are set to zero, and should therefore be default values.

What happens if the presented method is applied to an image that has very little illumination differences, or perhaps even none? The image should, of course, remain more or less unchanged, as shown in Fig. 12 in the top row. The input image to the left is processed by the method with the previously mentioned parameters set to zero. Once again, we get the washed-out appearance of the medium-sized objects in the middle image. To the right, both these parameters are set to 1.0. A cut section through the image running through the black object is shown at the bottom, where the left (input image) is the blue curve, the middle image (corrupted) is the green curve, and the rightmost image is represented by the red curve, which is close to the blue curve. The reason why it is not identical to the blue curve is that the band pass filter cannot extract the objects perfectly. We have not tried to find the best band pass filter in this paper, but rather to show the main idea. The filter used was a simple DOG filter, and we are aware that other band pass filters might give an even better result.

It should be noted that the checkerboard test image is rather extreme in view of its square pulse content, compared to aerial photos which contain more variations. Figure 13 shows a cut section (going downwards from the top) of the rightmost image from the top of Fig. 9. The top graph, where the uneven illumination is clearly visible, represents the original image. The middle graph shows the result of applying the original method with

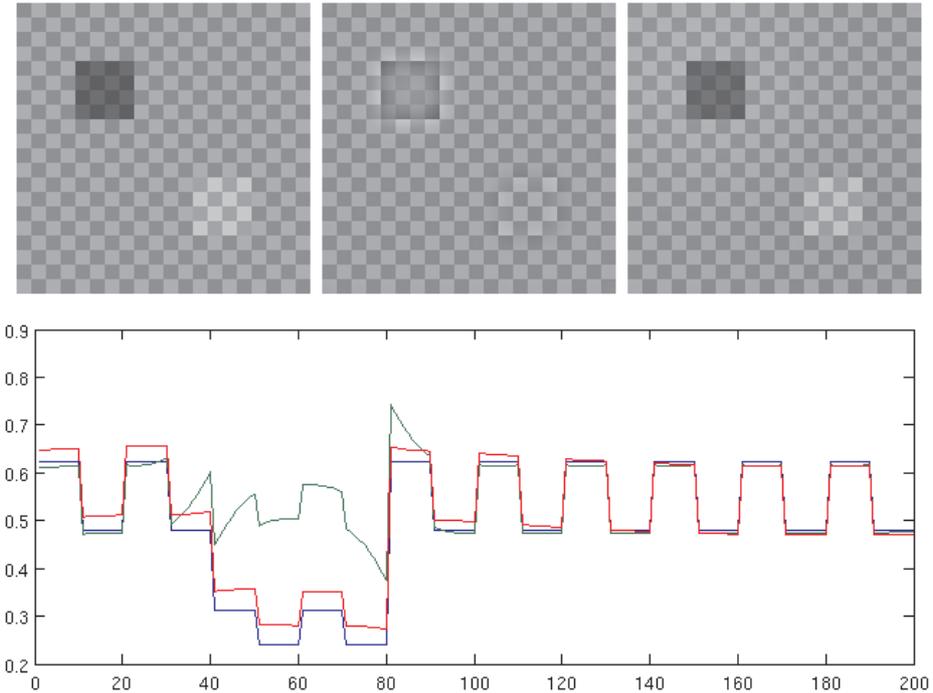


Fig. 12. Top: An image with perfectly even illumination is shown on the left, while the middle and rightmost images show the results obtained after processing it with the considered algorithm for different settings of the parameters affecting preservation of medium-sized objects. Bottom: The images from left to right are represented by blue, green and red curves, respectively, which show a cut through the black object. The left and right images become very similar, as indicated by the blue and red curves.

no preservation of medium-sized objects. That image is depicted in Fig. 9, on the right in the middle row. The overall illumination is improved, which can be seen by comparing the blue line with the red line (inserted for the sake of comparison). The bottom graph shows the result of applying the new method (the output image can be seen in Fig. 9, on the right in the bottom row). We can see there how the river Arno, represented by the low flat area to the right, is now much darker compared to the original method. The green line has been inserted to make the comparison easier, and the dark river now cuts through the green line, while it was lifted up by the original method. We can also note that the darker field at the top of the input image has become a bit darker after applying the proposed method, just as it should (see the leftmost part of the curve).

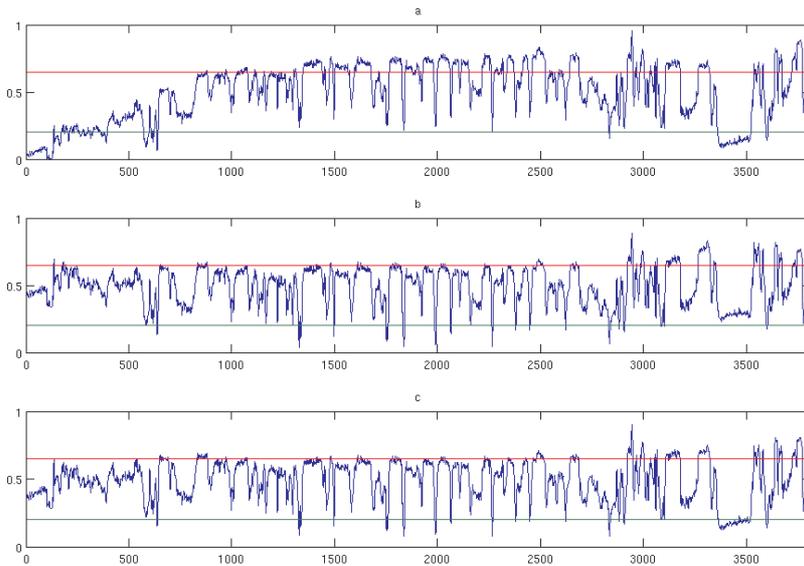


Fig. 13. Top: a cut section of the original image (see Fig. 9 on the right). Middle: the same cut section after applying the original method. Bottom: The proposed method has been applied, and the river (flat area to the right) is darker, as it should be, i.e. it is not lifted up above the inserted green reference line.

The difference between the middle and bottom curves might seem very small, but the difference in appearance is really noticeable in Fig. 9. We can also note that the sides of the river are not lifted up in the bottom curve as much compared to the red line as is the case in the middle curve. Hence, the annoying ripple effect is removed by the new method. This ripple effect is more visible in the image on the left in Fig. 9, where it is noticeable as a darker edge around the whiter field.

Finally, we have used one parameter to avoid contrast enhancement in already high contrast areas, as described at the end of section 3.1. Hence there are a number of parameters to be adjusted by the user. However, they can be set to standard default values, as explained, making it easy for the user to then adjust them in order to obtain an optimal image.

The proposed method is relatively fast in matlab (2.4 s for a scale factor of 10 and 3 iterations for the leftmost image in Fig. 9). However, a GPU implementation would make it possible to adjust parameters more or less in real time even for such large images as we have used here, making it suitable also for commercial applications.

6. Conclusions

The proposed algorithm uses luminance mapping for illumination correction by computing the local mean and deviation. These are obtained by Gaussian filtering, which is a computationally costly process. Therefore, we have proposed to carry out these operations on downsampled versions of the input image. The resulting blurred images are subsequently upsampled to full size with little loss of accuracy. Large objects in the images, having a different intensity than their surroundings, are removed by band pass filtering in order to avoid their washing out, which is what illumination correction generally does. They can be preserved by removing them before correction and adding afterwards. By iterating this process more than once, including removing, correction and adding, a better result is obtained as the band pass filtered image tends to contain some low variations. Several examples shown in the paper prove that the method can be used for both gray scale and colour images with good results.

References

- 1968**
- [1] Oppenheim, A.; Schafer, R.; Stockham, T., Jr.: Nonlinear filtering of multiplied and convolved signals. *Audio and Electroacoustics, IEEE Transactions on*, Volume: 16 Issue:3, pp. 437-466.
- 1983**
- [2] Burt, P., J. and Adelson, E., H.: A multiresolution spline with application to image mosaics. *Journal ACM Transactions on Graphics (TOG)*, Volume 2 Issue 4.
- [3] Fahnestock, J. D. and Schowengerdt, R.A.: Spatially variant contrast enhancement using local range modification, *Optical Engineering* 22, pp. 378-381.
- 1987**
- [4] Pizer, S. M., Amburn, E. P., Austin, J. D., Cromartie, R., Geselowitz, A., Greer, T., Romeny, B. T. H., Zimmerman, J. B. and Zuiderveld, K.: Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*. Volume 39, Issue 3, September, pp. 355-368.
- 1994**
- [5] Puff, D. T., Pisano, E. D., Muller, K. E., Johnston, R. E., Hemminger, B. M., Burbeck, C. A., McLelland, R. and Pizer, S. M.: A Method for Determination of Optimal Image Enhancement for the Detection of Mammographic Abnormalities. *Journal of Digital Imaging*, Vol 7, No 4 (November), pp 161-171.
- [6] Vrhel, M. J. and Trussell, H. J.: Filter considerations in color correction. *IEEE Trans. Image Processing*, vol. 3, pp. 147-161, Mar.
- 1995**
- [7] Laine, A.; Jian Fan; Wuhai Yang: Wavelets for contrast enhancement of digital mammography. *Engineering in Medicine and Biology Magazine, IEEE*, Volume: 14 Issue:5. pp. 536 - 550.
- 1997**
- [8] Reeves, T.H., Jernigan, M.E.: Multiscale-based image enhancement. *Electrical and Computer Engineering. IEEE Canadian Conference on*. Issue Date: 25-28 Vol 2. pp. 500-503.
- 2000**
- [9] Young, I. T. Shading Correction: Compensation for Illumination and Sensor Inhomogeneities, *Current Protocols in Cytometry*. pp 1-12.
- 2001**
- [10] Hertzmann, A., Jacobs, C. E., Oliver, N., Curless, B, and Salesin, D. H.: Image Analogies. *SIG-GRAPH 2001 Conference Proceedings*. pp. 327-340.

2002

- [11] Ko, J., Kim, E. and Byun, H.: A simple illumination algorithm for face recognition. Proceeding of PRICAI, pp. 532-541.
- [12] Sidney F. R.: Applied photographic optics, 3rd ed., Focal Press.
- [13] Yoon, J. H., Ro, Y. M.: Enhancement of the Contrast in Mammographic Images, using the Homomorphic Filter Method. INF. & SYST., VOL.E85-D, NO.1 JANUARY LETTER. pp. 298-303.

2003

- [14] Leong F. J. W-M., Brady, M., O'D McGee J.: Correction of uneven illumination (vignetting) in digital microscopy images J Clin Pathol. pp. 619 - 621.
- [15] Pérez, P., Gangnet, M. and Blake, A.: Poisson Image Editing, Journal ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH, Volume 22 Issue 3, pp. 313-318
- [16] Tzelepis, N. and Nakos, B.: A Study on the Lighting Factors affecting Relief Presentation, Proceedings of the 21st International Cartographic Conference (ICC), pp. 1343-1350.
- [17] Zhu, J., Liu, B. and Schwartz, S. C. General illumination correction and its application to face normalization. Proceeding of ICASSP. pp 133-136.

2004

- [18] Levin, A., Zomet, A., Peleg, S., Weiss, Y.: Seamless Image Stitching in the Gradient Domain. In ECCV (4) pp. 377-389.
- [19] Yu. W. Practical anti-vignetting methods for digital cameras. IEEE Trans. on Cons. Elect., 50. Pp. 975-983.

2005

- [20] Goldman, D. B. and Chen, J. H.: Vignette and Exposure Calibration and Compensation. Proceedings of ICCV '05, pp. 89-906.
- [21] Pajares, G., Ruz, J. J. and de la Cruz, J. J.: Performance Analysis of Homomorphic Systems for Image Change Detection. Pattern Recognition and Image Analysis, LNCS, Volume 3522/2005, pp. 563-570.
- [22] Rocchini, D. and Di Rita, A.: Relief effects on aerial photos geometric correction. Applied Geography 25. pp. 159-168.
- [23] Yucong Guo, Xingming Zhang, Huangyuan Zhan, and Jing Song, A Novel Illumination Normalization Method for Face Recognition. Advances in Biometric Person Authentication, International Workshop on Biometric Recognition Systems, IWBRIS 2005, Beijing, China, Volume 3781, pp. 23-30.

2006

- [24] Zheng, Y., Lin, S., Kang. S. B. : Single-Image Vignetting Correction. Computer Vision and Pattern Recognition, IEEE Computer Society Conference. pp. 461-468.

2007

- [25] Agus, O.; Ozkan, M.; Aydin, K.: Elimination of RF Inhomogeneity Effects in Segmentation, Proceedings of the 29th Annual International Conference of the IEEE EMBS, pp. 2081-2084.
- [26] Ardizzone, E., Pirrone, R., La Bua, S., Gambino, O.: Volumetric Bias Correction. Computer Vision/Computer Graphics Collaboration Techniques: MIRAGE 2007. LNCS, vol. 4418, pp. 525-533.

2008

- [27] Ashiba, H. I., Awadalla, K. H., El-Halfawy S. M., and Abd El-Samie F. E.-S.: Homomorphic enhancement of infrared images using the additive wavelet transform. Progress In Electromagnetics Research C, Vol. 1, pp. 123-130.
- [28] Sonka, M., Hlavac, V. and Boyle, R.: Image Processing, Analysis, and Machine Vision. Thomson Learning, USA, pp.38.

2010

- [29] ERDAS Field Guide, pp. 525 (2010)
http://www.erdas.com/Libraries/Tech_Docs/ERDAS_Field_Guide.sflb.ashx

2011

- [30] Hast, A.; Marchetti, A.: Retrospective Illumination Correction of Greyscale Historical Aerial Photos. International Conference on Image Analysis and Processing, Poster. pp. 275-284.
- [31] Russ, J.C.: The Image Processing Handbook, Sixth Edition, CRC Press. pp. 370.

A METHOD OF CONSTRUCTING PHYLLOTAXICALLY ARRANGED MODULAR MODELS BY PARTITIONING THE INTERIOR OF A CYLINDER OR A CONE

Cezary Stepień

Institute of Computer Science, Warsaw University of Technology, Poland
cst@ii.pw.edu.pl

Abstract. The paper describes a method of partitioning a cylinder space into three-dimensional subspaces, congruent to each other, as well as partitioning a cone space into subspaces similar to each other. The way of partitioning is of such a nature that the intersection of any two subspaces is the empty set. Subspaces are arranged with regard to phyllotaxis. Phyllotaxis lets us distinguish privileged directions and observe parastichies trending these directions. The subspaces are created by sweeping a changing cross-section along a given path, which enables us to obtain not only simple shapes but also complicated ones. Having created these subspaces, we can put modules inside them, which do not need to be obligatorily congruent or similar. The method ensures that any module does not intersect another one. An example of plant model is given, consisting of modules phyllotaxically arranged inside a cylinder or a cone.

Key words: computer graphics, modeling, modular model, phyllotaxis, cylinder partitioning, cone partitioning, genetic helix, parastichy.

1. Introduction

Phyllotaxis is the manner of how leaves are arranged on a plant stem. The regularity of leaves arrangement, known for a long time, still absorbs the attention of researchers in the fields of botany, mathematics and computer graphics. Various methods have been used to describe phyllotaxis. A historical review of problems referring to phyllotaxis is given in [7]. Its connections with number sequences, e.g. Fibonacci sequence, and with problems of symmetry, the golden ratio and logarithmic spiral have been discussed in a number of papers. A few theories of biological processes and biomechanical phenomena resulting in appropriate phyllotactic patterns were put forward ([9–14]). Douady and Couder [2, 5] showed that phyllotactic patterns similar to those observed in botany can also emerge as a result of activity of other physical processes. This approach is continued in [16], where a magnetic cactus is presented. The influence of basic parameters on a kind of phyllotactic pattern of both theoretical models and real plants is described in [4, 15]. The paper [8] contains a uniform description of spiral, jugate and whorl patterns on the basis of Helmholtz equation. Papers [1, 3] present an application of phyllotactic patterns to plant modeling in computer graphics. The website [17] shows that phyllotaxis can

be an inspiration for architects. The aim of this paper is to propose a new method of constructing objects being a set of modules arranged inside a cylinder or a cone in such a manner that they satisfy the rules of phyllotaxis. The modules are analogs of leaves and for this reason they cannot intersect. In order to obtain the appropriate result, a conception of subspaces is introduced. The subspaces fill to capacity the space of a model. Each subspace can contain exactly one module. The subspaces are obliged to satisfy appropriate conditions in reference to other subspaces while modules being analogs of leaves should be in appropriate relations, each to its own subspace.

In Chapter 2, fundamental conceptions referring to phyllotaxis are introduced, such as genetic helix or parastichy. In Chapter 3, the flat model of two- and three-directional phyllotaxis is introduced. Chapter 4 describes phyllotaxis on cylindrical and conical surfaces. Chapter 5 describes partitioning the interior of a cylinder and a cone. In Chapter 6, a method of arranging modules using subspaces is presented and Chapter 7 contains a conclusion.

2. Genetic helix and parastichies

To describe phyllotaxis, an arrangement of modules on a cylindrical surface of the radius R is often studied. The cylinder axis lies on the z axis of the coordinate system. As a result of the growing process, the module $i + 1$ emerges on the surface of the cylinder in such a manner that it is rotated by a fixed angle $\Delta\alpha$ and moved by a fixed distance Δz along z axis, both in reference to the module i .

Let $i = 0, \pm 1, \pm 2, \dots$. Knowing the origin coordinates of the module i , i.e. its angle α_i and its axis position z_i we can calculate the position of the next point from the equations: $\alpha_{i+1} = \alpha_i + \Delta\alpha$, $z_{i+1} = z_i + \Delta z$, so that we obtain a sequence of points $Q = \{\mathbf{q}_i\}$ lying on a helix. We call this helix a *genetic helix* and $\Delta\alpha$ —an *angle of divergence* [7].

Let a natural number N be given. Having the sequence Q , create a subsequence Q_j^N , ($j = 0, 1, \dots, N - 1$), choosing elements $\mathbf{q}_j, \mathbf{q}_{j\pm N}, \mathbf{q}_{j\pm 2N}, \dots$. Connect these points using a line being a helix. We call this helix *parastichy of the order N* and mark them with a symbol P_j^N .

If two numbers N_1 and N_2 ($N_1 \neq N_2$) are given instead of the number N , then we can study two sets of parastichies $\{P_j^{N_1}\}$ and $\{P_k^{N_2}\}$, where $j = 0, 1, \dots, N_1 - 1$ and $k = 0, 1, \dots, N_2 - 1$. Parastichies belonging to the same set do not intersect in opposition to parastichies belonging to different sets, as the latter ones intersect in certain points of Q .

Figure 1(a) represents a genetic helix, Fig. 1(b) and 1(c) show two examples of sets $\{P_j^{N_1}\}$ and $\{P_j^{N_2}\}$ and Fig. 1(d) shows an example of three sets of parastichies.

One can cut open the cylindrical surface and study the genetic helix in a two-dimensional space $S^{<\alpha\zeta>}$, i.e. in a space with the coordinate system $O\alpha\zeta$, where $\alpha \in \langle -\pi, \pi \rangle$ and $\zeta \in \langle -\infty, +\infty \rangle$. In this case α is the angle and ζ stands for z . On

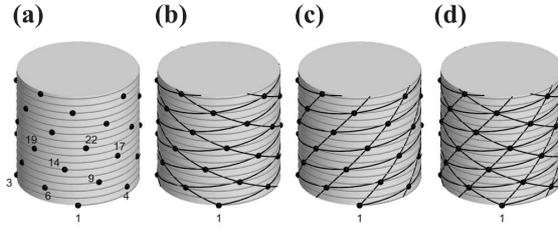


Fig. 1. Genetic helix (a) and parastichies belonging to sets: (b) $\{P_j^3\}, \{P_k^5\}$, (c) $\{P_j^3\}, \{P_k^8\}$, and (d) $\{P_j^3\}, \{P_k^5\}, \{P_l^8\}$.

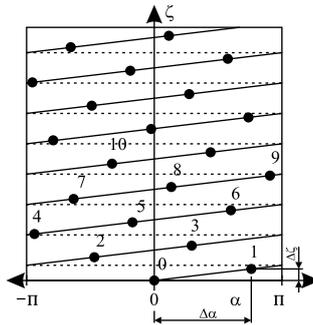


Fig. 2. The genetic helix on the cut open and unrolled cylinder surface.

a conical surface, the dependence of z upon ζ is slightly more complicated, which is described in Chapter 4.

In the case of two-dimensional space $S^{<\alpha\zeta>}$, the genetic helix and parastichies change into sets of line sections moved along the z axis by a constant value [1].

3. Flat model of phyllotaxis

In order to define a genetic helix in the space $S^{<\alpha\zeta>}$, we need two parameters $\Delta\alpha$ and $\Delta\zeta$. We calculate coordinates of points \mathbf{q}_i using equations:

$$\alpha_i = \Delta\alpha i - 2\pi k,$$

$$\zeta_i = \Delta\zeta i,$$

where $k: \alpha_i \in \langle -\pi, \pi \rangle$ and $i = 0, \pm 1, \pm 2, \dots$

The genetic helix for $\zeta \geq 0$ is shown in Figure 2. Notice that the genetic helix passes through the point \mathbf{q}_0 of coordinates $(\alpha, \zeta) = (0, 0)$.

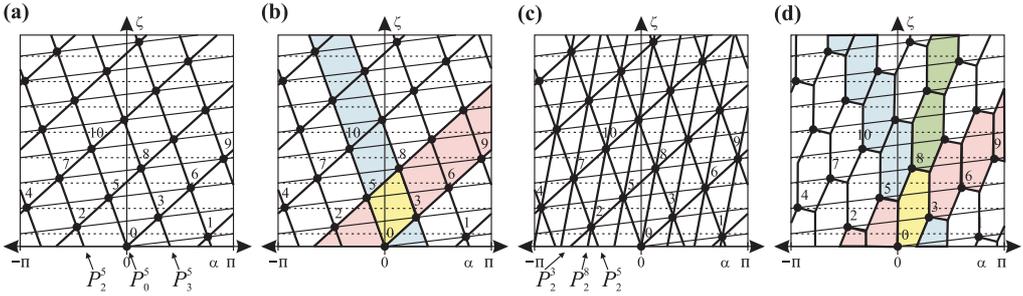


Fig. 3. The space with parastichies belonging to sets: (a) $\{P_j^3\}, \{P_k^5\}$ and (c) $\{P_j^3\}, \{P_k^5\}, \{P_l^8\}$ and partitioning the space into parallelograms and hexagons.

Figure 3(a) shows parastichies belonging to sets $\{P_j^{N_1}\}$ and $\{P_k^{N_2}\}$ where $N_1 = 3$ and $N_2 = 5$. For each parastichy of the order $N \geq 3$, it is possible to show two parastichies belonging to the same set and being in neighborhood with it. In this figure, we see that the parastichy P_0^5 is a neighbor of P_2^3 and P_3^5 . In accordance with two directions constituted by two sets of parastichies one can divide the space $S^{<\alpha\zeta>}$ into parallelograms of identical shapes and orientations filling all the space (Figure 3(b)). The parallelograms, due to their contact by their sides, create strips in accordance with the parastichies directions. We call such partitioning a *two-directional* one.

Analogically, we can consider three sets of parastichies $\{P_j^{N_1}\}, \{P_k^{N_2}\}$ and $\{P_l^{N_1+N_2}\}$. Such case was studied in [1]. The set $\{P_l^{N_1+N_2}\}$ states the third direction, as it is shown in Figure 3(c) for the sets $\{P_j^3\}, \{P_k^5\}$ and $\{P_l^8\}$. Then we can divide the space into hexagons of identical shapes and orientations completely filling it out. In Figure 3(d) we can observe strips parallel to these three directions and so we call such partitioning a *three-directional* one.

Instead of $\{P_l^{N_1+N_2}\}$ we can take into account the set $\{P_l^{N_1-N_2}\}$ ($N_1 - N_2 \geq 1$) [1], but that case can be easily reduced to the case described above.

Below, we study two- and three directional partitionings into areas of identical shapes and orientations, but we do not demand their sides to be sections of straight lines.

Firstly, we will consider a three directional partitioning. Let there be given the point \mathbf{q}_i lying in the place, where parastichies $P_j^{N_1}, P_k^{N_2}$ and $P_l^{N_1+N_2}$ intersect. We also consider points lying on neighboring parastichies, namely $\mathbf{q}_{i+N_1}, \mathbf{q}_{i+N_2}$, and $\mathbf{q}_{i+N_1+N_2}$.

Assume that the point \mathbf{q}_i is associated with two points $\mathbf{v}_i^{<1>}, \mathbf{v}_i^{<2>}$ and with a line L_i consisting of three open line segments $L_i^{<1>}, L_i^{<2>}$ and $L_i^{<3>}$, as it is shown in Figure 4(a). These line segments can be fragments of straight lines, curves or polylines. We assume below that for each two points \mathbf{q}_{i_1} and \mathbf{q}_{i_2} these lines are congruent in couples:

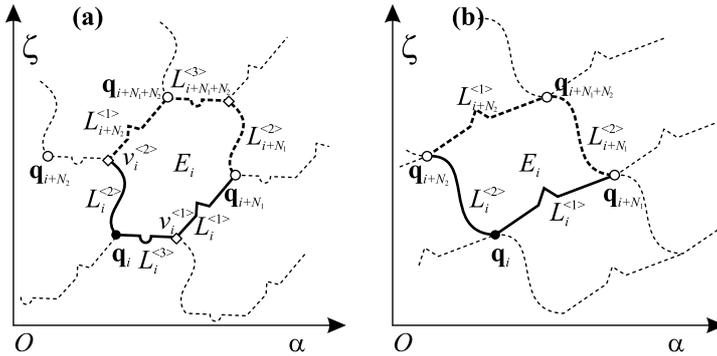


Fig. 4. Example of (a) three-directional partitioning and (b) two directional one.

$$L_{i_1}^{<1>} \equiv L_{i_2}^{<1>}, \quad L_{i_1}^{<2>} \equiv L_{i_2}^{<2>}, \quad L_{i_1}^{<3>} \equiv L_{i_2}^{<3>}. \quad (1)$$

From this assumption it appears that $\mathbf{v}_i^{<1>} = \mathbf{v}_{i-N_2}^{<2>}$ and $\mathbf{v}_i^{<2>} = \mathbf{v}_{i+N_2}^{<1>}$.

The segments $L_i^{<1>}$, $L_i^{<2>}$ and $L_i^{<3>}$ together with $L_{i+N_2}^{<1>}$, $L_{i+N_1}^{<2>}$ and $L_{i+N_1+N_2}^{<3>}$ limit a certain area. We denote the interior of this area with \underline{E}_i . Using \underline{E}_i , we can define a set E_i , which will be helpful in the next part of this paper:

$$E_i = \underline{E}_i \cup L_i^{<1>} \cup L_i^{<2>} \cup L_i^{<3>} \cup \{\mathbf{v}_i^{<1>}\} \cup \{\mathbf{q}_i\}.$$

The areas constructed in this manner are congruent, which results from Eq. (1) (of course, if one takes into account periodicity of the surface in relation to α). Moreover, none of these areas overlap each other and all of them fill the space $S^{<\alpha>}$ completely.

One can understand a two directional partitioning as a particular case of a three directional one, where the length of the segment $L_i^{<3>}$ has been reduced to null. Then $\mathbf{v}_i^{<1>} = \mathbf{q}_i$, $\mathbf{v}_i^{<2>} = \mathbf{q}_{i+N_2}$ etc., so E_i is enclosed by segments $L_i^{<1>}$, $L_i^{<2>}$ and $L_{i+N_2}^{<1>}$, $L_{i+N_1}^{<2>}$.

Reducing the length of $L_i^{<3>}$ to null causes the direction stated by parastichies from the set $P_i^{N_1+N_2}$ to become more difficult to observation, as particular areas contact with each other not along a line but only in one point. For example, areas E_i and $E_{i+N_1+N_2}$ are in contact not along $L_{i+N_1+N_2}^{<3>}$ but only in the point $\mathbf{q}_{i+N_1+N_2}$ (Fig. 4(b)).

4. Phyllotaxis on a cylindrical surface and on a conical one

Below, we describe a cylindrical surface as a space in the coordinate system $Oxyz$ and mark it by $S_{\text{cyl}}^{<xyz>}$. Similarly we mark a conical surface in the system $Oxyz$ by $S_{\text{cone}}^{<xyz>}$.

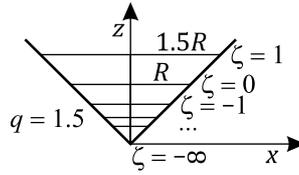


Fig. 5. Side view of a cone; there are shown cross-sections for various values of ζ .

If a point $\mathbf{q}(\alpha, \zeta) \in S^{<\alpha\zeta>}$ is given, then we can calculate the coordinates of its equivalent in the space $S^{<xyz>}_{cyl}$ using the following formulas:

$$x = R \cos \alpha, \quad y = R \sin \alpha, \quad z = Z\zeta, \tag{2}$$

where R and Z are predefined.

Analogically, we can calculate coordinates of $\mathbf{q}(\alpha, \zeta)$ in the space $S^{<xyz>}_{cone}$ using the formulas:

$$x = Rq^\zeta \cos \alpha, \quad y = Rq^\zeta \sin \alpha, \quad z = Zq^\zeta, \tag{3}$$

where R , Z and q are predefined.

After applying Eq. (2), the space $S^{<\alpha\zeta>}$ transforms into a cylindrical surface having the following properties:

- (1) For $\zeta \in (-\infty, +\infty)$ $z \in (-\infty, +\infty)$ and if $\zeta = 0$, then $z = 0$.
- (2) A cross-section of the cylinder surface is a circle of the radius R .

On the other hand, after applying Eq. (3), the space $S^{<\alpha\zeta>}$ transforms into a cone surface. Their properties are as follows:

- (1) For $\zeta \in (-\infty, +\infty)$ $z \in (0, +\infty)$ and if $\zeta \rightarrow -\infty$, then $z \rightarrow 0$.
- (2) If $\zeta = 0$ we obtain a cross-section being a circle of the radius R located on the plane $z = 1$.
- (3) If $\zeta = 1$ we obtain a cross-section being a circle of the radius Rq located on the plane $z = q$ (Fig. 5).

Let us study two areas in the space $S^{<\alpha\zeta>}$ described in Chapter 3. We denote them by E_{i_1} and E_{i_2} . We know that $E_{i_1} \equiv E_{i_2}$ (where \equiv is a symbol of congruency) and that they have an identical orientation. It means that there exists a translation described by two constants c_α, c_ζ such that if a point $\mathbf{p}^{<1>}(\alpha_1, \zeta_1) \in E_{i_1}$ exists, then there also exists the point $\mathbf{p}^{<2>}(\alpha_2, \zeta_2) \in E_{i_2}$ described by the equation:

$$\mathbf{p}^{<2>}(\alpha_2, \zeta_2) = \mathbf{p}^{<1>}(\alpha_1 + c_\alpha, \zeta_1 + c_\zeta).$$

After transforming $\mathbf{p}^{<1>}$ and $\mathbf{p}^{<2>}$ into the space $S^{<xyz>}_{cyl}$ by Eq. (2), we obtain $\mathbf{p}^{<1>}_{cyl} = (R \cos \alpha, R \sin \alpha, Z\zeta)$ and $\mathbf{p}^{<2>}_{cyl} = (R \cos(\alpha + c_\alpha), R \sin(\alpha + c_\alpha), Z(\zeta + c_\zeta))$. This means that $\mathbf{p}^{<1>}_{cyl} \equiv \mathbf{p}^{<2>}_{cyl}$ (are congruent), which in this case is a superposition of translation

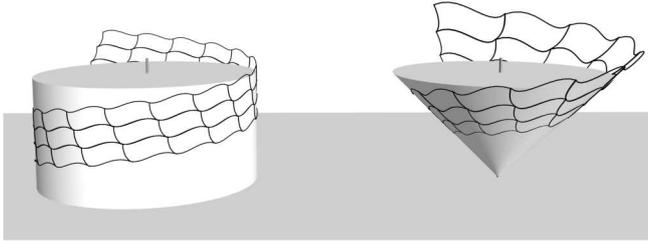


Fig. 6. Covering a cylindrical surface with congruent sets and a conical surface with similar ones; the sets of each kind completely cover their respective spaces.

along the z axis by c_ζ and rotation by an angle c_α . It is evident that the mentioned sets are also congruent in the space S_{cyl}^{xyz} , which we write down as $E_{i_1} \equiv E_{i_2}$.

After transforming these same points $\mathbf{p}^{<1>}$ and $\mathbf{p}^{<2>}$ from $S^{<\alpha\zeta>}$ into S_{cone}^{xyz} by Eq. 3), we obtain $\mathbf{p}_{\text{cone}}^{<1>} = (Rq^{\zeta_1} \cos \alpha_1, Rq^{\zeta_1} \sin \alpha_1, q^{\zeta_1})$ and $\mathbf{p}_{\text{cone}}^{<2>} = (Rq^{\zeta_1+c_\zeta} \cos(\alpha_1 + c_\alpha), Rq^{\zeta_1+c_\zeta} \sin(\alpha_1 + c_\alpha), q^{\zeta_1+c_\zeta})$. So we see that $\mathbf{p}_{\text{cone}}^{<1>} \sim \mathbf{p}_{\text{cone}}^{<2>}$, where \sim stands for similarity, which in this case is a result of superposition of scaling with the scale factor q^{c_ζ} and rotation by an angle c_α . It appears that, for $S_{\text{cone}}^{<xyz>}$, $E_{i_1} \sim E_{i_2}$. Using Eq. (3) results in complete covering the cone surface $S_{\text{cone}}^{<xyz>}$ with similar areas (Fig. 6).

5. Partitioning the interior of a cylinder and a cone

Let us introduce a three-dimensional space $S^{<\alpha\zeta\rho>}$, which is understood as a generalization of $S^{<\alpha\zeta>}$. This allows us to consider $S^{<\alpha\zeta>}$ as a cross-section of $S^{<\alpha\zeta\rho>}$ for a given ρ , namely $S^{<\alpha\zeta>}(\rho)$. Similarly, as it was described in Chapter 3, we divide the space $S^{<\alpha\zeta>}(\rho)$ into sets $E_i(\rho)$ using the genetic helix (passing through the point $\mathbf{q}_0(\rho)$). As $\mathbf{q}_0(\rho)$ is dependent on ρ , it means that appropriate parastichies and segments of lines being borders of the areas $E_i(\rho)$ are dependent on ρ too. We express that by using the notation $L_{i+N_2}^{<1>}(\rho)$, $L_{i+N_1}^{<2>}(\rho)$ and $L_{i+N_1+N_2}^{<3>}(\rho)$.

The dependence of $\mathbf{q}_0(\rho)$ on ρ implies that the set $\{\mathbf{q}_0(\rho)\}$ is a segment of a certain line, which we call a *path*. For a given $\{\mathbf{q}_0(\rho)\}$, segments of lines describing borders of $E_0(\rho)$ are denoted by $L_{N_2}^{<1>}(\rho)$, $L_{N_1}^{<2>}(\rho)$ and $L_{N_1+N_2}^{<3>}(\rho)$, because $i = 0$.

Taking into account how the points $\mathbf{q}_i(\rho)$ depend on $\mathbf{q}_0(\rho)$ we notice that for each $i_1, i_2 \in \{0, \pm 1, \pm 2, \dots\}$ and for a fixed ρ the paths $\{\mathbf{q}_{i_1}(\rho)\}$ and $\{\mathbf{q}_{i_2}(\rho)\}$ are congruent, which we show by using the notation $\{\mathbf{q}_{i_1}(\rho)\} \equiv \{\mathbf{q}_{i_2}(\rho)\}$.

Consider the set of points $V^{<\alpha\zeta\rho>} \in S^{<\alpha\zeta\rho>}$, defined as follows: $\alpha \in \langle -\pi, \pi \rangle$, $\zeta \in \langle -\infty, +\infty \rangle$, $\rho \in \langle r_{\min}, r_{\max} \rangle$ (Fig. 7(a)). Let us divide $V^{<\alpha\zeta\rho>}$ into subsets $V_i^{<\alpha\zeta\rho>}$,

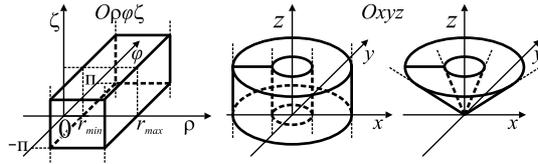


Fig. 7. From left: sets $V_i^{<\alpha\zeta\rho>}$, $V_{cyl}^{<xyz>}$ and $V_{cone}^{<xyz>}$.

which are further referred to as subspaces. Each subspace is defined as follows:

$$V_i^{<\alpha\zeta\rho>} = \bigcup_{\rho \in (r_{min}, r_{max})} E_i(\rho) \tag{4}$$

This equation shows that $E_i(\rho)$ is a cross-section of $V_i^{<\alpha\zeta\rho>}$ for a given ρ . Taking into account the congruency of $E_i(\rho)$, we see that subspaces $V_i^{<\alpha\zeta\rho>}$ are congruent too, completely fill $S^{<\alpha\zeta\rho>}$ and do not intersect each other. In fact, in order to create $V_i^{<\alpha\zeta\rho>}$, we simply sweep a variable cross-section $E_i(\rho)$ along a path $\{\mathbf{q}_i(\rho)\}$. This manner of description of three-dimensional solids is known in computer graphics as *sweep representation* [6].

Below, we consider a transformation of points $\mathbf{q}(\alpha, \zeta, \rho) \in V_i^{<\alpha\zeta\rho>}$ into the space $S_{cyl}^{<xyz>}$ or $S_{cone}^{<xyz>}$, respectively, in accordance with the following equations: for a cylinder

$$x = \rho \cos \alpha, \quad y = \rho \sin \alpha, \quad z = Z\zeta \tag{5}$$

and for a cone

$$x = \rho q^\zeta \cos \alpha, \quad y = \rho q^\zeta \sin \alpha, \quad z = Zq^\zeta \tag{6}$$

In the space $S_{cyl}^{<xyz>}$, the set $V^{<\alpha\zeta\rho>}$ obtains a shape of a drilled cylinder, which is marked by the symbol $V_{cyl}^{<xyz>}$ (Fig. 7(b)), while in the space $S_{cone}^{<xyz>}$, it obtains a shape of drilled cone which is marked by $V_{cone}^{<xyz>}$ (Fig. 7(c)). Subspaces being results of transformations $V_i^{<\alpha\zeta\rho>}$ are marked by $V_{cyl,i}^{<xyz>}$ and $V_{cone,i}^{<xyz>}$.

Figure 8 represents a space $V_i^{<\alpha\zeta\rho>}$ in the case when $L_i^{<1>}(\rho)$, $L_i^{<2>}(\rho)$ and the path $\{\mathbf{q}_0(\rho)\}$ are segments of straight lines. Moreover, two corresponding subspaces: $V_{cyl,i}^{<xyz>}$ and $V_{cone,i}^{<xyz>}$ are shown. Figure 9 explains how the subspaces $V_{cyl,i}^{<xyz>}$ and $V_{cone,i}^{<xyz>}$ can fill a cylindrical space and a conical one for both two- and three-directional partitioning.

Below, there are more complicated examples of two-directional partitioning. One can easily imagine analogous examples for a three-directional case comparing these examples with Fig. 6. Figure 10(a) illustrates a subspace $V_i^{<\alpha\zeta\rho>}$, which has its path being a segment of straight line, while $L_i^{<1>}(\rho)$, $L_i^{<2>}(\rho)$ are segments of curves changing

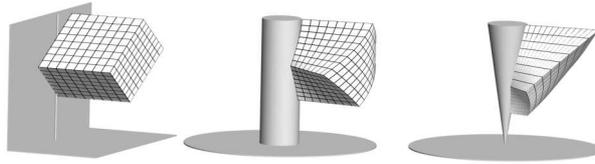


Fig. 8. From left: an example of a subspace $V_i^{<\alpha\zeta\rho>}$ and two corresponding subspaces $V_{cyl,i}^{<xyz>}$ and $V_{cone,i}^{<xyz>}$.

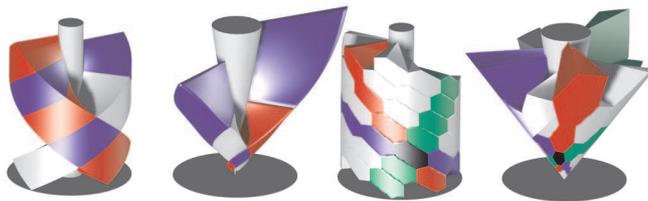


Fig. 9. Examples of a cylinder space and a cone one filled by subspaces $V_{cyl,i}^{<xyz>}$ and $V_{cone,i}^{<xyz>}$ for both the two- and the three-directional partitioning.

their shapes depending on ρ . Fig. 10(b) shows a subspace $V_{cyl,i}^{<xyz>}$ being a result of transformation for the input subspace from Fig. 10(a). Figure 10(c) shows a cylinder subspace constructed by these same lines $L_i^{<1>(\rho)}$, $L_i^{<2>(\rho)}$ and the path being a curve lying on a plane $\alpha = const$. In Fig. 10(d) there is a similar example but the path is a curve lying on a plane $\zeta = const$.

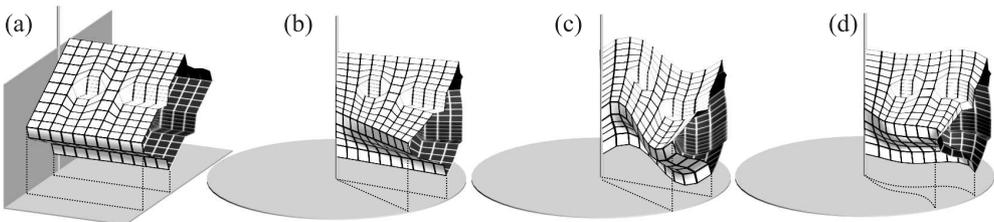


Fig. 10. Examples: (a) a subspace $V_i^{<\alpha\zeta\rho>}$ and various subspaces $V_{cyl,i}^{<xyz>}$ (b, c, d); the front faces have been removed in order to reveal the interiors.



Fig. 11. Filling the cylinder interior by subspaces $V_{\text{cyl},i}^{\langle xyz \rangle}$; left: three models where modules originate respectively from Fig. 10(b), (c) and (d), right: a three-directional model.

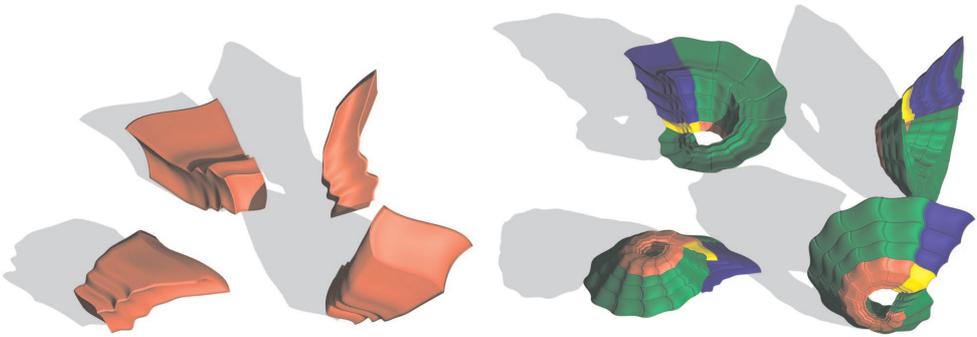


Fig. 12. An example of a subspace $V_{\text{cone},i}^{\langle xyz \rangle}$ (left) and the result of filling a cone interior (right).

Figure 11 (left) shows a manner of filling the cylinder space by subspaces $V_{\text{cyl},i}^{\langle xyz \rangle}$ from Fig. 10(b), (c) and (d). Subspaces lying on a certain chosen parastichy $P_j^{N_1}$ and on a certain $P_k^{N_2}$ have been distinguished respectively by colors: red and blue, whereas a subspace located on the intersection of both these parastichies—by yellow. On the right side of this figure a model with three-directional filling is shown.

Figure 12 shows an example of a certain subspace $V_{\text{cone},i}^{\langle xyz \rangle}$ and a result of filling a cone interior with subspaces similar to $V_{\text{cone},i}^{\langle xyz \rangle}$.

6. Arranging non-intersecting modules inside a cylinder or a cone

When natural objects having their modules arranged in accordance with the phyllotaxis rules are modeled, it is often essential for them not to intersect each other. In order to meet this requirement we can use the idea of subspaces $V_i^{<\alpha\zeta\rho>}$, $V_{cyl,i}^{<xyz>}$ and $V_{cone,i}^{<xyz>}$, ($i = 0, \pm 1, \pm 2, \dots$) discussed in Chapter 5.

Let us associate a certain set $M_i^{<\alpha\zeta\rho>}$, called *module*, with each subspace $V_i^{<\alpha\zeta\rho>}$. Assume that the following relation is true:

$$M_i^{<\alpha\zeta\rho>} \subseteq V_i^{<\alpha\zeta\rho>} \tag{7}$$

This relation remains true also after transforming a module $M_i^{<\alpha\zeta\rho>}$ into the cylindrical space or into the conical one. Thus for a cylinder

$$M_{cyl,i}^{<xyz>} \subseteq V_{cyl,i}^{<xyz>} \tag{8}$$

and for a cone

$$M_{cone,i}^{<xyz>} \subseteq V_{cone,i}^{<xyz>} \tag{9}$$

In the case of a cylinder, if we assume that all the modules are congruent, then we can work out the shape of a module in the space $S^{<\alpha\zeta\rho>}$. Next, after doing the test described by Eq. (7), we should copy this module and put the duplicates in the appropriate places of this space. In the end, the whole group of modules is transformed into the cylinder space.

Alternatively we can work out the shape of a module in the space $S_{cyl}^{<xyz>}$, execute the test in accordance with Eq. (8) and, in the end, copy the module to put its duplicates in places being occupied by the subspaces $V_{cyl,i}^{<xyz>}$. We should remember that it is not necessary to restrict our activity to the case when modules are congruent to each other, in contrast to the case of subspaces. It is only essential for each module to satisfy Eq. (8). This ensures that the area of applications of the described method is wider.

In the case of cones, variants of the procedure are analogous but requirements relating to congruency of the modules in the space $S_{cyl}^{<xyz>}$ should be replaced with requirements relating to the similarity in $S_{cone}^{<xyz>}$. In addition, the test described by Eq. (9) should be applied instead of the test in Eq. (8).

Figure 13 shows objects resembling leaves arranged inside a cylinder. The corresponding subspaces are shown too. Figure 14(a) represents a cylindrical model, consisting of 160 modules with shapes that vary depending on their positions along the z axis. Here, the subspaces from Fig. 13(d) have been used. Fig. 14(b) represents a conical model consisting of the same modules. Figure 15 shows a model of inflorescence consisting of two parts: the upper one—conical and the lower one—cylindrical. Each part contains

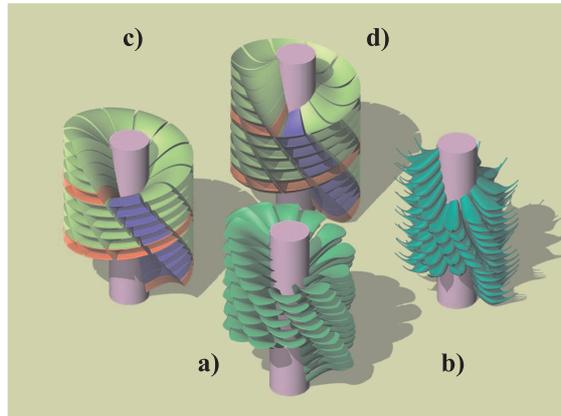


Fig. 13. Two models, (a) and (b), consisting of non-intersecting modules, each resembling a leaf on a stem, and two sets of subspaces corresponding with these models, shown in (c) and (d), respectively.

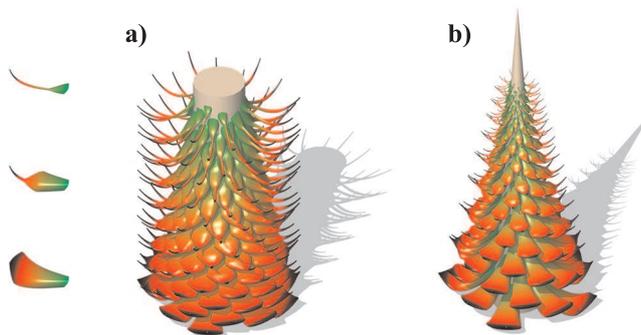


Fig. 14. A cylindrical model (a) consisted of modules having their shapes dependent on the module position along the model axis, and a conical one (b) consisted of the same modules.

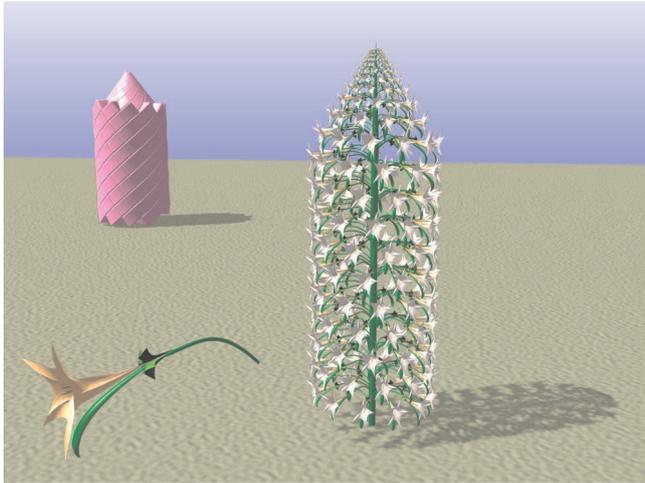


Fig. 15. On the right, an inflorescence consisting of two parts: the conical upper part and the cylindrical lower one; on the left, a module is shown in the foreground while in the background we can see partitioning a cone and a cylinder into subspaces appropriate for this model.

120 non-intersecting modules arranged in concordance with the rules described herein. Each module consists of a flower, a leaf and a stalk.

To sum up, in order to construct a model, namely the set $\{M_{\text{cyl},i}^{\langle xyz \rangle}\}$ or $\{M_{\text{cone},i}^{\langle xyz \rangle}\}$ one should perform the following steps:

Begin

- 1) Determine the phyllotaxis parameters, namely the parameters of the genetic helix and parameters of parastichies in dependence on a chosen partitioning: two- or three-directional;
- 2) Determine a path $\mathbf{q}_0(\rho)$ in the space $S^{\langle \alpha \zeta \rho \rangle}$;
- 3) Determine lines $L_0^{\langle 1 \rangle}(\rho)$, $L_0^{\langle 2 \rangle}(\rho)$, and for the three-directional partitioning also $L_0^{\langle 3 \rangle}(\rho)$;
- 4) Construct an area $E_0(\rho)$ and next, basing on this area and on the path, construct a subspace $V_0^{\langle \alpha \zeta \rho \rangle}$ using the sweep representation method; choose Variant A, B or C;

Variant A:

- 5(a) Construct the module $M_0^{\langle \alpha \zeta \rho \rangle}$ in the manner which satisfies relation $M_0^{\langle \alpha \zeta \rho \rangle} \subseteq V_0^{\langle \alpha \zeta \rho \rangle}$;

- 6(a) Copy $M_0^{<\alpha\zeta\rho>}$ and arrange duplicates along the axes α and ζ in accordance with the phyllotaxis parameters, which ensure that each subspace $V_i^{<\alpha\zeta\rho>}$ containing its own module $M_i^{<\alpha\zeta\rho>}$ precisely adjoin its neighboring subspaces;
- 7(a) Using Eq. (5) or Eq. (6), transform the obtained set of the modules $\{M_i^{<\alpha\zeta\rho>}\}$ into the cylinder space $S_{\text{cyl}}^{<xyz>}$ or into the cone one $S_{\text{cone}}^{<xyz>}$; Go to L1; (End of Variant A);

Variant B—for a cylinder:

- 5(b) Using Eq. (5), transform $V_0^{<\alpha\zeta\rho>}$ into the cylinder space $S_{\text{cyl}}^{<xyz>}$ to obtain $V_{\text{cyl},0}^{<xyz>}$;
- 6(b) Construct the module $M_{\text{cyl},0}^{<xyz>}$ in the manner which satisfies the relation $M_{\text{cyl},0}^{<xyz>} \subseteq V_{\text{cyl},0}^{<xyz>}$;
- 7(b) Copy $M_{\text{cyl},0}^{<xyz>}$, next rotate and move in accordance with the phyllotaxis parameters, in order to obtain the set $\{M_{\text{cyl},i}^{<xyz>}\}$; Go to L1; (End of Variant B);

Variant C—for a cone:

- 5(c) Using Eq. (6), transform $V_0^{<\alpha\zeta\rho>}$ into the cone space $S_{\text{cone}}^{<xyz>}$ to obtain $V_{\text{cone},0}^{<xyz>}$;
- 6(c) Construct the module $M_{\text{cone},0}^{<xyz>}$ in the manner which satisfies the relation $M_{\text{cone},0}^{<xyz>} \subseteq V_{\text{cone},0}^{<xyz>}$;
- 7(c) Copy $M_{\text{cone},0}^{<xyz>}$, next rotate, scale and move in accordance with the phyllotaxis parameters, in order to obtain the set $\{M_{\text{cone},i}^{<xyz>}\}$; (End of Variant C);

L1: End

7. Conclusion

The method described in this paper makes it possible to construct modules and arrange them in accordance with phyllotaxis rules in such a way that they do not intersect each other. One can create the modules using the conception of a subspace in the space $V^{<\alpha\zeta\rho>}$ or alternatively in the respective spaces $V_{\text{cyl}}^{<xyz>}$ or $V_{\text{cone}}^{<xyz>}$. A diversity of shapes is a result of creation subspaces using the so called *sweep representation* method by moving a cross-section along a given path. The curvilinear paths are allowed as well as such cross-sections that change their shapes along the path. The appropriate requirements related to the cross-section shape in the space $V^{<\alpha\zeta\rho>}$ ensure the congruency of subspaces inside a cylinder as well as similarity inside a cone and at the same time the absence of mutual intersections of subspaces. The advantage of the described method consists in the following: if there is given a certain partitioning into subspaces then one

can place inside these subspaces modules differing from each other in shapes, as it can be observed in nature. The method enables us to construct complicated models composed of certain parts being cylinders and at the same time of other parts being cones (Fig. 15). This approach can be useful in practical applications. The author hopes that the method described herein can be applicable to modeling not only modular plants but also other objects of a cylindrical or conical shape, consisting of regularly arranged modules.

References

1990

- [1] Prusinkiewicz P., Lindenmayer A.: *The Algorithmic Beauty of Plants*. Springer Verlag.

1992

- [2] Douady S., Couder Y.: Phyllotaxis as a Physical Self-Organized Growth Process. *Physical Review Letters*. Vol. 68.
 [3] Fowler D.R., Prusinkiewicz P., Battjes J.: A collision-based model of spiral phyllotaxis. *Proc. of the 19th Annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH'92*.

1994

- [4] Zagórska-Marek B.: Phyllotaxic diversity in Magnolia flowers, *Acta Soc. Bot. Poloniae* 63:117-137.

1996

- [5] Douady S., Couder Y.: Phyllotaxis as a Dynamical Self Organizing Process. *Journal of Theoretical Biology* 178:255-274.
 [6] Foley I. et al.: *Computer Graphics: Principles and Practice*. Addison-Wesley.

1997

- [7] Adler I., Barabe D., Jean R.V.: A History of the Study of Phyllotaxis. *Annals of Botany* 80:231-244.

1998

- [8] Cummings F.W., Strickland J.C.: A Model of Phyllotaxis. *Journal of Theoretical Biology* 192:531-544.

2000

- [9] Hargittai I., Pickover C.A.: *Spiral symmetry*. World Scientific Publishing.

2004

- [10] Kappraff J.: Growth in Plants: A Study in Number. *Forma* 19:335-354.

2005

- [11] Shipman P.D., Newell A.C.: Polygonal planforms and phyllotaxis on plants. *Journal of Theoretical Biology* 236:154-197.

2006

- [12] Smith R.S., Kuhlemeier C., Prusinkiewicz P.: Inhibition fields for phyllotactic pattern formation: a simulation study. *Canadian Journal of Botany* 84(11):1635-1649.
 [13] Ciszak Ł., Stepień C.: A dynamic model of phyllotaxis for application in computer graphics. *Proceedings of the XII National Conference Application of Mathematics to Biology and Medicine*, Koninki, Poland, pp. 31-36.

2008

- [14] Newell A.C., Shipman P.D., Sun Z.: Phyllotaxis: Cooperation and competition between mechanical and biochemical processes. *Journal of Theoretical Biology* 251:421-439.
- [15] Zagórska-Marek B., Szpak M.: Virtual phyllotaxis and real plant model cases. *Functional Plant Biology* 35:1025-1033.

2009

- [16] Nisoli C., Gabor N.M., Lammert P.E., Maynard J.D., Crespi V.H.: Static and Dynamical Phyllotaxis in Magnetic Cactus. *Physical Review Letters* 102.

2014

- [17] Swiss Re H, 30 St Mary Axe. <http://www.fosterandpartners.com/Projects/1004/Default.aspx> (retrieved Feb. 1, 2014).

ROBUST METHOD FOR CAMERA SELF-CALIBRATION BY AN UNKNOWN PLANAR SCENE

A. Baataoui¹, N. El Akkad¹, A. Saaidi^{1,2}, K. Satori¹, Med. Masrar¹

¹*LIAN, Department of Mathematics and Computer Science, Faculty of Sciences,
Dhar El Mehraz, Sidi Mohamed Ben Abdellah University, B.P 1796, Atlas, Fez, Morocco.*

²*LIMAO, Department of Mathematics, Physics and Computer Science,
Polydisciplinary Faculty of Taza, Sidi Mohamed Ben Abdellah University, B.P 1223, Taza, Morocco.
baataoui.aziz@gmail.com, nabil_abdo80@yahoo.fr, saaidi.abde@yahoo.fr,
khalidsatorim3i@yahoo.fr, masrar.m@hotmail.com.*

Abstract. In this paper, we present a method of self-calibration of a CCD camera with varying intrinsic parameters by an unknown planar scene. The advantage of our method is to reduce the number of images (two images) to estimate the parameters of the camera used. Moreover, the self-calibration equations become related to the number of matched points (very numerous and easy to detect) and not to the number of images, because the use of a large number of the images requires a high execution time. On the other hand, we base on the matched points which are numerous to estimate the projection matrices and the homographies between images. The latter are used with the images of the absolute conic to formulate a system of non-linear equations (self-calibration equations depend on the number of matched couples). Finally, the intrinsic parameters of the camera can be obtained by minimizing a non-linear cost function proceeding from two steps: initialization and optimization. The experiment results show the robustness of our algorithms in terms of stability and convergence.

Keywords: self-calibration, equilateral triangle, absolute conic, homography, varying intrinsic parameters.

1. Introduction

The camera is the main element in many applications of computer vision. The estimation of camera parameters is an important step in this kind of applications. Generally, the estimation procedure can be performed according to two strategies. The first is called the calibration: it consists to determine the intrinsic and extrinsic parameters using a known object (calibration pattern) [12, 14, 15, 25]. The latter can be three-dimensional (3D calibration) or plane (2D calibration). The second strategy is called self-calibration. It allows determining the intrinsic and extrinsic parameters without any prior knowledge of the scene. Several works are based on the self-calibration of cameras from 3D scenes [4–8, 13, 24, 31], or they are based on planar scenes [11, 18, 26, 28] to automatically determine the intrinsic and extrinsic parameters.

In this work, we are interested on the self-calibration of a camera with varying intrinsic parameters by any planar scene. We mention that from two points of the 2D scene

we can obtain an equilateral triangle (the third point of the triangle can be obtained in a unique way). Our method is based on the rotation of a fixed reference associated with the planar scene to determine the transformation matrix between the vertices of the different equilateral triangles. This transformation characterizes the strong point of our approach. The latter resides in the use of a large number of matches which are the projections of the points (two vertices of the equilateral triangles) of the planar scene in the couples of images. This projection is used with the homography matrices to formulate a system of linear equations. The resolution of the latter allows obtaining the projection matrices. After detecting the interest points by the Harris algorithm [3] and the matching of these points by the correlation measure ZNCC [16, 22], the homography between the two images is estimated from four matches by using RANSAC algorithm [2]. The relationships between the projection matrices, homographies and the images of the absolute conic proved a non-linear cost function. The minimization of this function by the Levenberg-Marquardt [1] allows obtaining the intrinsic parameters of the used camera.

In addition, the importance of this work resides on the one hand in the use of fewer images (two images) instead of using more images [10] to estimate the cameras parameters. On the other hand, it resides in the formulation of self-calibration equations from the matches (numerous and easy to detect) and not of the number of the images (requires a high processing time). The self-calibration steps of our approach are presented in Figure 1.

The paper is organized as follows: Section 2 presents a survey of related works. The camera model used in this work and the image of the absolute conic is treated in Section 3. The vision system is described in Section 4. The tools for self-calibration are presented in Section 5. The self-calibration equations are elaborated in Section 6. Experiments are presented in Section 7, and the conclusions in Section 8.

2. Survey of Related Works

In literature we find several methods which treat the problem of self-calibration. Two categories of these methods can be distinguished: *i*) the methods that use cameras characterized by constant parameters. *ii*) those which use cameras characterized by varying parameters.

i) With the assumption of constant parameters we find several works. The first major work of self-calibration was treated in [4], the authors have proposed an algorithm based on two steps: In the first step, they found the epipolar transformation by the method of Sturm (this method is based on projective invariants) and the other method is based on the generalization of the essential matrix. In the second step of the computation, they used the Kruppa equations [13] which link the epipolar transformation to the image of the absolute conic. Subsequently, in [7] the authors treated the self-calibration of camera

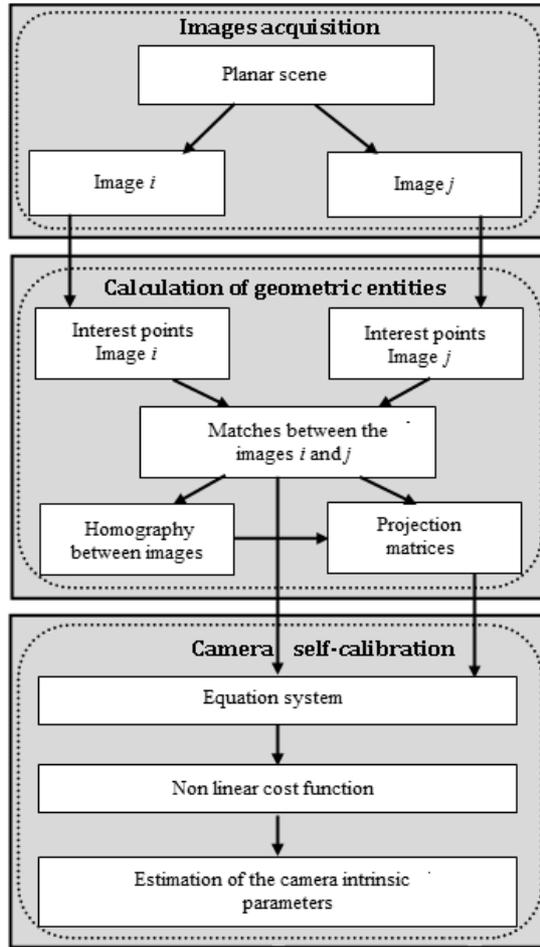


Fig. 1. Steps of camera self-calibration

by using the absolute dual quadric to recover the Euclidean structure. With the same assumption, in [11] the authors have proposed an algorithm based on the projection of two circular points of the planar scene in each image plan (five images at least), together with the estimation of the homography between each couple of images to determine the camera parameters. On the other hand, in [8] the authors have incorporated the so-called module in the stratification approach to upgrade projective structures to affine and finally recover the absolute conic and improve structures Euclidean constraint. Further work

with the same assumption, using 2D or 3D scenes assuming contain specific objects (parallelogram, circle, triangle...) that allow to exploit some geometric constraints to estimate camera parameters. For example, in [26], the authors have proposed a method of self-calibration plane based on the use of a parallelogram. They have used the matched points to estimate all the projection matrices of this parallelogram. These matrices are operated with homographies between images to estimate the intrinsic parameters.

ii) In recent years, the researchers have proposed new methods of camera self-calibration with varying parameters. They have supposed assumptions about the scene (2D or 3D), camera movement (circular, pure rotation), and intrinsic parameters (zero skew and known aspect ratio) for estimating the camera parameters. A new method of self-calibration of camera characterized by the varying intrinsic parameters is treated in [29]. It is based on the quasi-affine reconstruction. After this reconstruction, the authors have estimated the homography of the plane at infinity and they have used it with some constraints on the images of the absolute conic to determine the intrinsic the cameras parameters. On the other hand, a robust method of self-calibration of the cameras characterized by varying parameters is treated in [31]. The last method is based on the projection of three points of the scene on the plans of images and the relation between the matchings to formulate a non-linear cost function. The resolution of the latter allows obtaining the intrinsic parameters of the cameras used. In [23] a method of self-calibration of a camera with varying parameters is based on a circular movement of the camera. The homography of the plane at infinity is determined from two constraints: the first considers that rotation angle between two views of the camera is known, and the second considers that the pixels are square. After obtaining the homography, the intrinsic camera parameters are easily determined. In [17], the authors have considered a self-calibration problem of a moving camera whose intrinsic parameters are known, except the focal length which may vary freely across different views. Furthermore, the focal length's values depend only on the camera's motions. The authors gave a complete catalog of critical motion sequences, which is used to determine these sequences from stereo systems with variable focal. With the assumption of varying parameters, in [27] the authors have presented a practical algorithm for self-calibrating of a camera with varying intrinsic parameters. For each view, the authors suggest minimizing a non-linear least square to establish the matrix of intrinsic parameters. The minimization procedure begins by an initialization to give a first estimate of the focal distance; therefore the estimation of the intrinsic parameters is performed by an algorithm with several iterations. In each iteration, one parameter is estimated by assuming some constraints on the other parameters. A recent method [30] is based on relative distances to estimate the camera parameters. The latter are obtained from the resolution of a non-linear equation system which is formulated by using the invariant relative distance and the homography that transforms the projective reconstruction to metric reconstruction.

The method presented in this work is a development of the work treated in [28] and almost similar to the work treated in [26] and [31]. In [26] and [28], the authors considered that the camera used has constant intrinsic parameters and requires at least three images to calibrate the camera. Moreover, the authors assumed the constraints $\tau = 0$ and $\varepsilon = 0$. On the contrary, the present method uses any camera (characterized by varying intrinsic parameters), in addition, two images are sufficient to calibrate the camera and no constraint on the intrinsic parameters of the camera used. Furthermore, the only difference between this method and the one treated in [31] resides in the objects used in planar scenes: the method presented in [31] is based on the projection of a parallelogram on the plans of images and the relationship between the matches. On the contrary, the present method is based only on the relations between the vertices of the triangles used in the planar scene.

3. Camera Model and the Image of The Absolute Conic

Our approach is based on the pinhole model of the camera to transform a point of the planar scene to its projection in the image.

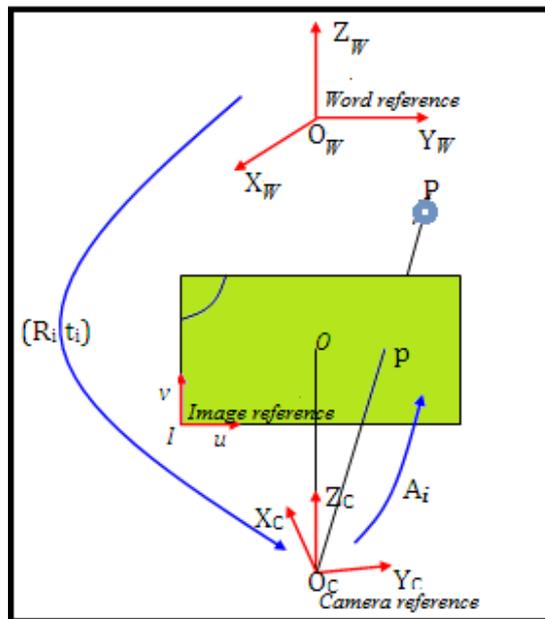


Fig. 2. Pinhole model of camera

The projection of each point P of the scene in image i can be described by a 3×4 matrix: L_i which is expressed by the following formula:

$$p \sim L_i P \quad (1)$$

The matrix L_i can be written as follows: $L_i = A_i(R_i t_i)$; with:

- $(R_i t_i)$ represents the matrix of extrinsic parameters, where R_i the rotation matrix and t_i is the translation vector in the space.
- A_i is the matrix of intrinsic parameters, which is:

$$A_i = \begin{pmatrix} f_i & \tau_i & u_{0i} \\ 0 & \varepsilon_i f_i & v_{0i} \\ 0 & 0 & 1 \end{pmatrix} \quad (2)$$

With, f_i represents the focal length for the view i , $i = 1$ or 2 (in our case), ε_i is the aspect ratio, (u_{0i}, v_{0i}) are the coordinates of the principal point in the image i and τ_i is the image skew.

The Image of the Absolute Conic (IAC), denoted by ω_i , is an imaginary point conic directly related to the camera internal matrix A_i in formula (2) via $\omega_i = (A_i A_i^T)^{-1}$:

$$\omega_i = \frac{1}{\varepsilon_i^2 f_i^4} \begin{pmatrix} \varepsilon_i^2 f_i^2 & -\tau_i \varepsilon_i f_i & -u_{0i} \varepsilon_i^2 f_i^2 + v_{0i} \tau_i \varepsilon_i f_i \\ * & f_i^2 + \tau_i^2 & -v_{0i} f_i^2 + u_{0i} \tau_i \varepsilon_i f_i - v_{0i} \tau_i^2 \\ * & * & \varepsilon_i^2 f_i^4 + v_{0i}^2 f_i^2 + (u_{0i} \varepsilon_i f_i - \tau_i v_{0i})^2 \end{pmatrix} \quad (3)$$

Where the three lower triangular elements are replaced by “*” to save space, since ω_i is symmetric. The estimation of intrinsic parameters of the camera used automatically gives the elements of the matrix ω_i and vice versa.

4. Vision System

In this work, we consider an unknown planar scene. On the plan of the scene, we consider n points P_r , with $r = 1 \dots n$, or $n \in N^*$ and O is a point different from the points P_r and it is in the same plane which contains the points P_r . For each segment $[OP_r]$, there exists a unique point M_r in the scene plan such that $OP_r M_r$ is an equilateral triangle having an angle $P_r \hat{O} M_r > 0$. We associate to each triangle $OP_r M_r$ a reference (O, X_r, Y_r, Z_r) with $P_r \in (OX_r)$ and (OZ_r) is perpendicular to the plan containing the triangle $(OP_r M_r)$ (Figure 3).

We denote by (O, X_1, Y_1) the fixed reference in the scene plan. Let (O, X_r, Y_r) with $r = 2 \dots n$ denote moving references according to (O, X_1, Y_1) . These references are associated with equilateral triangles $OP_r M_r$ with $r = 2 \dots n$. They are obtained by a simple rotation of the fixed reference around (OZ_1) axis. In addition to that, the passage

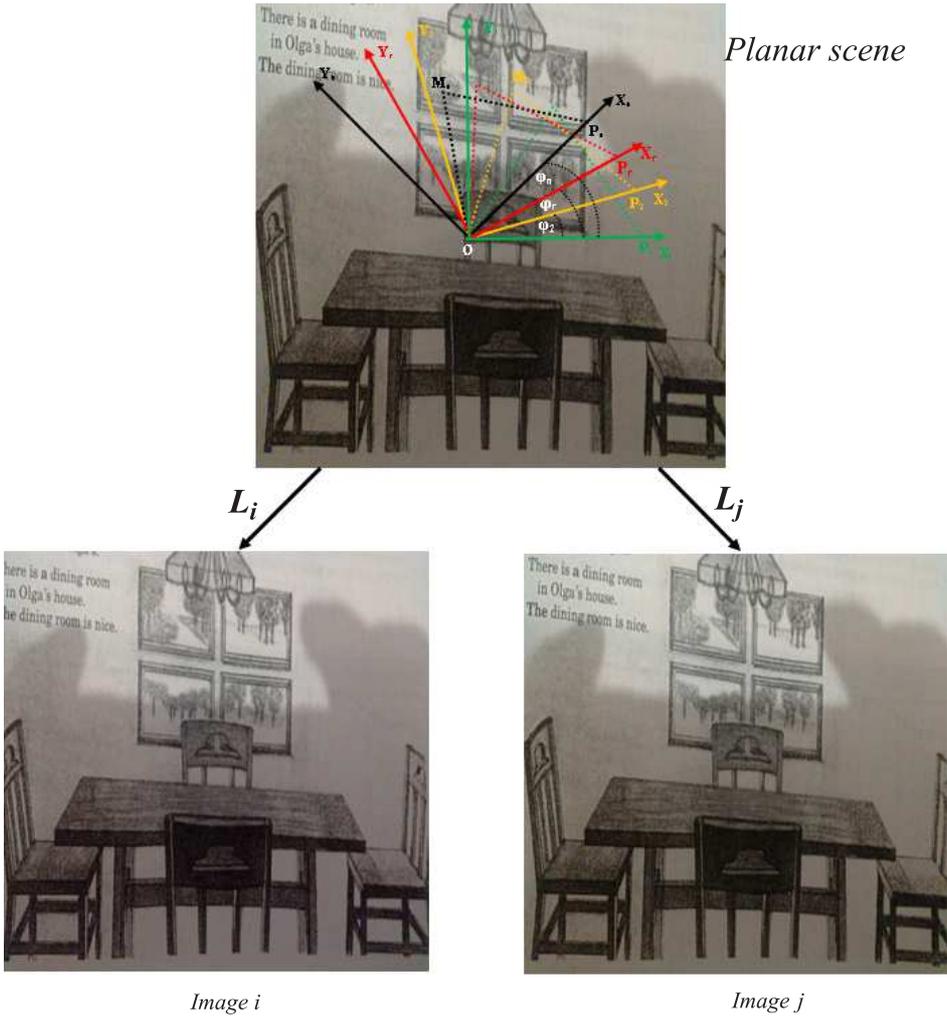


Fig. 3. Vision system

from fixed reference (O, X_1, Y_1) to the moving reference (O, X_r, Y_r) is performed by using the rotation matrix which is given as follows:

$$\mathfrak{R}(\varphi_r) = \begin{pmatrix} \cos(\varphi_r) & -\sin(\varphi_r) & 0 \\ \sin(\varphi_r) & \cos(\varphi_r) & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{with } r = 2 \dots n \quad (4)$$

With φ_r the rotation angle which allows obtaining the moving reference (Figure 4).

Figure 4 shows the system used: the planar scene, the fixed reference, the moving references and the equilateral triangles.

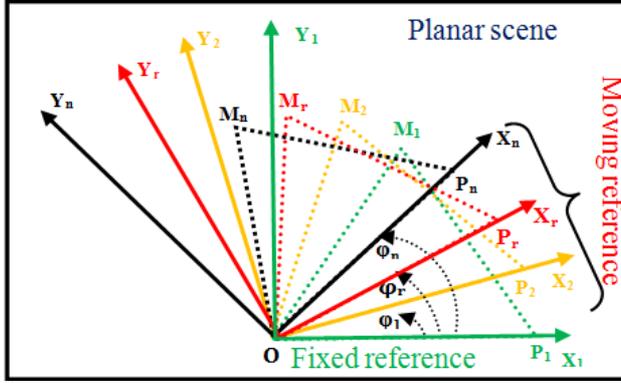


Fig. 4. System used

The homogeneous coordinates of points P_r , M_r and O in the moving reference (O, X_r, Y_r) are respectively $(a_r, 0, 1)^T$, $(\frac{a_r}{2}, \frac{\sqrt{3}}{2}a_r, 1)^T$ and $(0, 0, 1)^T$. With a_r ($a_r = OP_r$) represents the length of the equilateral triangle OP_rM_r . These coordinates can be rewritten as follows:

$$(a_r, 0, 1)^T = S_r(1, 0, 1)^T \quad (5)$$

$$\left(\frac{a_r}{2}, \frac{\sqrt{3}}{2}a_r, 1\right)^T = S_r(0, 1, 1)^T \quad (6)$$

With:

$$S_r = \begin{pmatrix} a_r & \frac{a_r}{2} & 0 \\ 0 & \frac{\sqrt{3}}{2}a_r & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (7)$$

The coordinates corresponding to points P_r and M_r in the fixed reference can be calculated as follows: $\mathfrak{R}(\varphi_r)P_r$ and $\mathfrak{R}(\varphi_r)M_r$.

5. Self-calibration Tools

The self-calibration procedure used in our approach is the following: detecting the interest points by Harris algorithm, setting a matching of the interest points by the correlation measure ZNCC, calculating the homography between images using the RANSAC

algorithm, determining the projection matrix of the scene by the resolution of a linear system, and estimating the intrinsic parameters of the camera used by minimizing of a non-linear cost function.

5.1. Matching and interest points

The matching of the image points can be established in two steps: The first step is to extract the interest points of the two images i and j . In the literature, there are several algorithms to extract interest points [3, 9, 20, 21], this article uses the Harris algorithm [3]. The second step is to find for each interest point of the image their correspondent in the image by measuring correlation ZNCC [16, 22], and then we eliminate the false matches by using the RANSAC algorithm [2].

5.2. Homography between images

The homography is a 3×3 transformation of matrix linking the matches points between the images i and j , it is expressed as follows:

$$p_{jr} \sim H_{ij}p_{ir} \quad (8)$$

Where p_{ir} and p_{jr} are respectively the projection of a point P_r of the scene in the images i and j . With $r = 1 \dots n$, H_{ij} is the homography matrix between the images i and j . The homography matrix is calculated by the RANSAC algorithm [2]. The latter allows estimating the geometric entity (homography) from four matches between the images i and j .

5.3. Projection matrices of the segments $[OP_r]$

The objective of this section is to estimate the projection matrices L_{ir} and L_{jr} for each segment $[OP_r]$ of the scene, in the two images i and j (Figure 5). Knowing that the degree of freedom of these matrices is eight, therefore, we need at least eight equations to calculate these two matrices which will be used in the self-calibration equations.

The projection of the different points expressed in the fixed reference of the scene in the images i and j is given by the formula (1). The coordinates of the points P_r in the fixed reference are given by $\mathfrak{R}(\varphi_r)P_r$, and they are given by $S_r(1, 0, 1)^T$ in the moving reference (O, X_r, Y_r) .

The projection of the points O, P_1, P_2, \dots, P_n in the images i and j is performed by the following formulas:

$$Image\ i \Rightarrow \begin{cases} (u_{i0}, v_{i0}, 1)^T & \sim L_{ir}(0, 0, 1)^T \\ (u_{ip_r}, v_{ip_r}, 1)^T & \sim L_{ir}(1, 0, 1)^T \end{cases} \quad (9)$$

$$Image\ j \Rightarrow \begin{cases} (u_{j0}, v_{j0}, 1)^T & \sim L_{jr}(0, 0, 1)^T \\ (u_{jp_r}, v_{jp_r}, 1)^T & \sim L_{jr}(1, 0, 1)^T \end{cases} \quad (10)$$

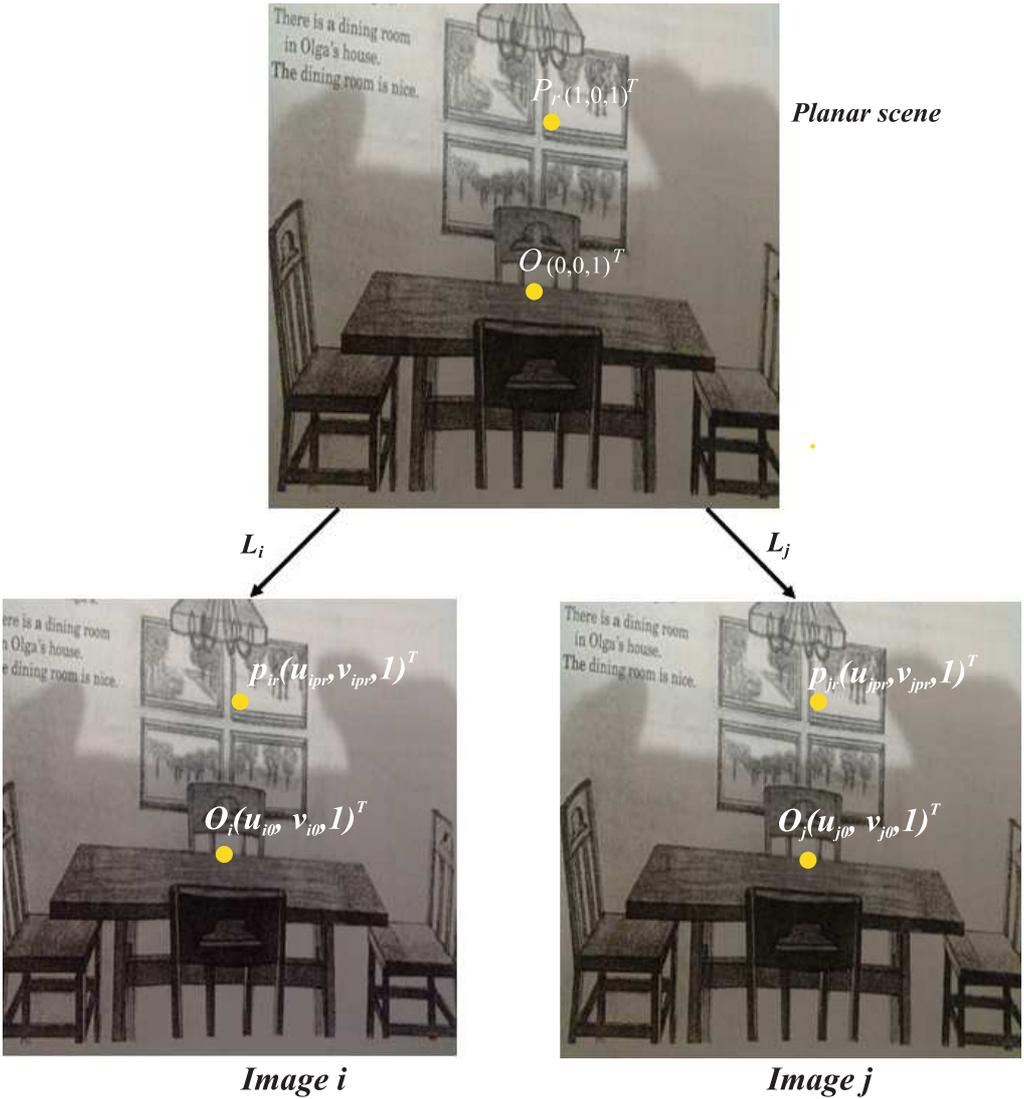


Fig. 5. Projection of points of the scene in the images i and j

With the points $(u_{i0}, v_{i0}, 1)^T$ and $(u_{iP_r}, v_{iP_r}, 1)^T$ are respectively the projection of the points O and P_r in the image i , the points $(u_{j0}, v_{j0}, 1)^T$ and $(u_{jP_r}, v_{jP_r}, 1)^T$ are respectively the projection of the points O and P_r in the image j and L_{i_r} , L_{i_r} represent the

projection matrices of the points (O, P_r) in the images i and j respectively, they are expressed as follows:

$$L_{ir} = A_i R_i \begin{pmatrix} 1 & 0 \\ 0 & 1 & R_i^T t_i \\ 0 & 0 \end{pmatrix} \mathfrak{R}(\varphi_r) S_r \tag{11}$$

$$L_{jr} = A_j R_j \begin{pmatrix} 1 & 0 \\ 0 & 1 & R_j^T t_j \\ 0 & 0 \end{pmatrix} \mathfrak{R}(\varphi_r) S_r \tag{12}$$

With $\mathfrak{R}(\varphi_r)$ and S_r are given respectively by the formulas (4) and (7).

Let S_r^* denotes the matrix is defined as follows: $S_r^* = \mathfrak{R}(\varphi_r) S_r$. Using the formulas (4) and (7), S_r^* can be expressed as follows:

$$S_r^* = \begin{pmatrix} a_r \cos(\varphi_r) & \frac{a_r}{2} \cos(\varphi_r) - \frac{\sqrt{3}}{2} a_r \sin(\varphi_r) & 0 \\ a_r \sin(\varphi_r) & \frac{a_r}{2} \sin(\varphi_r) + \frac{\sqrt{3}}{2} a_r \cos(\varphi_r) & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{13}$$

The matrices $H_i = A_i R_i \begin{pmatrix} 1 & 0 \\ 0 & 1 & R_i^T t_i \\ 0 & 0 \end{pmatrix}$ and $H_j = A_j R_j \begin{pmatrix} 1 & 0 \\ 0 & 1 & R_j^T t_j \\ 0 & 0 \end{pmatrix}$ are respectively the homographies which permit to project the plane of the scene, in the images i and j , therefore the formulas (11) and (12) become:

$$L_{ir} = H_i S_r^* \tag{14}$$

$$L_{jr} = H_j S_r^* \tag{15}$$

From Equations (14) and (15) we deduce that:

$$L_{jr} = H_{ij} L_{ir} \tag{16}$$

With H_{ij} is the homography between the images i and j such that:

$$H_{ij} = H_j H_i^{-1} \tag{17}$$

From the formulas (11), (12) and (13) we can deduce that:

$$L_{ir} = A_i R_i (G_r R_i^T t_i) \tag{18}$$

$$L_{jr} = A_j R_j (G_r R_j^T t_j) \tag{19}$$

With:

$$G_r = \begin{pmatrix} a_r \cos(\varphi_r) & \frac{a_r}{2} \cos(\varphi_r) - \frac{\sqrt{3}}{2} a_r \sin(\varphi_r) \\ a_r \sin(\varphi_r) & \frac{a_r}{2} \sin(\varphi_r) + \frac{\sqrt{3}}{2} a_r \cos(\varphi_r) \\ 0 & 0 \end{pmatrix} \quad (20)$$

The formula (9) gives four equations according to the L_{ir} elements.

The formulas (10) and (16) give:

$$\begin{cases} (u_{j0}, v_{j0}, 1)^T \sim H_{ij} L_{ir} (0, 0, 1)^T \\ (u_{jp_r}, v_{jp_r}, 1)^T \sim H_{ij} L_{ir} (1, 0, 1)^T \end{cases} \quad (21)$$

The system (21) gives four other equations according to the L_{ir} elements;

Therefore, we can obtain the eight L_{ir} elements from expressions (9) and (21).

The L_{jr} projection matrix is estimated from the formula (16).

6. Camera Self-Calibration

The expression (18) gives:

$$A_i^{-1} L_{ir} = R_i (G_r R_i^T t_i) \quad (22)$$

The formula (22) can be written as follows:

$$L_{ir}^T \omega_i L_{ir} = \begin{pmatrix} G_r^T G_r & G_r R_i^T t_i \\ t_i^T R_i G_r & t_i^T t_i \end{pmatrix} \quad (23)$$

With $\omega_i = (A_i A_i^T)^{-1}$ is the projection of the absolute conic ($\Omega \sim I_3$) in the image i .

Proceeding in the same way, we can show that:

$$L_{jr}^T \omega_j L_{jr} = \begin{pmatrix} G_r^T G_r & G_r R_j^T t_j \\ t_j^T R_j G_r & t_j^T t_j \end{pmatrix} \quad (24)$$

With $\omega_j = (A_j A_j^T)^{-1}$ is the projection of the absolute conic ($\Omega \sim I_3$) in the image j .

The expression (20) gives:

$$G_r^T G_r = \begin{pmatrix} a_r^2 & \frac{a_r^2}{2} \\ \frac{a_r^2}{2} & a_r^2 \end{pmatrix} \quad (25)$$

From the formulas (23) and (24) we can deduce that the four upper left coefficients ($G_r^T G_r$) of $L_{ir}^T \omega_i L_{ir}$ and $L_{jr}^T \omega_j L_{jr}$ are identical.

We note by Q_{ir} and Q_{jr} the matrices that represent respectively the four upper left coefficients ($G_r^T G_r$) of the two matrices $L_{ir}^T \omega_i L_{ir}$ and $L_{jr}^T \omega_j L_{jr}$.

From expressions (23) and (24), we conclude that:

$$Q_{ir} \sim Q_{jr} \tag{26}$$

We put:

$$Q_{ir} = \begin{pmatrix} q_{1ir} & q_{2ir} \\ q_{2ir} & q_{1ir} \end{pmatrix} \tag{27}$$

and

$$Q_{jr} = \begin{pmatrix} q_{1jr} & q_{2jr} \\ q_{2jr} & q_{1jr} \end{pmatrix} \tag{28}$$

From the equations (26), (27) and (28) we can deduce the following equalities between images i and j :

$$\frac{q_{1ir}}{q_{2ir}} = \frac{q_{1jr}}{q_{2jr}} \quad \text{with } r = 1 \dots n \tag{29}$$

Where n represents the number of matches between images i and j .

The expression (29) gives:

$$q_{1ir}q_{2jr} - q_{1jr}q_{2ir} = 0 \quad \text{with } r = 1 \dots n \tag{30}$$

Therefore, for each couple (p_{ir}, p_{jr}) , we obtain one equations. Then we need at least ten matching couples to estimate the ten parameters of the camera used. Indeed, with this approach we detect an important number of matched points couples (n matches) which provides a large number of equations (n equations). Furthermore, the importance of this approach lies in the fact that the self-calibration equations become related to the couples of the matched points.

The system (30) is non-linear; therefore, we will minimize the following non-linear cost function:

$$\min_{(\omega_i, \omega_j)} \sum_{i=1}^{m-1} \sum_{j=i+1}^m \sum_{r=1}^n (q_{1ir}q_{2jr} - q_{1jr}q_{2ir})^2 \tag{31}$$

With m represents the number of images, and n represents the number of matches.

To solve the function (31) we use the Levenberg-Marquardt algorithm [1]. The latter requires an initialization step. For this, we assume that some conditions are satisfied on the vision system.

- The pixels are squared therefore $\varepsilon_i = \varepsilon_j = 1$ and $\tau_i = \tau_j = 0$.
- The principal point is in the center of the images, therefore: $u_{0i} = v_{0i} = u_{0j} = v_{0j} = 256$ (because the size of images used is 512×512). And the focal lengths (f_i, f_i) are estimated from the expression (30) (by replacing the parameters by their values in this expression).

7. Experimentation

7.1. Simulations

In this section, we realize a simulation of a sequence of ten 512×512 images of an unknown planar scene to show the performance and robustness of our approach. We estimated the camera parameters by a classical method of calibration from a planar pattern, the parameters obtained are as follows: $f = 1230$, $\varepsilon = 0.93$, $u_0 = 261$ and $v_0 = 254$. In the

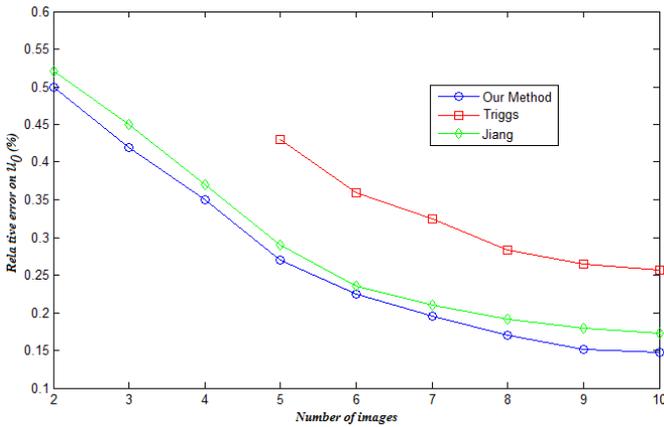


Fig. 6. Relative error on u_0 (%) according to number of images

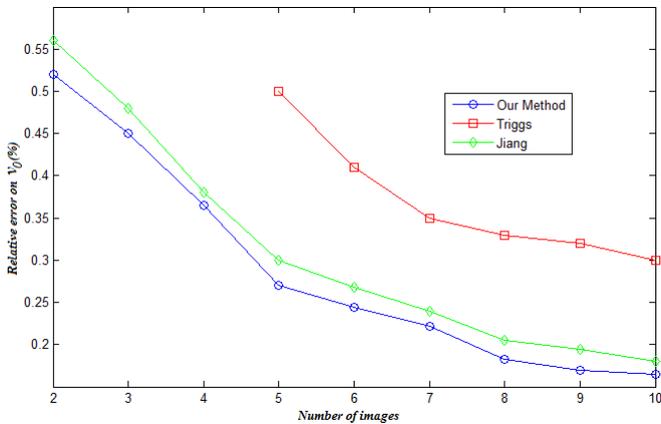


Fig. 7. Relative error on v_0 (%) according to number of images

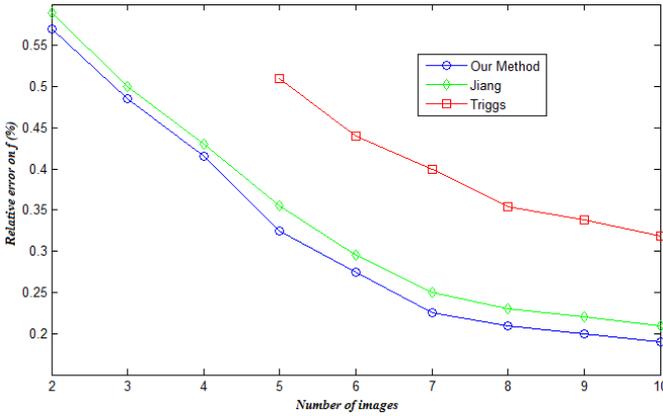


Fig. 8. Relative error on f (%) according to number of images

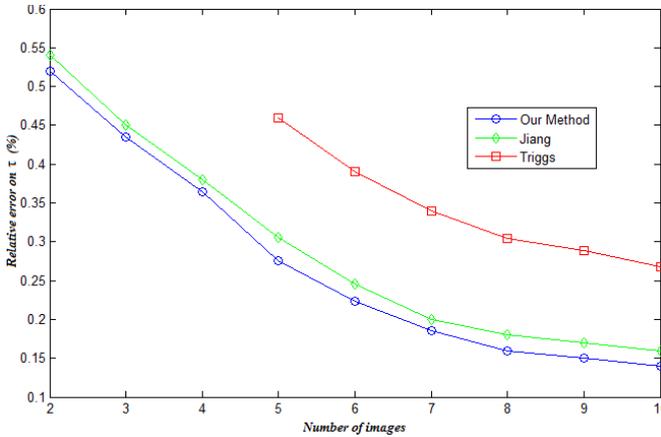


Fig. 9. Relative error on τ (%) according to number of images

first time, we have carried the detection of interest points by Harris algorithm [3], and we have matched these points by the ZNCC correlation function [16, 22] aiming to estimate the homographies (from the 4 matches) between couples of images by the RANSAC algorithm [2]. These homographies are used with the projection of the planar scene points in the images to determine the different projection matrices. The resolution of a non-linear equation system which is formulated from the points of the planar scene and its projections in the images (the matches) by Levenberg-Marquardt algorithm [1] allows

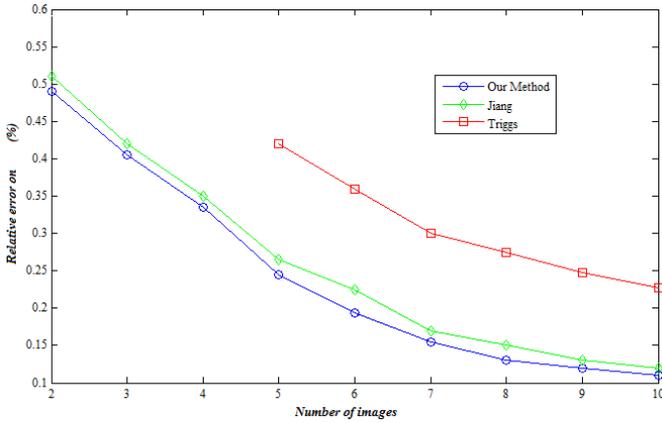


Fig. 10. Relative error on ε (%) according to number of images

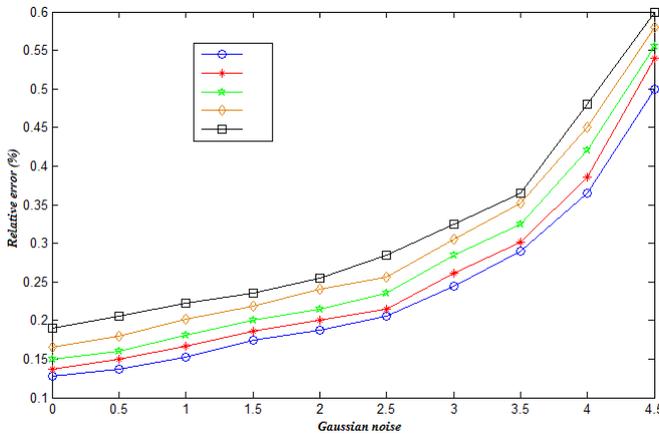


Fig. 11. Relative error on $(u_0, v_0, \varepsilon, \tau)$ and f (%) according to Gaussian noise

to treat the intrinsic parameters of the different cameras used. For this, we compare our method with two other efficient methods which are Triggs [11] and Jiang [29]. In this simulation, we discussed the influence of the number of images used on the relative errors corresponding to u_0 , v_0 , ε , τ and f (represented respectively in Figures 6, 7, 8, 9 and 10) by our approach and the approaches of Triggs [11] and Jiang [29].

We also carry out a second simulation to test the performance of our method with respect to noise. To do this, we add to all image pixels a Gaussian white noise with

standard deviation σ ($0 < \sigma \leq 4.5$) with a step of 0.5. For each noise level, we calculate the relative errors corresponding to u_0 , v_0 , ε and f (represented in Figure 11) by our approach.

7.2. Analysis of the simulations

The Figures 6, 7, 8, 9 and 10 show that relative errors corresponding to u_0 , v_0 , ε , τ and f determined by our method decrease almost linearly if the number of images is between 2 and 5. They decrease slowly when the number of images used is between 5 and 8. They become almost stable if the number of images exceeds 8, but when the number of images increases, the parameters to be estimated become very numerous. Consequently, the calculations become more complex, which allows increasing the program execution time.

On the other hand, the Figure 11 shows that the relative errors to u_0 , v_0 , ε , τ and f remain almost stable when the noise value is between 0 and 2.5. They increase slowly if the noise is between 2.5 and 3.5, and they increase quickly if the noise becomes greater than 3.5.

The analysis of the results obtained in Figures 6, 7, 8, 9 and 10, shows that the relative errors corresponding to the parameters u_0 , v_0 , ε , τ and f obtained by our method are similar to those calculated by the method of Jiang [29], and they are a little different to those obtained by the method of Triggs [11]. Triggs uses more than four images to estimate the intrinsic parameters. On the contrary, our method estimates the parameters of the camera used from two images only.

7.3. Real data

Ten 512×512 images of an unknown planar scene are taken by a digital camera characterized by varying parameters from different views to confirm the robustness of the approach presented in this paper. Two (among ten) are shown in Figure 12(a). The interest points and the matches between these two images are shown respectively in Figure 12(b) and Figure 12(c).

For the choice of equilateral triangles in the images shown in Figure 12(c), we denote by (o_i, o_j) a matching between the images i and j , such that (o_i, o_j) are the projections of the origin of the fixed reference in the images i and j . Moreover, let (p_{ri}, p_{rj}) a couple of matching between the images i and j such that p_{ri} and p_{rj} are the projections of the second vertex of the triangle respectively in the images i and j . There exists a unique point M_r in the scene plan such that OP_rM_r is an equilateral triangle, the point M_r is not used in practice (we only projected two points o and P_r in the images, because their correspondents in these images represent a couple of matching).

In our approach, the estimation of the intrinsic parameters is based on the couples of matched points. To obtain efficient solutions, we have performed a regularization phase.

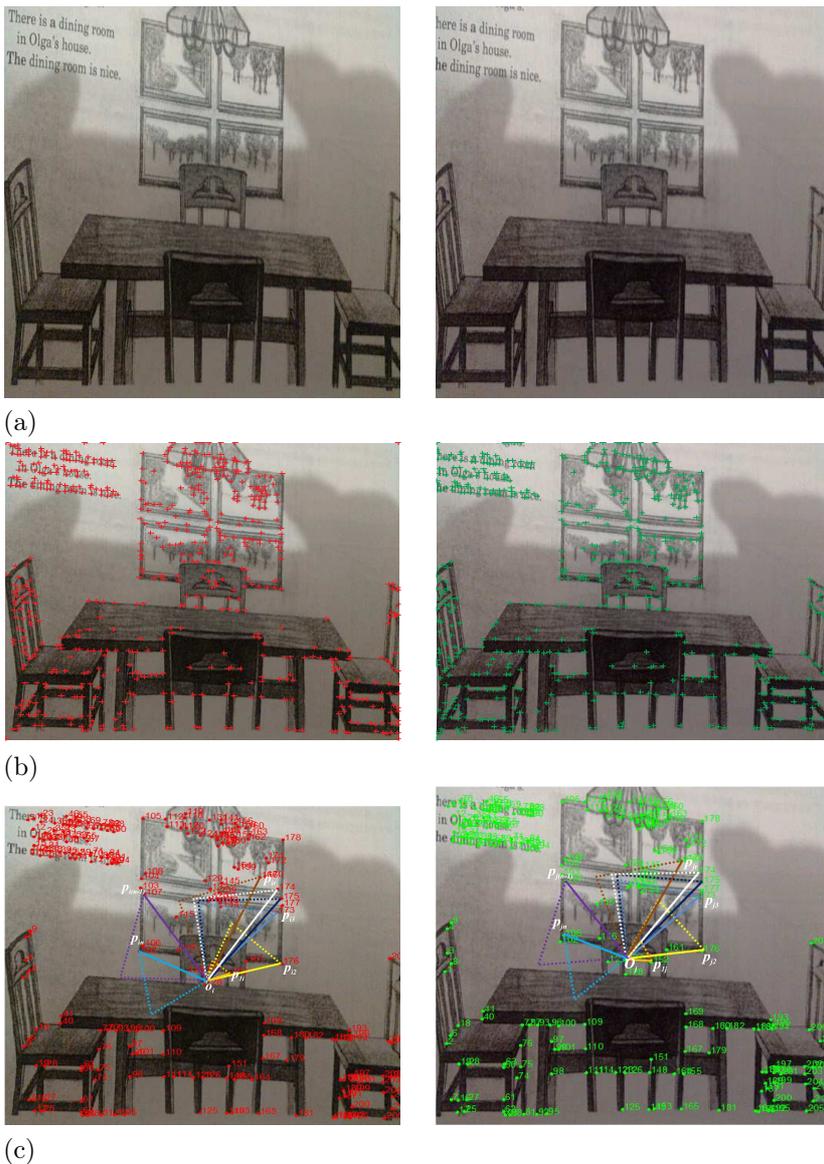


Fig. 12. (a) The two images of the planar scene. (b) The interest points detected by Harris. (c) The matches between the couple of images.

Indeed, the couples of matched points in this phase contain false matches; we eliminate them by the constraint that checks the formula (8).

The projection of the points of the planar scene in the two images allows estimating the geometric entities (the homographies and the projection matrices). Afterwards, the solution of a non-linear system of equations (formula (31)) allows estimating the elements the image of the absolute conic and finally the intrinsic parameters of the camera.

The Table 1 below represents the intrinsic parameters estimated by our approach.

Tab. 1. The Results of Intrinsic Camera Parameters Estimated by the Two Methods

		f	ε	τ	u_0	v_0
The present method	Image 1	1237	0.93	0.04	258	262
	Image 2	1233	0.95	0.03	260	254
	Image 3	1247	0.92	0.01	253	259
	Image 4	1251	0.94	0.02	257	261
Jiang	Image 1	1249	1	0	251	259
	Image 2	1243	0.93	0.05	263	261
	Image 3	1255	0.95	0.02	259	264
	Image 4	1267	0.91	0.04	251	260

According to the experiments results on real data: the two images presented in Figure 12(a) and the eight other images, we conclude that our approach gives the results closer to those obtained by Jiang [29]. This shows, on the one hand, the accuracy of the approach presented in this present work. In the other hand, our algorithms converge rapidly to the optimal solution, because we estimated the five parameters of the camera for each view, knowing that we did not use any constraint on the intrinsic parameters of the camera, on the contrary, Jiang assumes that $\tau = 0$ and $\varepsilon = 1$ in the first image. These constraints influence directly the results of self-calibration. Our method and Jiang's can calibrate the camera from only two images, but the suggested method presents the advantage does not require any constraint on the self-calibration system, compared to Jiang method which requires the constraints on intrinsic parameters of the camera.

8. Conclusion

In this work, the problem of the self-calibration of cameras with varying intrinsic parameters has been addressed by using an unknown planar scene. This approach is based on the use of equilateral triangles assumed in the planar scene and the transformation matrix between them. The projection of vertices of equilateral triangles in the planes of the images and , and the relationship between images of absolute conic for each pair of

images allow formulating a non-linear cost function. The minimization of this function by the Levenberg-Marquardt algorithm provides the intrinsic parameters of the cameras used. The advantages of this method are the use of any camera (no constraints on the intrinsic parameters) and two images of the planar scene are sufficient to calibrate the cameras used. The found experiments results are satisfactory, which shows the robustness and reliability of our approach.

References

- 1977**
- [1] J. More : The Levenberg-Marquardt Algorithm, Implementation and Theory. In G. A. Watson, editor, Numerical Analysis, Lecture Notes in Mathematics 630. Springer-Verlag, pp. 105-116.
- 1981**
- [2] M. A. Fischler and R. C. Bolles : Random sample consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Graphics and Image Processing, pp. 381-395.
- 1988**
- [3] C. Harris and M. Stephens : A combined Corner et Edge Detector. 4th Alvey vision Conference. pp. 147-151.
- 1992**
- [4] O. Faugeras, Q.-T. Luong, S. Maybank : Camera Self-Calibration: Theory and Experiments . In: G. Sandini (Ed.), Proceedings of the European Conference on Computer Vision, Springer-Verlag, pp. 321-334.
- [5] S. J. Maybank and O. D. Faugeras : A Theory of Self Calibration of a Moving Camera. International Journal on Computer Vision, pp. 123-151.
- 1994**
- [6] R. I. Hartley : Self-Calibration From Multiple Views with a Rotating Camera. In ECCV, pp. 471-478.
- 1996**
- [7] A. Heyden, K. Aström : Euclidean Reconstruction from Constant Intrinsic Parameters. In: Proc. International Conference on Pattern Recognition, pp. 339-343.
- [8] M. Pollefeys, L.V. Gool, M. Oosterlinck : The Modulus Constraint: A New Constraint for Self-Calibration. In: Proc. Int. Conf. on Pattern Recognition, pp. 349-353.
- 1997**
- [9] S. M. Smith and J. M. Brady : SUSAN - A New Approach to Low Level Image Processing. International Journal of Computer Vision, 45-78.
- 1998**
- [10] M. Pollefeys, R. Koch and L. Van Gool : Self-Calibration and Metric Reconstruction in Spite of Varying and Unknown Internal Camera Parameters. ICCV, pp. 90-95.
- [11] B. Triggs : Autocalibration from Planar Sequences. In Proceedings of 5th European Conference on Computer Vision, pp. 89-105.
- 1999**
- [12] P. F. Sturm and S. J. Maybank : On Plane-Based Camera Calibration: A General Algorithm, Singularities, Applications. In Proceedings of the CVPR-IEEE, pp. 432-437.
- 2000**
- [13] P. Sturm : A Case Against Kruppa's Equations for Camera Self-Calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1199-1204.

- [14] Z. Zhang : A Flexible New Technique for camera Calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence. pp. 1330-1334.
- 2001**
- [15] M. Wilczkowiak, E. Boyer, P. Sturm : Camera Calibration and 3D Reconstruction from Single Images Using Parallelepipeds. In ICCV, Vancouver, pp. 142-148.
- 2002**
- [16] M. Lhuillier and L. Quan : Quasi-Dense Reconstruction from Image Sequence. ECCV.
- [17] P. Sturm : Critical motion sequences for the self-calibration of cameras and stereo systems with variable focal length. Image and Vision Computing, 20(5-6): 415-426.
- 2003**
- [18] P. Gurdjos and P. Sturm : Methods and Geometry for Plane-Based Self-Calibration. CVPR, pp. 491-496.
- [19] P. Liu, J. Shi, J. Zhou and L. Jiang : Camera Self-Calibration Using the Geometric Structure in Real Scenes. In: Proceedings of the Computer Graphics International, pp. 262-265.
- 2004**
- [20] D. G. Lowe : Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, 91-110.
- 2006**
- [21] B. Bouda, Lh. Masmoudi and D. Aboutajdine. A New Grey Level Corner Detection Based on Electrostatic Model. ICGST-GVIP, 21-26.
- [22] A. Saaidi, H. Tairi and K. Satori : Fast Stereo Matching Using Rectification and Correlation Techniques. ISCCSP, 2nd International Symposium on Communications, Control and Signal Processing.
- 2008**
- [23] Y. Li, Y.S. Hung, S. Lee : A Stratified Self-Calibration Method for Circular Motion in Spite of Varying Intrinsic Parameters. Image and Vision Computing, pp. 731-739.
- [24] A. Saaidi, A. Halli, H. Tairi and K. Satori : Self -Calibration Using a Particular Motion of Camera. In WSEAS Transaction on Computer Research, pp. 295-299.
- 2009**
- [25] J. Kima, I. Kweon : Camera calibration based on arbitrary parallelograms. Computer Vision and Image Understanding, pp. 1-10
- [26] A. Saaidi, A. Halli, H. Tairi and K. Satori : Self-Calibration Using a Planar Scene and Parallelogram. ICGST-GVIP, pp. 41-47.
- 2011**
- [27] A. El-attar, M. Karim, H. Tairi, S. Ionita : A Robust Multistage Algorithm for Camera Self-Calibration Dealing with Varying Intrinsic Parameters. JATIT, pp. 46-54.
- 2012**
- [28] A. Baataoui, I. El Batteoui, A. Saaidi and K. Satori : Camera Self-Calibration by an Equilateral Triangle. International Journal of Computer Applications, pp. 29-34.
- [29] Z. Jiang, S. Liu : Self-calibration of Varying Internal Camera Parameters Algorithm Based on Quasi-affine Reconstruction. Journals of Computers, pp. 774-778.
- [30] Y. Shang, Z. Yue, M. Chen and Q. Song : A New Method of Camera Self-Calibration Based on Relative Lengths. Information Technology Journal, pp. 376-379.
- 2013**
- [31] N. El Akkad, M. Merras, A. Saaidi, and K. Satori : Robust Method for Self-Calibration of Cameras Having the Varying Intrinsic Parameters. JATIT, pp. 57-67.

A CURVATURE-TENSOR-BASED PERCEPTUAL QUALITY METRIC FOR 3D TRIANGULAR MESHES

Fakhri Torkhani, Kai Wang, and Jean-Marc Chassery

Gipsa-lab, CNRS UMR 5216, Université de Grenoble

11 rue des Mathématiques, F-38042 Saint Martin d'Hères, France

{fakhri.torkhani, kai.wang, jean-marc.chassery}@gipsa-lab.grenoble-inp.fr

Abstract. Perceptual quality assessment of 3D triangular meshes is crucial for a variety of applications. In this paper, we present a new objective metric for assessing the visual difference between a reference triangular mesh and its distorted version produced by lossy operations, such as noise addition, simplification, compression and watermarking. The proposed metric is based on the measurement of the distance between curvature tensors of the two meshes under comparison. Our algorithm uses not only tensor eigenvalues (i.e., curvature amplitudes) but also tensor eigenvectors (i.e., principal curvature directions) to derive a perceptually-oriented tensor distance. The proposed metric also accounts for the visual masking effect of the human visual system, through a roughness-based weighting of the local tensor distance. A final score that reflects the visual difference between two meshes is obtained via a Minkowski pooling of the weighted local tensor distances over the mesh surface. We validate the performance of our algorithm on four subjectively-rated visual mesh quality databases, and compare the proposed method with state-of-the-art objective metrics. Experimental results show that our approach achieves high correlation between objective scores and subjective assessments.

Key words: 3D triangular mesh, perceptual quality, human visual system, objective metric, curvature tensor, visual masking.

1. Introduction

Three-dimensional (3D) triangular meshes have become the *de facto* standard for digital representation of 3D objects, and by now have found wide use in various applications, such as digital entertainment, medical imaging and computer-aided design [30]. As a rule, 3D triangular meshes undergo some lossy operations, like simplification, compression and watermarking. Although these operations are necessary to speed up and facilitate the transmission, storage and rendering of 3D meshes, or to enforce the copyright protection, they inevitably introduce distortion to the original, unprocessed mesh. However, such distortion might degrade the quality of service associated with the mesh model. Since end users of 3D triangular meshes are often human beings, it is thus important to derive some means to faithfully evaluate the degree of *visual distortion* introduced to a 3D mesh. Directly asking human subjects to evaluate the visual distortion is obviously unpractical in most real-world applications, as such subjective evaluation is time-consuming and

costly. Therefore, it is necessary to develop objective metrics (i.e., software tools) that can accurately predict the result of a subjective visual quality assessment of the for 3D triangular meshes [39].

Although during the last decade we have seen tremendous advance in objective image visual quality assessment [19, 34], the research on objective *mesh visual quality* (MVQ) assessment is still at its early stage, with very few metrics proposed [39]. A possible way to evaluate the perceptual quality of 3D meshes is to apply image quality metrics on 2D images generated through 3D model rendering under several pre-selected viewing positions. The first problem with this so-called *image-based* approach is how to select the viewing positions that the 2D image projections are generated from. In our opinion, a both optimum and automatic selection of such viewing positions is a very difficult problem in itself. Furthermore, researchers wonder whether it is appropriate to use 2D image quality metrics to evaluate the visual quality of 3D meshes. In order to answer this question, Rogowitz and Rushmeier [10] investigated the reliability of the image-based approach by conducting a series of experiments to compare the perceived quality of simplified 3D meshes (as presented in a series of continuous viewing positions) and of their corresponding 2D image projections. The experimental results seem to imply that the perceptual quality of 3D meshes is in general not equivalent to the visual quality of their 2D image projections.

Inspired and motivated by the results of Rogowitz's and Rushmeier's experiments, the research community is now paying more attention to the development of *model-based* MVQ metrics. This approach suggests that it would be more reasonable to relate MVQ directly to the 3D shape of the mesh model than to its 2D image projections. Similarly to the fact that PSNR (peak signal-to-noise ratio) and MSE (mean squared error) fail to capture the visual quality of an image [29], it is not surprising to see that classical mesh geometric distances (e.g., root mean squared error and Hausdorff distance) [6, 11] have been demonstrated to be irrelevant to human visual perception, and thus failing to predict the visual difference between an original mesh (also called *reference* mesh) and the distorted one [39]. In Fig. 1, we show an example where the Hausdorff distance (*HD*) fails to provide correct MVQ assessments: The Hausdorff distance between the original mesh in Fig. 1(a) and the distorted mesh in Fig. 1(b) is smaller than that between the original mesh and the distorted mesh in Fig. 1(c); however, conversely, the model of Fig. 1(c) has obviously better perceptual quality than the one shown in Fig. 1(b).

In order to develop an effective objective MVQ metric, it is necessary to make use of perceptually relevant features, and take into account some important properties of the *human visual system* (HVS). In this paper, we choose surface curvature amplitudes and principal curvature directions as perceptually relevant features. As discussed later, both features are important properties of the mesh surface that can be derived from eigendecomposition of the curvature tensor. Meanwhile, we integrate some HVS properties in the metric, in particular the *visual masking* effect. In the case of MVQ assessment and

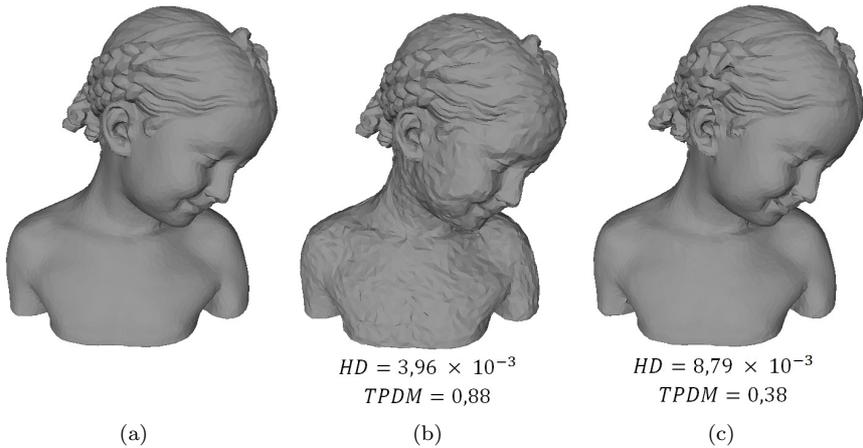


Fig. 1. (a) Original Bimba model; (b) distorted Bimba with noise added in smooth regions; and (c) distorted Bimba with noise added in rough regions. The distortions introduced for Hausdorff distances (HD) [6, 11] and $TPDM$ (Tensor-based Perceptual Distance Measure, proposed in this paper) are given below the distorted models.

as illustrated in Fig. 1, this effect mainly means that distortions in mesh surfaces are usually more visible in smooth regions than in rough regions. This paper is an extension of the earlier work presented in [36], and our contributions can be summarized as follows.

- Development of an effective model-based approach to the assessment of visual mesh quality based on a novel distance measure between mesh curvature tensors.
- Use of not only curvature amplitudes, but also of principal surface directions (which have been shown perceptually important, as discussed later) to define the curvature tensor distance.
- Integration of some HVS features in the metric: We introduce a roughness-based weighting of the local curvature tensor distance to simulate the visual masking effect, and a processing step similar to the *divisive normalization transform* to mimic an important neural mechanism, known as *adaptive gain control* [27, 34].
- The source code of our MVQ metric is freely available on-line at: <http://www.gipsa-lab.fr/~fakhri.torkhani/software/TPDM.rar>.

The proposed metric, named $TPDM$ (Tensor-based Perceptual Distance Measure), has been extensively tested on four subjectively-rated visual mesh quality databases, and has been compared with state-of-the-art objective MVQ metrics. The experimental results show that our metric achieves high correlation between objective scores and subjective assessments. As a simple example to illustrate its effectiveness, $TPDM$ provides perceptually coherent assessments for the visual quality of the distorted meshes shown

in Figs. 1(b) and 1(c) (a lower *TPDM* value implies a better visual quality of the distorted model): The *TPDM* distance between the original mesh and the distorted mesh in Fig. 1(b) is 0.88, while that between the original mesh and the distorted mesh in Fig. 1(c) is 0.38. The proposed metric has the potential to be used, for instance, in the benchmarking of a variety of mesh processing algorithms (e.g., compression, watermarking, remeshing, etc.), or to guide the design of new perceptually-oriented algorithms.

The remainder of this paper is organized as follows. Related work on model-based MVQ assessment and our motivation for the design of *TPDM* are presented in Section 2. Section 3 details the pipeline of the proposed MVQ metric. The experimental results are presented in Section 4, including performance evaluation of *TPDM*, comparison with state-of-the-art MVQ metrics, and two simple examples to illustrate potential applications of the proposed metric. Finally, we draw the conclusion and suggest several future work directions in Section 5.

2. Related Work and Motivation

2.1. Model-based MVQ assessment

During the last decade, there has been increasing interest in the research on perceptual quality assessment of 3D meshes. To our knowledge, the first perceptually-oriented model-based MVQ metric was introduced by Karni and Gotsman [9] for the evaluation of their mesh compression algorithm. That metric is actually a weighted combination of root mean squared errors in vertex positions and errors in mesh Laplacian coordinates. Based on the fact that the local smoothness measure has a more important visual effect, Sorkine et al. [14] enhanced Karni and Gotsman’s metric by assigning a greater weight to the errors in mesh Laplacian coordinates. Although initially proposed as by-products for evaluating mesh compression algorithms, these two metrics have triggered promising studies focusing on the perceptual quality assessment of 3D meshes.

Corsini et al. [22] developed two perceptual metrics for the visual quality assessment of watermarked meshes, named, respectively, $3DWPM_1$ and $3DWPM_2$. The visual distortion is evaluated in these two metrics as the roughness difference between the original and watermarked meshes. Two roughness measures were proposed: The first one is based on statistics (within multiscale local windows) of dihedral angles over the mesh surface, while the second roughness measure is defined as the geometric difference between a mesh model and its carefully smoothed version.

A physically-inspired MVQ metric was proposed by Bian et al. [25]. They considered 3D meshes as objects with elasticity, and assumed that the visual difference between a pair of meshes is related to the strain energy that is required to induce the deformation between them. It was shown in [25] that this metric was effective in assessing small

visual differences between meshes with constant connectivity (i.e., the same adjacency relationship between the bmesh vertices).

The research on MVQ assessment could benefit from the much more fruitful literature on image visual quality assessment. Following this line of research, Lavoué et al. [17] proposed a metric called structural mesh distortion measure (*MSDM*), which can be considered as an extension of the well-known structure similarity index for 2D images [16] to the case of 3D triangular meshes. *MSDM* relates the visual degradation to the alteration of local statistics (i.e., mean, variance and covariance) of mesh curvature amplitudes. An improved multiscale version *MSDM2* [32] has been proposed, which also integrates a vertex matching preprocessing step to allow the comparison of two meshes with different vertex connectivities.

Váša and Rus [37] proposed a dihedral angle mesh error (*DAME*) metric to compare triangular meshes sharing the same connectivity. *DAME* relies on the oriented surface dihedral angles to evaluate the perceptual distortion, and integrates the visual masking effect, as well as the visibility model which accounts for the probable viewing positions and directions of the models to be compared. This metric is computationally efficient, and has a good correlation with subjective assessment.

Recently, Wang et al. [38] introduced the fast mesh perceptual distance (*FMPD*) measure. This metric is based on a local roughness measure derived from the Gaussian curvature of the mesh surface. *FMPD* estimates the perceptual distance as the difference between the global roughness values of the two meshes under comparison. Therefore, the metric does not require a mesh correspondence or registration preprocessing step, and can be applied to compare meshes with different connectivities. *FMPD* is in essence a *reduced-reference* metric, since only the global roughness of the original mesh and some parameter values are required to carry out the visual quality assessment of a distorted mesh. In contrast, *full-reference* MVQ metrics, such as *MSDM*, *MSDM2* and *DAME*, require the availability of full information about the original mesh.

A summary of the most representative model-based MVQ metrics, along with our *TPDM* metric, is presented in Tab. 1. In that table, we also list two popular mesh geometric distances *RMS* (root mean squared error) and *HD* (Hausdorff distance). Despite their poor correlation with human visual perception [39], nowadays *RMS* and *HD* are still largely used in the evaluation of various mesh processing algorithms. The metrics are summarized according to three different algorithmic aspects: the mesh feature used for MVQ assessment, the information about the reference mesh required (i.e., whether the metric is classified as full-reference or reduced-reference one), and whether the metric requires the two meshes under comparison to have the same connectivity. Quantitative comparisons of these metrics are presented in Section 4.

Tab. 1. Summary of model-based mesh perceptual quality metrics.

Metric	Feature	Information about reference mesh	Connectivity constraint
<i>RMS</i> [6, 11]	Surface-to-surface distance	Full-reference	No
<i>HD</i> [6, 11]	Surface-to-surface distance	Full-reference	No
<i>3DWPM</i> [22]	Global roughness	Reduced-reference	Yes
<i>MSDM</i> [17]	Curvature amplitude	Full-reference	Yes
<i>MSDM2</i> [32]	Multiscale curvature amplitude	Full-reference	No
<i>DAME</i> [37]	Dihedral angle	Full-reference	Yes
<i>FMPD</i> [38]	Global roughness	Reduced-reference	No
<i>TPDM</i>	Curvature tensor distance	Full-reference	No

2.2. Motivation for tensor-based MVQ assessment

MSDM2 has a good correlation with subjective scores [32], though by considering only the modification in mesh curvature amplitudes. We argue that a modification in the *principal surface directions* as defined by the *orthogonal* directions of minimum and maximum curvatures is also important for MVQ assessment. As shown in Fig. 2, maximum and minimum curvature directions represent salient structural features of the surface and thus should be visually important. Indeed, when drawing a 3D object, one strategy of caricaturists is to draw strokes on these lines of curvatures [23]. For example, we may expect that the drawings of either trained artists or untrained amateurs, when asked to complete a line drawing of the Bimba model shown in Fig. 2(a), would be similar to the images shown in Figs. 2(b) and 2(c). The perceptual importance of principal surface directions have been noticed by computer graphics and geometry processing experts: They have been successfully used for describing [3] and illustrating [8] complex 3D objects, as well as for guiding a high-performance anisotropic remeshing algorithm [12].

Motivated by the above observation, in this paper we introduce a new MVQ metric *TPDM* which makes use of more information that can be extracted from mesh curvature tensors, i.e., both the curvature amplitudes and the principal surface directions. In Section 4, we will show that experimentally our tensor-based metric achieves a high correlation with the subjective scores of mesh visual quality. In particular, the use of information on principal surface directions in *TPDM* appears to help improve the assessment performance when compared to the state-of-the-art metric *MSDM2* that uses only curvature amplitudes. Before presenting the technical details of the proposed metric in Section 3, in the next subsection we will briefly introduce a technique for the estimation of mesh curvature tensors, and explain how to obtain curvature amplitudes and principal curvature directions from the tensor.

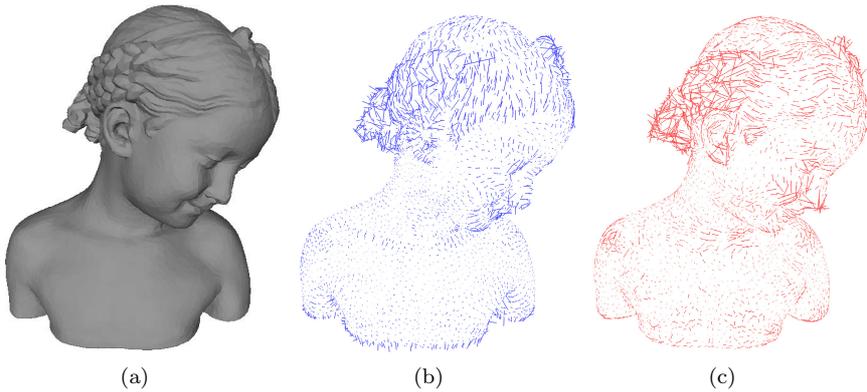


Fig. 2. (a) Bimba’s model; (b) maximum curvature directions of Bimba scaled by maximum curvature values; and (c) minimum curvature directions of Bimba scaled by minimum curvature values.

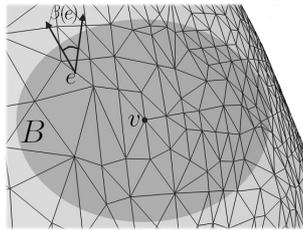


Fig. 3. Geometric elements used to compute the curvature tensor.

2.3. Curvature tensor estimation

Estimation of the mesh curvature tensor is a well-researched problem. So far, the most popular estimation technique has been the one from Cohen-Steiner and Morvan [13]. Based on the solid foundation of normal cycle theory, they derived an elegant per-vertex curvature tensor estimation. Tensors computed on edges are averaged on a geodesic disk window B of user-defined size to obtain the curvature tensor \mathcal{T} on each vertex v :

$$\mathcal{T}(v) = \frac{1}{|B|} \sum_{\text{edges } e} \beta(e) |e \cap B| \bar{e} \bar{e}^t, \tag{1}$$

where $|B|$ is the area of the geodesic disk, $\beta(e)$ is the signed angle between the normals of the two triangles incident to edge e , $|e \cap B|$ is the length of the part of e inside B , \bar{e} and \bar{e}^t are the unit vector in the direction of e and its transpose (cf. Fig. 3), respectively. The minimum and maximum curvature amplitudes (denoted by κ_{\min} and κ_{\max}), respectively)

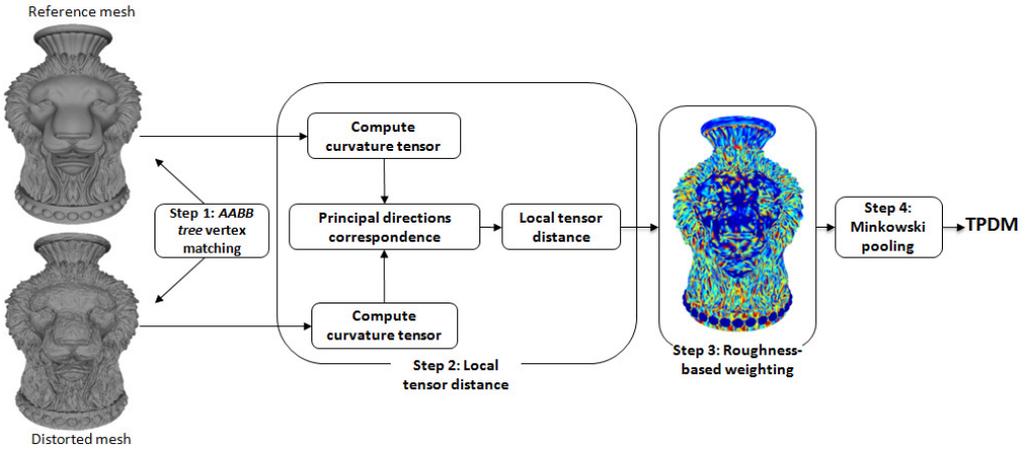


Fig. 4. Block diagram of the pipeline of *TPDM* (Tensor-based Perceptual Distance Measure). In the roughness map within the block of “Step 3: Roughness-based weighting”; warmer colors represent larger values (i.e., where the local surface is rougher).

are the absolute values of the two non-zero eigenvalues of the tensor \mathcal{T} , and the principal surface directions are the associated two eigenvectors (denoted by 3D vectors γ_{\min} and γ_{\max} , respectively). In Section 3, we will derive a perceptually-oriented distance between curvature tensors by incorporating the information from both their eigenvalues and their eigenvectors, and will use this distance to conduct the MVQ assessment.

3. MVQ Assessment Based on Curvature Tensor Distance

An overview of the processing pipeline for the proposed MVQ metric *TPDM* is illustrated in Fig. 4. First of all, in order to compare two meshes with potentially different connectivities, we perform a preprocessing step of vertex matching between the two meshes under comparison, based on the AABB tree data structure implemented in the CGAL library [35]. The objective of this step is to find, for each vertex of the reference mesh, a corresponding point on the surface of the distorted mesh. The second step is to compute a curvature tensor at each vertex of the two meshes, and then to derive the distance between the tensors of each vertex in the reference mesh and its counterpart in the distorted mesh. Both the curvature amplitudes and the principal curvature directions are involved in the tensor distance calculation. Before this local tensor distance computation, a correspondence relationship has to be established between the principal curvature directions and the curvature amplitudes of the two tensors to be compared, so as to determine for each vertex how to exactly calculate the tensor distance. In the

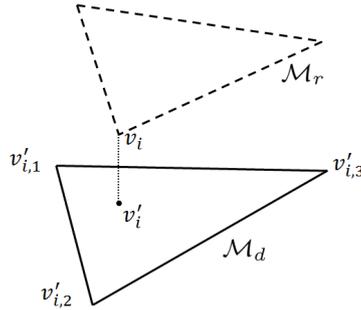


Fig. 5. Projection of a vertex v_i of the reference mesh \mathcal{M}_r onto the surface of the distorted mesh \mathcal{M}_d . Here the projection is v'_i , a point on the triangular facet that is constituted of $v'_{i,1}$, $v'_{i,2}$ and $v'_{i,3}$. For the sake of simplicity, we only show one single facet in \mathcal{M}_r and in \mathcal{M}_d .

third step, this local tensor distance is weighted by two roughness-based factors in order to account for the visual masking effect of HVS, which is vital for conducting a correct MVQ assessment. In the fourth and last step, we use a surface-weighted Minkowski pooling of the local *TPDM* distances to obtain a global *TPDM* value. In what follows, we will present the technical details of each of the four steps.

3.1. Vertex matching preprocessing

In order to establish the correspondence between the vertices of the two meshes, analogously as in the preprocessing step in *MSDM2* [32], we use the AABB tree data structure implemented in the CGAL library [35] to perform a fast and simple vertex projection from the reference mesh \mathcal{M}_r to the surface of the distorted mesh \mathcal{M}_d . As a result of this matching step, each vertex v_i in \mathcal{M}_r is assigned a corresponding point v'_i on the surface of \mathcal{M}_d . Note that v'_i is in general not a vertex of \mathcal{M}_d , but a point on a certain triangular facet T'_i of \mathcal{M}_d composed of three vertices $v'_{i,1}$, $v'_{i,2}$ and $v'_{i,3}$ (cf. Fig. 5). In this general case, the local *TPDM* distance associated to v_i , denoted by $LTPDM_{v_i}$, is computed as the barycentric interpolation [2] of the three local perceptual distances, between v_i and $v'_{i,1}$, v_i and $v'_{i,2}$, and finally v_i and $v'_{i,3}$, respectively:

$$LTPDM_{v_i} = \sum_{k=1}^3 b_k(v'_i) LPD_{v_i, v'_{i,k}}, \tag{2}$$

where $LPD_{v_i, v'_{i,k}}$ is the local perceptual distance between v_i and the k -th vertex of the triangular facet T'_i that contains the projection v'_i , and $b_k(v'_i)$ is the k -th barycentric coordinate [2] of v'_i within T'_i . The next two subsections focus on how to derive the local perceptual distance $LPD_{v_i, v'_{i,k}}$.

3.2. Local tensor distance

The derivation of the local perceptual distances $LPD_{v_i, v'_{i,k}}$ is based on the local tensor distances between v_i and $v'_{i,k}$, denoted by $LTD_{v_i, v'_{i,k}}$ for $k = 1, 2, 3$. For this purpose, we first estimate the curvature tensors at each vertex of \mathcal{M}_r and \mathcal{M}_d . The tensors on the two vertices v_i and $v'_{i,k}$ are hereafter denoted by \mathcal{T}_{v_i} and $\mathcal{T}_{v'_{i,k}}$, respectively. They are computed using Equation (1), with a local window established as the intersection of the mesh surface and the Euclidean sphere [13] that is centered on the vertex and has, experimentally, a radius equal to 0.5% of the bounding box diagonal of \mathcal{M}_r .

As mentioned earlier, we want to use the differences between both the curvature amplitudes (i.e., tensor eigenvalues) and the principal curvature directions (i.e., tensor eigenvectors) to derive the local tensor distance. We should first of all decide between which curvature amplitudes/directions we will compute the differences. The straightforward choice is to derive the difference between the minimum curvature of \mathcal{T}_{v_i} and the minimum curvature of $\mathcal{T}_{v'_{i,k}}$ (the same for the differences in the maximum curvature amplitudes, in the minimum curvature directions and in the maximum curvature directions). Then we can combine the obtained four differences to derive the local tensor distance. However, we find that this simple “min→min, max→max” correspondence between the tensor elements results in poor MVQ assessments, especially in the situations where the principal directions are severely disturbed after medium and strong distortions, and where the mesh contains a large portion of locally isotropic regions (in these regions, the values of minimum and maximum curvature amplitudes are close to each other, so that the minimum and maximum curvatures may change roles even after small-amplitude distortions).

Motivated by this observation, we adopt another rule for establishing the correspondence relationship between the curvature amplitudes/directions of \mathcal{T}_{v_i} and $\mathcal{T}_{v'_{i,k}}$, which is based on the *minimum angular distance criterion* between the principle curvature directions. More precisely, for γ_{\min} (i.e., the minimum curvature direction) of \mathcal{T}_{v_i} , we find the principal direction of $\mathcal{T}_{v'_{i,k}}$ that has the smallest angular distance to it (this direction is denoted by γ'_1), and then relate γ_{\min} to γ'_1 . Accordingly, κ_{\min} (i.e., the minimum curvature amplitude) of \mathcal{T}_{v_i} is related to the curvature amplitude associated to γ'_1 (denoted by κ'_1). That is, if γ'_1 is the minimum (or maximum) curvature direction of $\mathcal{T}_{v'_{i,k}}$, then κ'_1 is the minimum (or maximum) curvature amplitude of $\mathcal{T}_{v'_{i,k}}$. Under the proposed correspondence rule, γ'_1 can be either the minimum or the maximum curvature direction of $\mathcal{T}_{v'_{i,k}}$, as long as this minimum or maximum curvature direction has the smallest angular distance to γ_{\min} . Similarly, the following correspondence relationships are established: $\kappa_{\max} \rightarrow \kappa'_2$ and $\gamma_{\max} \rightarrow \gamma'_2$. It is easy to see that γ'_1 and γ'_2 (and κ'_1 and κ'_2 , respectively) are distinct principal curvature directions (distinct principal curvature amplitudes, respectively) of $\mathcal{T}_{v'_{i,k}}$. In practice, the above correspondence, which is based

on the minimum angular distance criterion, yields better MVQ assessment results than the straightforward “min→min, max→max” correspondence discussed above.

The local tensor distance is computed for each pair of v_i and $v'_{i,k}$ as

$$LTD_{v_i, v'_{i,k}} = \frac{\theta_{\min}}{(\pi/2)} \delta_{\kappa_{\min}} + \frac{\theta_{\max}}{(\pi/2)} \delta_{\kappa_{\max}}, \quad (3)$$

where $\theta_{\min} \in [0, \pi/2]$ is the angle between the curvature lines of γ_{\min} and γ'_1 (similarly, $\theta_{\max} \in [0, \pi/2]$ is the angle between the lines of γ_{\max} and γ'_2), and $\delta_{\kappa_{\min}}$ is a Michelson-like contrast [1] of the curvature amplitudes κ_{\min} and κ'_1 , i.e., $\delta_{\kappa_{\min}} = \left| \frac{\kappa_{\min} - \kappa'_1}{\kappa_{\min} + \kappa'_1 + \varepsilon} \right|$ with ε a stabilization constant fixed as 5% of the average mean curvature of \mathcal{M}_r (similarly, $\delta_{\kappa_{\max}} = \left| \frac{\kappa_{\max} - \kappa'_2}{\kappa_{\max} + \kappa'_2 + \varepsilon} \right|$). Both the differences in the curvature amplitudes and in the principal surface directions are involved in the derivation of the local tensor distance. Besides its perceptual relevance [4, 5], another reason to use the Michelson-like contrast to evaluate the difference between curvature amplitudes is that in this way both the difference the principal directions (after the normalization by a factor of $\pi/2$) and in the curvature amplitudes are in the same range of $[0, 1]$, so that these two kinds of differences can be easily combined together.

3.3. Roughness-based weighting of local tensor distance

For the development of an effective MVQ metric, we should take into account some HVS features, in particular the visual masking effect [20]. In the context of MVQ assessment, this effect mainly means that the same distortion is less visible in rough regions of the mesh surface than in the smooth regions. In order to account for the visual masking effect, our solution is to modulate the values of $LTD_{v_i, v'_{i,k}}$ by two roughness-based weights (the rougher the local surface is, the smaller the weights are). The local perceptual distance between v_i and $v'_{i,k}$, which incorporates the visual masking effect, is computed as:

$$LPD_{v_i, v'_{i,k}} = RW_i^{(\gamma)} \cdot RW_i^{(\kappa)} \cdot LTD_{v_i, v'_{i,k}}, \quad (4)$$

with $RW_i^{(\gamma)}, RW_i^{(\kappa)} \in [0.1, 1.0]$. They are, respectively, the roughness-based weights derived from the principal surface directions and the curvature amplitudes in the 1-ring neighborhood of v_i . For $RW_i^{(\gamma)}$, we first project all the principal curvature directions at the 1-ring neighbors of v_i on the tangent plane of v_i , and then take the sum of the two angular standard deviations of the projected minimum and maximum curvature directions as the local roughness value. This value is then linearly mapped to $[0.1, 1.0]$ to obtain $RW_i^{(\gamma)}$: The higher the sum of the two angular standard deviations is, the lower the mapped value is. Similarly, to get $RW_i^{(\kappa)}$, we compute the ratio of the Laplacian for the mean curvatures in the 1-ring neighborhood of v_i and the mean curvature at v_i as the

local roughness measure. The Laplacian of the mean curvatures in the 1-ring neighborhood of v_i describes the local variation of the mesh curvature amplitudes. In our metric, we use the cotangent-based mesh Laplacian due to its solid theoretical foundation and its excellent performance in practical applications [18, 24]. This curvature-amplitude-based roughness value is then linearly mapped to $[0.1, 1.0]$ to obtain $RW_i^{(\kappa)}$: The higher the roughness is, the lower the mapped value is. It is worth mentioning that the derivation of the roughness weight $RW_i^{(\kappa)}$ includes a divisive normalization (i.e., the normalization of the Laplacian for mean curvatures by the mean curvature on v_i itself), which is similar to that in the neural mechanism of HVS that partially explains the visual masking effect [27, 34]. Also note that vertices in isotropic regions, i.e., where κ_{\min} and κ_{\max} are close to each other, are treated differently. An anisotropy coefficient ρ_{v_i} is first computed as:

$$\rho_{v_i} = \frac{\kappa_{\max}^{(v_i)} - \kappa_{\min}^{(v_i)}}{\kappa_{\max}^{(v_i)} + \kappa_{\min}^{(v_i)} + \epsilon}, \quad (5)$$

where $\kappa_{\min}^{(v_i)}$ and $\kappa_{\max}^{(v_i)}$ are, respectively, the minimum and maximum curvatures at vertex v_i , and ϵ is a stabilization constant set as 5% of the average mean curvature of \mathcal{M}_r . We consider that the vertices at which ρ_{v_i} is smaller than 0.5 belong to relatively isotropic regions. For these vertices, we set $RW_i^{(\gamma)}$ equal to 1, and accordingly the final roughness-based weight is determined by the value of $RW_i^{(\kappa)}$. The reason is that in isotropic regions, the principal curvature directions are not well-defined and their estimation is not reliable. It is therefore safer to use only the curvature-amplitude-based roughness weight $RW_i^{(\kappa)}$ for the local tensor distance modulation. A roughness map that combines both weights $RW_i^{(\gamma)}$ and $RW_i^{(\kappa)}$ is shown in the ‘‘Step 3’’ block of Fig. 4, where warmer colors represent higher roughness values (i.e., lower roughness-based weights).

Finally, as described by Equation (2), the local tensor-based perceptual distance measure at vertex v_i , $LTPDM_{v_i}$, is computed as the barycentric interpolation of the three local perceptual distances $LPD_{v_i, v'_{i,1}}$, $LPD_{v_i, v'_{i,2}}$ and $LPD_{v_i, v'_{i,3}}$.

3.4. Global perceptual distance

The global tensor-based perceptual distance measure $TPDM$ from the reference mesh \mathcal{M}_r to the distorted mesh \mathcal{M}_d is computed as the *weighted Minkowski sum* of the local distances $LTPDM_{v_i}$, $i = 1, 2, \dots, N$:

$$TPDM = \left(\sum_{i=1}^N w_i |LTPDM_{v_i}|^p \right)^{\frac{1}{p}}, \quad (6)$$

where $w_i = s_i / \sum_{i=1}^N s_i$ with s_i one third of the total area of all the incident facets of v_i , and $p = 2.5$. The surface-based weighting can, to some extent, enhance the stability of

the metric to the variation of vertex sampling density over the mesh surface. Compared to the standard mean-squared error where $p = 2.0$, the choice of $p = 2.5$ can increase the importance of the local high amplitude distances of in the calculation of the global perceptual distance [15]. This is perceptually relevant, since the part of mesh with high-amplitude distortion experimentally attracts more attention from human observers, and thus has more impact on the result of subjective assessment.

4. Experimental Results

4.1. Performance evaluation and comparisons

In order to verify its efficacy, the proposed metric *TPDM* has been extensively tested and compared with the existing metrics on four subjectively-rated visual mesh quality databases:

- The LIRIS/EPFL general-purpose database¹ [17]: Contains 4 reference meshes and the total of 84 distorted models. The distortion types include noise addition and smoothing, applied either locally or globally to the reference mesh. Subjective evaluations were made by 12 observers.
- The LIRIS masking database² [26]: Contains 4 reference meshes and the total of 24 distorted models. The local noise addition distortion included in this database was designed specifically for testing the capability of MVQ metrics to capture the visual masking effect. 11 observers participated in the subjective tests.
- The IEETA simplification database³ [28]: Contains 5 reference meshes and the total of 30 simplified models. 65 observers participated in the subjective study.
- The UWB compression database⁴ [37]: Contains 5 reference meshes and the total of 64 distorted models. Subjective evaluations were made by 69 observers.

TPDM has been compared with seven state-of-the-art metrics, i.e., the Hausdorff distance (*HD*) [6, 11], the root mean squared error (*RMS*) [6, 11], *3DWPM*₁ and *3DWPM*₂ [22], *MSDM*₂ [32], *DAME* [37] and *FMPD* [38]. The coherence between the objective values produced by the MVQ metrics and the mean opinion scores (*MOS*) provided by subjective databases is measured using two different correlation kinds: The Pearson linear correlation coefficient (*PLCC* or r_p), which measures the prediction accuracy of the objective metrics, and the Spearman rank-order correlation coefficient (*SROCC* or r_s), which measures the prediction monotonicity [19, 34]. Before computing the correlation values, especially the *PLCC*, it is recommended to conduct a psychometric fitting between the objective scores and the *MOS* values, in order to partially

¹<http://liris.cnrs.fr/guillaume.lavoue/data/datasets.html>

²<http://liris.cnrs.fr/guillaume.lavoue/data/datasets.html>

³<http://www.ieeta.pt/~sss/index.php/perceivedquality/repository>

⁴<http://compression.kiv.zcu.cz/>

remove the non-linearity between them. Another effect of psychometric fitting is that afterwards we obtain objective MVQ values belonging to the $[0, 1]$ interval, which are easier for the users to understand. In our tests, we apply a cumulative Gaussian psychometric function [7] for the fitting:

$$g(a, b, R) = \frac{1}{\sqrt{2\pi}} \int_{a+bR}^{\infty} e^{-(t^2/2)} dt, \quad (7)$$

where R is the raw *TPDM* value. The two parameters $a = -1.14$ and $b = 11.47$ are obtained through a non-linear least squares fitting (under Matlab, with the curve fitting toolbox) using the raw *TPDM* values and the corresponding *MOS* for the group of Dinosaur models in the general-purpose database. As shown in Fig. 6, the same psychometric function is used for models in other subjective databases. From the plots we can see that the fitted psychometric function has a good generalization capability for other databases, as the psychometric curve is close to the *TPDM-MOS* pairs.

The tables 2 to 5 present, respectively, the evaluation and comparison results for the general-purpose, masking, simplification and compression databases. Analogously to the comparisons in [39], the overall correlations (last two columns in the tables) are computed for the whole database (i.e., we compute the statistical linear or non-linear dependence between all objective scores and *MOS* values of all models in each database), except the compression database, where per-model averages are used. This is because the data acquisition procedure of the compression database does not take into account inter-model coherence [37, 39]. The results for the existing metrics shown in Tabs. 2 to 5 are either extracted from published papers [31, 37–39] and the related erratum [40], or collected from our own tests.

The geometric distance metrics *HD* and *RMS* in general fail to evaluate the perceptual mesh quality: The overall Pearson and Spearman correlations are quite low for all the four databases. *HD* even results in a negative correlation on some difficult models, i.e., *Jessy* and *James* from the compression database (cf. Tab. 5). In the following we will focus on the comparison of our approach with state-of-the-art perceptually-driven MVQ metrics for each database.

For the general-purpose database (cf. Tab. 2), *TPDM* has high *PLCC* and *SROCC* values for almost every individual model, as well as for the whole repository. *TPDM* has the highest overall *PLCC* and *SROCC* values among all the MVQ metrics tested (the last two columns). In particular, there is a noticeable improvement in terms of overall correlations compared to the second best metric *FMPD*. For example, the overall *SROCC* has improved from 81.9% for *FMPD* to 89.6% for *TPDM*. Since the general-purpose database has the highest number of distorted models among the four available databases, as well as a variety of distortion types, the high correlation values of *TPDM* for this database appear to be promising evidence for the good performance of *TPDM* in assessing visual mesh quality.

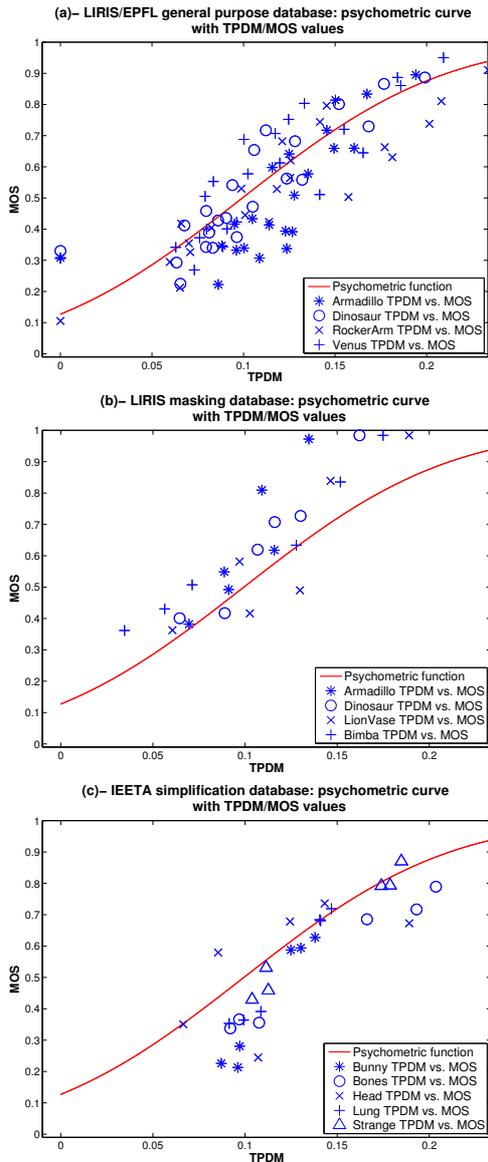


Fig. 6. Psychometric function curve plotted with *TPDM-MOS* pairs of all the reference and distorted models in: (a) the LIRIS/EPFL general-purpose database; (b) the LIRIS masking database; and (c) the IEETA simplification database.

Tab. 2. $PLCC$ (r_p) and $SROCC$ (r_s) (%) of different objective metrics on the general-purpose database.

Metric	Armadillo		Dinosaur		RockerArm		Venus		All models	
	r_p	r_s								
<i>HD</i> [6, 11]	54.9	69.5	47.5	30.9	23.4	18.1	8.9	1.6	11.4	13.8
<i>RMS</i> [6, 11]	56.7	62.7	0.0	0.3	17.3	7.3	87.9	90.1	28.1	26.8
<i>3DWPM</i> ₁ [22]	59.7	65.8	59.7	62.7	72.9	87.5	68.3	71.6	61.9	69.3
<i>3DWPM</i> ₂ [22]	65.6	74.1	44.6	52.4	54.7	37.8	40.5	34.8	49.6	49.0
<i>MSDM</i> ₂ [32]	85.3	81.6	85.7	85.9	87.2	89.6	87.5	89.3	81.4	80.4
<i>DAME</i> [37]	76.3	60.3	88.9	92.8	80.1	85.0	83.9	91.0	75.2	76.6
<i>FMPD</i> [38]	83.2	75.4	88.9	89.6	84.7	88.8	83.9	87.5	83.5	81.9
<i>TPDM</i>	78.8	84.5	89.0	92.2	91.4	92.2	91.0	90.6	86.2	89.6

Tab. 3. $PLCC$ (r_p) and $SROCC$ (r_s) (%) of different objective metrics on the masking database.

Metric	Armadillo		Bimba		Dinosaur		LionVase		All models	
	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s
<i>HD</i> [6, 11]	61.4	48.6	27.4	25.7	55.8	48.6	50.1	71.4	20.2	26.6
<i>RMS</i> [6, 11]	66.8	65.7	46.7	71.4	70.9	71.4	48.8	71.4	41.2	48.8
<i>3DWPM</i> ₁ [22]	64.6	58.0	29.0	20.0	67.3	66.7	31.1	20.0	31.9	29.4
<i>3DWPM</i> ₂ [22]	61.6	48.6	37.9	37.1	70.8	71.4	46.9	38.3	42.7	37.4
<i>MSDM</i> ₂ [32]	81.1	88.6	96.8	100	95.6	100	93.5	94.3	87.3	89.6
<i>DAME</i> [37]	96.0	94.3	88.0	97.7	89.4	82.9	99.5	100	58.6	68.1
<i>FMPD</i> [38]	94.2	88.6	98.9	100	96.9	94.3	93.5	94.3	80.8	80.2
<i>TPDM</i>	91.4	88.6	97.2	100	97.1	100	88.4	82.9	88.6	90.0

From the results for the masking database (cf. Tab. 3), we can see that in general *TPDM* captures the visual masking effect well, as reflected by the high individual and overall $PLCC$ and $SROCC$ values for that database. The same applies to the results for the general-purpose database: *TPDM* has the highest overall $PLCC$ and $SROCC$ values among all the metrics tested. Another observation is that although the reduced-reference metric *FMPD* performs quite well for each individual model, its overall $PLCC$ and $SROCC$ values are not that high when compared to the full-reference metrics *MSDM*₂ and *TPDM*. Hence, it seems that in order to capture the visual masking effect well in the case of MVQ assessment, it would be advantageous to conduct a precise vertex-to-vertex local analysis, with the availability of the full information about the reference mesh.

Compared to other connectivity-independent metrics (cf. Tab. 1), *TPDM* has comparable performance with *MSDM*₂ and *FMPD* for the simplification database (cf. Tab. 4). The results of *3DWPM*₁, *3DWPM*₂ and *DAME* are missing because those metrics have the mesh connectivity constraint, and therefore cannot be applied to compare two meshes with different connectivities. For the Head model, the correlation of *TPDM* is rather low. This is because *TPDM* has difficulties in distinguishing the quality of simplified Head meshes generated by different simplification algorithms but with the

Tab. 4. $PLCC$ (r_p) and $SROCC$ (r_s) (%) of different objective metrics on the simplification database.

Metric	Bones		Bunny		Head		Lung		Strange		All models	
	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s
<i>HD</i> [6, 11]	92.0	94.3	37.8	39.5	72.8	88.6	80.6	88.6	52.3	37.1	50.5	49.4
<i>RMS</i> [6, 11]	86.4	94.3	94.5	77.1	49.6	42.9	89.0	100	90.4	88.6	59.6	70.2
<i>MSDM2</i> [32]	98.3	94.3	98.1	77.1	88.9	88.6	92.3	60.0	99.0	94.3	89.2	86.7
<i>FMPD</i> [38]	96.0	88.6	98.0	94.3	70.4	65.7	95.5	88.6	96.0	65.7	89.3	87.2
<i>TPDM</i>	99.0	94.3	98.0	94.3	63.1	65.7	98.6	94.3	98.7	94.3	86.9	88.2

Tab. 5. $PLCC$ (r_p) and $SROCC$ (r_s) (%) of different objective metrics on the compression database.

Metric	Bunny		James		Jessy		Nissan		Helix		All models	
	r_p	r_s										
<i>HD</i> [6, 11]	34.1	52.2	-16.8	6.8	-23.6	12.5	14.4	23.6	45.1	46.4	10.6	28.3
<i>RMS</i> [6, 11]	34.2	20.9	14.0	10.8	0.0	14.8	17.8	29.7	46.9	44.6	22.0	24.1
<i>3DWPM₁</i> [22]	94.7	93.4	77.3	72.3	87.2	89.5	63.6	59.3	98.0	95.2	84.1	81.9
<i>3DWPM₂</i> [22]	96.0	91.2	76.9	65.3	86.9	85.9	56.3	67.6	95.5	94.3	82.3	80.9
<i>MSDM2</i> [32]	97.4	90.1	82.6	69.2	84.3	63.1	84.4	73.1	98.1	94.7	89.3	78.0
<i>DAME</i> [37]	96.8	93.4	95.7	93.4	84.4	70.5	93.9	75.3	96.6	95.2	93.5	85.6
<i>FMPD</i> [38]	94.2	89.6	95.3	91.2	63.3	60.0	92.4	77.5	98.4	90.8	88.8	81.8
<i>TPDM</i>	95.1	96.5	90.8	73.6	85.8	75.8	82.7	73.4	98.7	95.0	91.5	82.9

same vertex reduction ratio. *FMPD* is more or less affected by the same problem. The simplification database is a relatively simple dataset. *MSDM2*, *FMPD* and *TPDM* all have very high overall correlation on this database, and even *HD* and *RMS* have very good performance on some individual models. This observation implies the necessity of constructing a comprehensive subjectively-rated MVQ database which incorporates more models and more types of lossy operations affecting mesh connectivity, e.g., other simplification algorithms, remeshing or even subdivision.

For the compression database (cf. Tab. 5), *DAME* has the highest overall $PLCC$ and $SROCC$: 93.5% and 85.6% for *DAME* against respectively 91.5% and 82.9% for *TPDM*, the second best metric for this database. Initially, we encountered difficulties in testing *TPDM* on the James, Jessy and Nissan models, which consist of many spatially non-connected components (70, 138 and 212, respectively). Some of the components have null Euclidean distance to each other, so that the vertex matching preprocessing fails since a vertex may be incorrectly projected onto a different nearby component, even without any introduced distortion. A simple solution has been adopted to resolve this problem. We first identify such incorrectly projected vertices on the reference mesh by conducting the *TPDM* comparison between the reference mesh and itself. The incorrectly projected vertices are those with non-zero *LTPDM* values. When performing a

comparison with a distorted mesh, the *LTPDM* distances for these vertices are deduced from the distance values of its 1-ring neighbors, via simple median filtering. This also demonstrates that in order to conduct an effective MVQ assessment (especially for the development of full-reference MVQ metrics), it is important to develop a robust, or ideally perceptually-driven, mesh correspondence algorithm. The development of such an algorithm remains an open research problem.

All in all, *TPDM* shows quite good performance on all the available subjectively-rated mesh visual quality databases, as reflected by its high correlation with subjective scores on most individual models, as well as on the whole repositories. In particular, *TPDM* has the highest overall *SROCC* (the last column in the tables) on the general-purpose, masking and simplification databases, and it is the second best performing metric on the compression database. Furthermore, *TPDM* has always higher *SROCC* than *MSDM2* on all the four databases, and also under three cases out of four higher *PLCC* values (the exception is the simplification database). It appears that the injection of the information on the principal surface directions helps improve the MVQ assessment performance.

TPDM also allows us to obtain a perceptually coherent distance map between two meshes. Figure 7 illustrates the distance maps produced by *TPDM* and *RMS* between the original Bimba model and a distorted Bimba after uniform random noise addition. The map of *TPDM* is quite consistent with human perception (i.e., the perceived distortion is higher in smooth regions than in rough regions), while the map of *RMS* is purely geometric and fails to capture the visual masking effect.

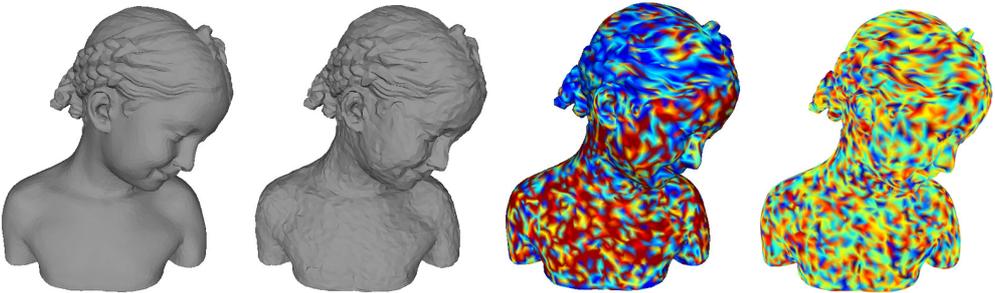


Fig. 7. From left to right: the original Bimba model, the distorted Bimba model after uniform random noise addition, the distance map of *TPDM* between the two meshes, and the distance map of *RMS*. In the distances maps, warmer colors represent higher local distance values.

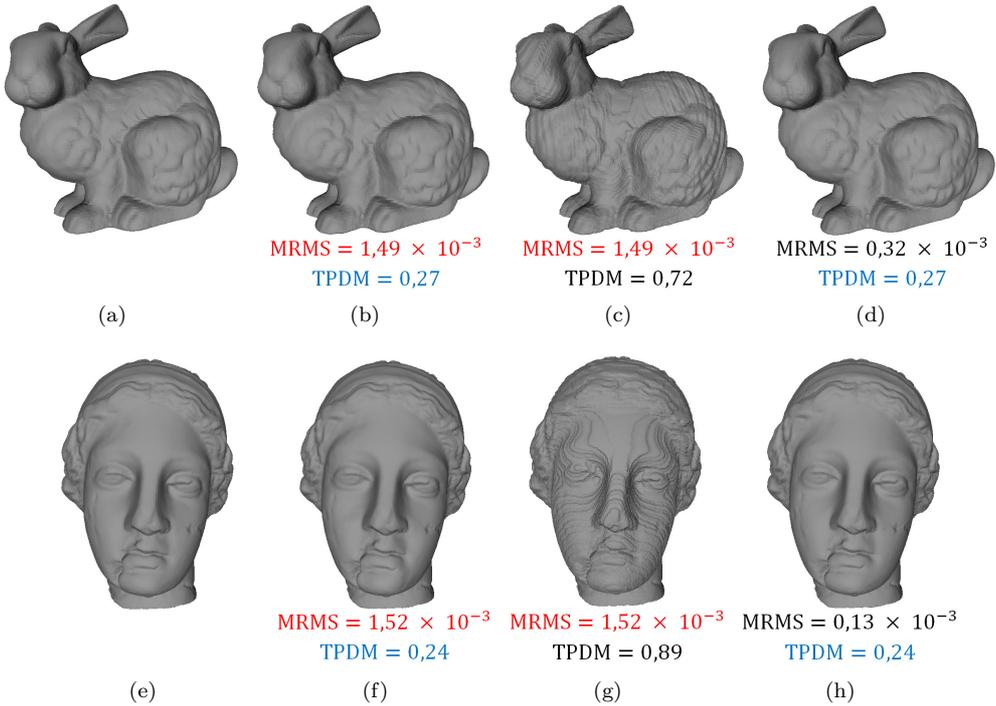


Fig. 8. Application of *TPDM* to the perceptual evaluation of robust mesh watermarking: (a) original Bunny model; (b) Bunny watermarked by using the method in [33]; (c) and (d) models watermarked by the method in [21] giving, respectively, the same *MRMS* and *TPDM* distortion as (b); (e) original Venus model; (f) Venus watermarked by the method in [33]; (g) and (h) models watermarked by the method in [21], giving, respectively, the same *MRMS* and *TPDM* distortion as (f). The *TPDM* values are those obtained after psychometric fitting.

4.2. Applications

In this subsection, we show two simple examples to illustrate the potential of *TPDM* in practical mesh applications. The two examples concern, respectively, the quality evaluation of watermarked meshes and the optimum quantization level selection for mesh vertex coordinates.

Figure 8 shows the potential application of our metric *TPDM* in the visual quality assessment of watermarked meshes or, more generally, in the benchmarking of robust mesh watermarking algorithms. Indeed, when comparing two robust watermarking algorithms, a common strategy is to first of all fix the amount of distortion induced by watermark embedding, and then compare the robustness of watermarks against a series

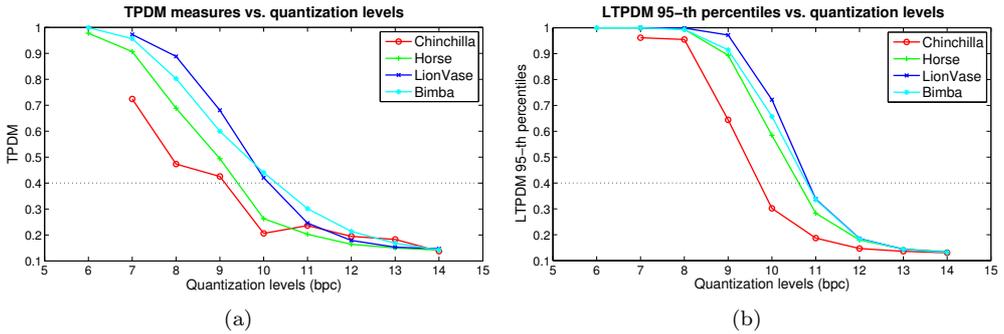


Fig. 9. Application of $TPDM$ for the selection of optimum quantization level of mesh vertex coordinates: (a) $TPDM$ vs. quantization levels, with threshold $\tau_{TPDM} = 0.40$; and (b) 95-th percentiles of $LTPDM$ vs. quantization levels, with threshold $\tau_{LTPDM} = 0.40$.

of attacks, such as noise addition, smoothing and simplification. The question comes to how to quantify the induced distortion. We consider that using classical mesh geometric distances (e.g., maximum root mean squared error, $MRMS$ [6, 11]) is not appropriate, at least when the watermarked meshes are used in applications having human beings as users. In Figs. 8(b) and 8(f) (second column of the figure) we show two watermarked models generated by the method of Wang et al. [33]. We show in Figs. 8(c) and 8(g) (third column) the corresponding watermarked models produced by the method of Cho et al. [21] that have exactly the same $MRMS$ as Wang et al.'s models. Despite having the same $MRMS$ values, the watermarked models produced by the two methods are of significantly different visual quality, and so it would be unfair to conduct robustness comparison on these models. In contrast, the comparison would be fair enough, at least for applications where the visual quality of watermarked mesh is very important, if we fix the amount of $TPDM$ distortion induced by the two methods. We show in Figs. 8(d) and 8(h) (last column) the watermarked models produced by the method of Cho et al. [21] that have exactly the same $TPDM$ values as the models of Wang et al. The watermarked meshes are of comparable and good visual quality.

The second potential application of $TPDM$ shown here is automatic selection of the optimum quantization level of mesh vertex coordinates, defined as the minimum number of bits allocated for representing each quantized coordinate that does not introduce unacceptable visual distortion. Vertex coordinate quantization is almost a mandatory step in lossy mesh compression, but the selection of optimum quantization level is in general mesh-dependent and tedious, which often requires efforts of human observers. $TPDM$ could help us facilitate this task. For this purpose, we introduce two kinds of thresholds, as explained below. The first threshold τ_{TPDM} is defined on the global $TPDM$ value, which guarantees good global visual quality of the mesh after quantization.

In Fig. 9(a) we plot $TPDM$ values versus quantization levels for four meshes of different geometric complexity, and the threshold (i.e., the maximum allowable $TPDM$ introduced by quantization) is fixed as 0.40. The second threshold τ_{LTPDM} is defined on the local tensor-based perceptual distance measures, so as to ensure precise control of the local distortion. For this local control, we first compute the 95-th percentile of the $LTPDM$ distances after quantization (i.e., the value below which 95 percent of $LTPDM$ distances may be found in a quantized mesh), and then compare this percentile value with τ_{LTPDM} . In Fig. 9(b) we show the 95-th percentiles of $LTPDM$ versus quantization levels, and the threshold τ_{LTPDM} is fixed as 0.40, the same value as τ_{TPDM} . Compared to the global threshold, the local threshold appears to be a more strict metric for the control of introduced distortion. We show the quantization results in Fig. 10. For Chinchilla and Bimba, both global and local thresholds select the same level as the optimum, i.e., 10 *bpc* (*bits per coordinate*) for Chinchilla and 11 *bpc* for Bimba, which are consistent with human perception. For Horse, after applying the global threshold, a relatively low quantization level 10 *bpc* is selected, which results in a mesh of rather globally acceptable visual quality, but with some high-amplitude local distortions (especially on the head); the local threshold is stricter and selects 11 *bpc* as the optimum, so as to avoid visually unacceptable local distortion. Another remark is that $TPDM$ may result in unstable assessment results for meshes with relatively few vertices. As shown in Fig. 9(a), for Chinchilla, which has 4307 vertices, the distorted mesh obtained after 10-bit quantization is of better quality than the mesh obtained after 11-bit quantization, according to the $TPDM$ values. Although the two meshes are of rather comparable visual quality (cf. Figs. 10(b) and 10(c)), we think that this is a drawback of the proposed metric, and an improvement on this point will be part of our future work.

5. Conclusion and Future Work

A new curvature-tensor-based approach to objective evaluation of visual mesh quality has been proposed. We show that it is beneficial to use the information on both the curvature amplitudes and the principal curvature directions for MVQ assessment. The local tensor distance that we propose may be found useful in other mesh applications, such as mesh segmentation and shape matching. Experimental results show that our $TPDM$ metric has high correlation with subjective scores and performance comparable with the best performing MVQ metrics proposed so far. Finally, two simple examples illustrate the potential applications of the proposed metric.

$TPDM$ implementation is freely available on-line at <http://www.gipsa-lab.fr/~fakhri.torkhani/software/TPDM.rar>. Future work will consist mainly in integrating more HVS features into the metric (e.g., the contrast sensitivity function), extending the metric to perceptual evaluation of triangular meshes with photometric properties (i.e., with color and textures), and developing a curvature-tensor-based visual quality

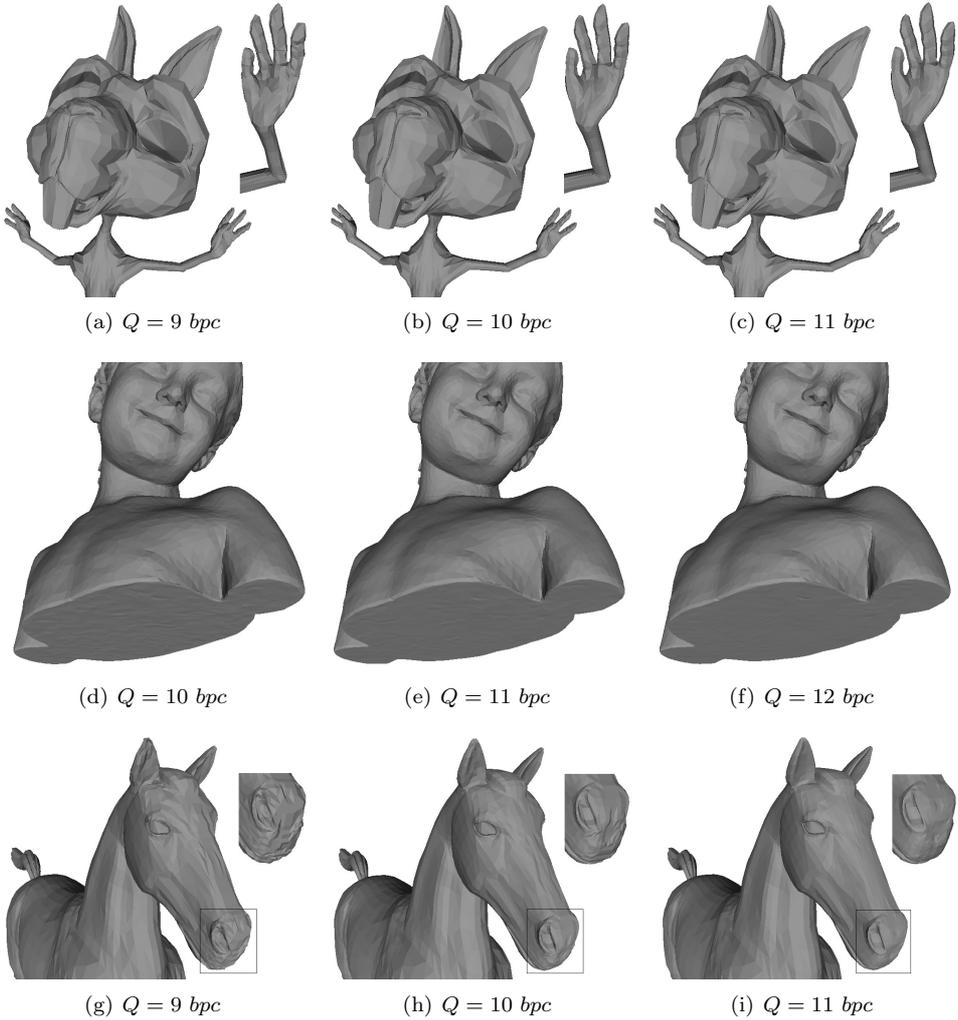


Fig. 10. Application of *TPDM* to the selection of optimum quantization level of mesh vertex coordinates. For Chinchilla (first row, (a)-(c)), both global and local thresholds select 10 *bpc* as the optimum level (Q in bits per coordinate). For Bimba (second row, (d)-(f)), both global and local thresholds select 11 *bpc* as the optimum level. For Horse (third row, (g)-(i)), the global threshold selects 10 *bpc* as the optimum level while the local threshold selects 11 *bpc* as the optimum level.

metric for dynamic meshes. In particular, for perceptual evaluation of dynamic meshes, it would be interesting to derive a spatial-temporal perceptually-oriented curvature tensor distance that accounts for both the spatial visual masking effect (as shown in this paper) and the temporal visual masking effect due to the movement of the 3D mesh.

References

- 1927**
[1] A. Michelson. *Studies in Optics*. Univ. of Chicago Press.
- 1969**
[2] H. S. M. Coxeter. *Introduction to Geometry, 2nd Ed.* Wiley.
- 1985**
[3] M. Brady, J. Ponce, A. L. Yuille, and H. Asada. Describing surfaces. *Comput. Vision, Graphics, and Image Process.*, 32(1):1–28.
- 1993**
[4] H. Kukkonen, J. Rovamo, K. Tiippana, and R. Näsänen. Michelson contrast, RMS contrast and energy of various spatial stimuli at threshold. *Vision Research*, 33(10):1431–1436.
- 1995**
[5] M. C. Morrone, D. C. Burr, and L. M. Vaina. Two stages of visual processing for radial and circular motion. *Nature*, 376(6540):507–509.
- 1998**
[6] P. Cignoni, C. Rocchini, and R. Scopigno. Metro: measuring error on simplified surfaces. *Comput. Graphics Forum*, 17(2):167–174.
- 2000**
[7] P. G. Engeldrum. *Psychometric Scaling: A Toolkit for Imaging Systems Development*. Imcotek Press.
[8] A. Hertzmann and D. Zorin. Illustrating smooth surfaces. In *Proc. of ACM Siggraph*, pages 517–526.
[9] Z. Karni and C. Gotsman. Spectral compression of mesh geometry. In *Proc. of ACM Siggraph*, pages 279–286.
- 2001**
[10] B.-E. Rogowitz and H.-E. Rushmeier. Are image quality metrics adequate to evaluate the quality of geometric objects. In *Proc. of Human Vision and Electronic Imaging*, pages 340–348.
- 2002**
[11] N. Aspert, D. Santa-Cruz, and T. Ebrahimi. MESH: measuring errors between surfaces using the Hausdorff distance. In *Proc. of IEEE Int. Conf. on Multimedia & Expo*, pages 705–708.
- 2003**
[12] P. Alliez, D. Cohen-Steiner, O. Devillers, B. Lévy, and M. Desbrun. Anisotropic polygonal remeshing. *ACM Trans. on Graphics*, 22(3):485–493.
[13] D. Cohen-Steiner and J. M. Morvan. Restricted Delaunay triangulations and normal cycle. In *Symp. on Computational Geometry*, pages 312–321.
[14] O. Sorkine, D. Cohen-Or, and S. Toledo. High-pass quantization for mesh encoding. In *Proc. of Eurographics/ACM Siggraph Symp. on Geometry Processing*, pages 42–51.
[15] Z. Wang and X. Shang. Spatial pooling strategies for perceptual image quality assessment. In *Proc. of IEEE Int. Conf. on Image Process.*, pages 2945–2948.
- 2004**
[16] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. on Image Process.*, 13(4):600–612.
- 2006**
[17] G. Lavoué, E. Drelie Gelasca, F. Dupont, A. Baskurt, and T. Ebrahimi. Perceptually driven 3D distance metrics with application to watermarking. In *Proc. of SPIE Electronic Imaging*, pages 63120L.1–63120L.12.

- [18] O. Sorkine. Differential representations for mesh processing. *Comput. Graphics Forum*, 25(4):789–807.
- [19] Z. Wang and A. C. Bovik. *Modern Image Quality Assessment*. Morgan & Claypool.
2007
- [20] B. G. Breitmeyer. Visual masking: past accomplishments, present status, future developments. *Advances in Cognitive Psychology*, 3(1-2):9–20.
- [21] J.-W. Cho, R. Prost, and H.-Y. Jung. An oblivious watermarking for 3-D polygonal meshes using distribution of vertex norms. *IEEE Trans. on Signal Process.*, 55(1):142–155.
- [22] M. Corsini, E. Drelie Gelasca, T. Ebrahimi, and M. Barni. Watermarked 3-D mesh quality assessment. *IEEE Trans. on Multimedia*, 9(2):247–256.
2008
- [23] F. Cole, A. Golovinskiy, A. Limpaecher, H.-S. Barros, A. Finkelstein, T. Funkhouser, and S. Rusinkiewicz. Where do people draw lines? *ACM Trans. on Graphics*, 27(3):88:1–88:11.
- [24] B. Vallet and B. Lévy. Spectral geometry processing with manifold harmonics. *Comput. Graphics Forum*, 27(2):251–260.
2009
- [25] Z. Bian, S.-M. Hu, and R. R. Martin. Evaluation for small visual difference between conforming meshes on strain field. *J. of Comput. Sci. and Technol.*, 24(1):65–75.
- [26] G. Lavoué. A local roughness measure for 3D meshes and its application to visual masking. *ACM Trans. on Appl. Perception*, 5(4):21:1–21:23.
- [27] Q. Li and Z. Wang. Reduced-reference image quality assessment using divisive normalization-based image representation. *IEEE J. Sel. Topics Signal Process.*, 3(2):202–211.
- [28] S. Silva, B. S. Santos, C. Ferreira, and J. Madeira. A perceptual data repository for polygonal meshes. In *Proc. of Int. Conf. in Visualization*, pages 207–212.
- [29] Z. Wang and A. C. Bovik. Mean squared error: love it or leave it? — A new look at signal fidelity measures. *IEEE Signal Process. Magazine*, 26(1):98–117.
2010
- [30] M. Botsch, L. Kobbelt, M. Pauly, P. Alliez, and B. Lévy. *Polygon Mesh Processing*. AK Peters.
- [31] G. Lavoué and M. Corsini. A comparison of perceptually-based metrics for objective evaluation of geometry processing. *IEEE Trans. on Multimedia*, 12(7):636–649.
2011
- [32] G. Lavoué. A multiscale metric for 3D mesh visual quality assessment. *Comput. Graphics Forum*, 30(5):1427–1437.
- [33] K. Wang, G. Lavoué, F. Denis, and A. Baskurt. Robust and blind mesh watermarking based on volume moments. *Comput. & Graphics*, 35(1):1–19.
- [34] Z. Wang and A. C. Bovik. Reduced- and no-reference image quality assessment. *IEEE Signal Process. Magazine*, 28(6):29–40.
2012
- [35] P. Alliez, S. Tayeb, and C. Wormser. 3D fast intersection and distance computation (AABB tree). In *CGAL User and Reference Manual*.
- [36] F. Torkhani, K. Wang, and J.-M. Chassery. A curvature tensor distance for mesh visual quality assessment. In *Proc. of Int. Conf. on Computer Vision and Graphics*, pages 253–263.
- [37] L. Váša and J. Rus. Dihedral angle mesh error: a fast perception correlated distortion measure for fixed connectivity triangle meshes. *Comput. Graphics Forum*, 31(5):1715–1724.
- [38] K. Wang, F. Torkhani, and A. Montanvert. A fast roughness-based approach to the assessment of 3D mesh visual quality. *Comput. & Graphics*, 36(7):808–818.
2013
- [39] M. Corsini, M. C. Larabi, G. Lavoué, O. Petřík, L. Váša, and K. Wang. Perceptual metrics for static and dynamic triangle meshes. *Comput. Graphics Forum*, 32(1):101–125. (Improved version of Eurographics 2012 state-of-the-art report).
- [40] G. Lavoué. Erratum of the Results of Mesh Visual Quality Metrics (Available on-line at <http://liris.cnrs.fr/glavoue/travaux/Erratum.html>).

BAG OF WORDS – QUALITY ISSUES OF NEAR-DUPLICATE IMAGE RETRIEVAL

Mariusz Paradowski, Mariusz Durak, Bartosz Broda
Institute of Informatics, Wrocław University of Technology, Poland

Abstract. This paper addresses the problem of large scale near-duplicate image retrieval. Issues related to visual words dictionary generation are discussed. A new spatial verification routine is proposed. It incorporates neighborhood consistency, term weighting and it is integrated into the Bhattacharyya coefficient. The proposed approach reaches almost 10% higher retrieval quality, comparing to other recently reported state-of-the-art methods.

Key words: spatial verification, vector space model, visual words, clustering.

1. Introduction

In recent ten years lots of research effort has been put to the problem of large scale near-duplicate image retrieval. Databases containing thousands or even millions of images are successfully processed by efficient retrieval algorithms. The basic and well known technique for efficient near-duplicate retrieval is *vector space model* [1, 5, 15]. Feature representation used in vector space model is called *bag of words* (BoW). Bag of words is a histogram (one-dimensional in most cases) representing a single image and constructed from many *visual words* [5]. A *visual word* is a single feature vector compressed to a single numeric value (group, cluster). Image similarity measurement is usually done by histogram comparison. Application of various histogram similarity measures [9] is possible. Efficient image retrieval systems are constructed on top an of inverted index structure [8]. Data related to visual words is stored in *inverted index files*. Files are read on demand, thus memory usage is kept low. Computational complexity of the retrieval process is less than $O(n)$ per image pair, where n is the number of visual words of the image.

1.1. Research background

Plain bag of words histogram comparison is not sufficient in many cases. Visual words context has to be taken into account (similarly as in Natural Language Processing) using spatial (context) analysis. Spatial analysis methods come in two major categories: global and local. Global methods find image transformations (usually affine or perspective) between images. Major approaches are: RANSAC [2] based methods for single transformation detection (e.g. [18]), Hough transformation for multiple object detection

(e.g. [7, 17]) and non-linear approaches (e.g. [16]). Local methods usually verify visual word neighborhood consistency, e.g. [4].

Spatial image analysis is a time consuming process, much slower comparing to simple histogram comparison. An efficient, spatially enforced retrieval scheme is hierarchical. First, a complete database of images is processed using histogram similarity measurement. Later on, a subset of best results is analyzed using spatial methods. In the paper we address both issues: efficient histogram comparison and spatial analysis of images.

1.2. Contribution

The contribution of the paper is the following. We propose an extended version of simple spatial verification routine [4]. We show that the commonly used *cosine* bag of words similarity measure [10, 18] is outperformed by *Bhattacharyya coefficient* and χ^2 distances. We discuss bag of words dictionary generation. We point out a problem in a commonly used experimental protocol.

The paper is organized as follows. The second section presents the complete retrieval scheme, together with our proposals. The third section demonstrates the experiments performed. The last section concludes our work.

2. Image retrieval method

The discussed retrieval method is based on the *vector space model* [1]. The vector space model assumes the existence of discrete terms (words), which describe both the database and the query. The model has been originally used in Natural Language Processing, where terms are naturally related to words or concepts. The model can be effectively used in image processing [5]. However the existence of discrete terms is not obvious. *Visual terms* also known as *visual words* have to be generated from visual features, which are usually continuous in nature. The process of visual terms generation can be modeled as a grouping problem and performed using variants of *k-means* method. Representation of a single image consists of many high dimensional feature vectors (key points, key regions) [6]. Each vector is converted into a visual term and thus the image representation becomes a bag of words.

One of the elementary issues in vector space model is a proper construction of *visual term dictionary*. Construction of the dictionary is an unsupervised grouping problem. There are many grouping approaches, but classic *k-means* have gained most popularity. Application of *k-means* is not accidental – it minimizes the total distance between cluster centers and data points. To speed up the grouping routine some researches have applied various modifications of *k-means*, including: *hierarchical k-means* and *approximate k-means* [10]. We revert to the original *k-means* with much success.

Bag of words representing a single image may be defined both in terms of vectors

and probability distributions, thus both vector similarity and PDF similarity measures may be used. For efficient retrieval the similarity function has to be calculated on sparse vector representation in $O(n+m)$ computational complexity, where n and m are number of non-zero elements in sparse *BoW* vectors. Only a subset of similarity measures follow this criterion. The widely recognized and used one is the *cosine similarity*:

$$\cos(x, y) = \sum_{i \in V} x_i y_i, \quad (1)$$

where: x and y are L2 normalized vectors representing weighted image BoW's, V is the set of visual words, x_i and y_i are BoW values of the i -th bin.

After the generation of the initial bag of words ranking, spatial verification takes place. It is a key issue in successful near-duplicate retrieval [10]. There are many approaches to spatial verification, including:

1. geometry-based approaches that reconstruct various transformations between two images, e.g., RANSAC [2, 10], Hough Transform [17];
2. topology-based approaches focused on various local pseudo-invariants, e.g. [3, 4].

Further post-processing methods are also available, such as *query expansion*. They are outside the scope of our research and we do not address them in the paper. However, it is worth noting that the presented approach is compatible with these routines.

2.1. Alternative bag of words similarity measurement

Let us first address the similarity measurement of two images in a vector space model. Many similarity measures may be used in the vector space model [9]. We have analyzed several of them. Two of them proven to be worth of interest.

A measure that is rarely used in image retrieval, but has proven to be interesting, is *Bhattacharyya coefficient* of two probability distributions. In this case each *BoW* representation becomes a discrete PDF (L1 normalized vector). Distribution similarity is defined as follows:

$$BC(x, y) = \sum_{i \in V} \sqrt{x_i y_i}, \quad (2)$$

Another measure worth noting is χ^2 histogram distance. There are several variants of χ^2 [9], we choose the symmetric one:

$$\chi^2(x, y) = \sum_{i \in V} \frac{(x_i - y_i)^2}{x_i + y_i}. \quad (3)$$

Both of the above similarity measures can be used in inverted-index retrieval approach. Usage of $BC(x, y)$ is straightforward, because only corresponding non-zero elements influence the final result. χ^2 distance integration is slightly more difficult. Let

us assume that x is the query image. In such case $V_x \subset V$ is the set of visual words belonging to x through which we iterate:

$$\begin{aligned} \chi^2(x, y) &= \sum_{i \in V} \frac{(x_i - y_i)^2}{x_i + y_i} = \sum_{i \in V} \frac{x_i^2}{x_i + y_i} + \sum_{i \in V} \frac{y_i^2}{x_i + y_i} - 2 \sum_{i \in V} \frac{x_i y_i}{x_i + y_i} = \\ &= \sum_{i \in V_x} \frac{x_i^2}{x_i + y_i} + \left(\sum_{i \in V} \frac{y_i^2}{y_i} - \sum_{i \in V_x} \frac{y_i^2}{y_i} + \sum_{i \in V_x} \frac{y_i^2}{x_i + y_i} \right) - 2 \sum_{i \in V_x} \frac{x_i y_i}{x_i + y_i}. \end{aligned} \quad (4)$$

The above equation shows that only one component per database image has to be pre-computed ($\sum_{i \in V} y_i$). All other are either equal to 0 or may be calculated only from $V_x \subset V$, during inverted index lookup.

2.2. Proposed spatial validation routine

Let us now describe the main research contribution of this paper. The proposed method is a spatial validation routine. The basic idea of the approach may be traced back to early works of Mohr and Schmid [3, 4], later incorporated into the vector space model by Sivic [5]. The main idea is to check spatial consistency of neighboring pairs. The higher the spatial consistency the better. We extend this idea using other well known techniques in the following ways:

1. neighboring pairs are *tf-idf* weighted instead of simple counting to incorporate their importance in vector space model (e.g. [12]),
2. neighborhood size is dynamic and is relative to key points size ratio,
3. neighboring pairs consistency is relative to key point size ratio and key point distance ratio,
4. the routine is integrated into *Bhattacharyya coefficient* giving a re-weighted BoW histogram.

First let us define a standard *tf-idf* weighted *Bhattacharyya coefficient*:

$$BC_{tfidf}(x, y) = \sum_{i \in V} \sqrt{tf(x_i) \cdot idf_i \cdot tf(y_i) \cdot idf_i}. \quad (5)$$

Term frequencies $tf(x_i)$ and $tf(y_i)$ represent histogram bin data. In spatial verification routine they have to be replaced by the neighborhood consistency function $SV(x_i, y_i)$:

$$BC_{SV}(x, y) = \sum_{i \in V} \sqrt{idf_i^2 \cdot SV(x_i, y_i)}. \quad (6)$$

Spatial verification function $SV(x_i, y_i)$ measures the consistency of all key point pairs belonging to the i -th bin. Let us define a function $N(a, b)$ which measures the consistency of a key point pair (a, b) . Consistency value has to be normalized, thus maximum

possible consistency $M(a, b)$ has to be defined. Functions $N(a, b)$ and $M(a, b)$ are defined further, by eqs. (9) and (10). Basic spatial consistency is defined as a ratio of measured consistency and maximum consistency (e.g. [12, 14]):

$$SV_{basic}(x_i, y_i) = \sum_{(a,b) \in P_i} \frac{|N(a, b)|}{|M(a, b)|}, \quad (7)$$

where: $P_i = x_i \times y_i$ – set of key point pairs belonging to i -th BoW bin, i.e., a Cartesian product of key point sets x_i and y_i .

The down side of such spatial consistency is that all neighboring pairs have the same contribution. As we know from the vector space model, some key points are more important than other ones. Their importance is measured by *tf-idf*. Following this idea, spatial verification with key point pair importance measurement is defined as:

$$SV_{ifidf}(x_i, y_i) = \sum_{(a,b) \in P_i} \frac{\sum_{(\alpha, \beta, k) \in N(a, b)} idf_k}{\sum_{(\alpha, \beta, k) \in M(a, b)} idf_k}, \quad (8)$$

where: (α, β) is the neighboring key point pair and it belongs to cluster k .

To address *scale-invariance* problem, neighborhood size for key point a and key point b should not be set equal. Useful information about scale can be extracted from square roots r_a and r_b of the areas of key points a and b . Thus neighborhood functions $N(a, b)$ and $M(a, b)$ are defined as:

$$N(a, b) = \{(\alpha, \beta) \in P : \|a, \alpha\| < \epsilon \wedge r_b \|b, \beta\| < r_a \epsilon\}, \quad (9)$$

and

$$M(a, b) = \{(\alpha, \beta) \in P : \|a, \alpha\| < \epsilon \vee r_b \|b, \beta\| < r_a \epsilon\}. \quad (10)$$

where: P is the set of all key point pairs, $\|\cdot, \cdot\|$ stands for Euclidean distance and ϵ is the distance limit. The above definition of $N(a, b)$ and $M(a, b)$ ensures that $|N(a, b)| \leq |M(a, b)|$ is always true.

Yet another information about quality of each key point pair can be extracted from the relative size of key points. Given that pair (a, b) has size ratio $\frac{r_a}{r_b}$, distances between key points ($\|a, \alpha\|$ and $\|b, \beta\|$) should follow the same ratio. Distance ratio agreement $ratio(a, b, \alpha, \beta)$ can be defined as:

$$ratio(a, b, \alpha, \beta) = \min \left(\frac{\|a, \alpha\| r_a}{\|b, \beta\| r_b}, \frac{\|a, \alpha\| r_b}{\|b, \beta\| r_a} \right). \quad (11)$$

Thus, final spatial verification routine $SV(x_i, y_i)$ is defined as:

$$SV(x_i, y_i) = \sum_{(a,b) \in P_i} \frac{\sum_{(\alpha, \beta, k) \in N(a, b)} idf_k \min \left(\frac{\|a, \alpha\| r_a}{\|b, \beta\| r_b}, \frac{\|a, \alpha\| r_b}{\|b, \beta\| r_a} \right)}{\sum_{(\alpha, \beta, k) \in M(a, b)} idf_k}. \quad (12)$$

Finally, *Bhattacharyya coefficient* with spatially weighted histogram bins is defined as:

$$BC_{SV}(x, y) = \sum_{i \in V} \sqrt{idf_i^2 \sum_{(a,b) \in P_i} \frac{\sum_{(\alpha,\beta,k) \in N(a,b)} idf_k \min\left(\frac{\|a,\alpha\|r_a}{\|b,\beta\|r_b}, \frac{\|a,\alpha\|r_b}{\|b,\beta\|r_a}\right)}{\sum_{(\alpha,\beta,k) \in M(a,b)} idf_k}}. \quad (13)$$

Let us now present the retrieval quality verification of all the discussed ideas.

3. Experimental verification

The proposed retrieval approach has been experimentally verified according to well established image retrieval protocols. Quality is measured using *mean average precision*. Reference queries and reference results are predefined. Precision–recall curves are shown. Two widely recognized image datasets are used in the process: Oxford5K [10] and Paris6K [11]. Our experiments address the following aspects of retrieval quality:

1. influence of number of *k-means* iterations during visual words dictionary generation,
2. influence of *BoW* similarity measurement,
3. a small flaw in the widely used experimental protocol,
4. proposed spatial verification routine,
5. cross-database visual words dictionary use.

3.1. Clustering and number of k-means iterations

The first of the addressed issues deals with the quality of BoW representation. It has been shown by researchers that high dimensional BoW gives better retrieval quality comparing to low dimensional ones. A standard approach based on *approximate k-means* have been used in several state-of-the-art papers [10, 18]. In our experiments we show that reverting to standard k-means can lead to higher quality. We also show that only a few k-means iterations are necessary to get satisfying clusters quality.

Modern hardware allows highly efficient implementation of k-means, using vector CPU and GPU processing (e.g. [13]). Thus, computational complexity of k-means method is no longer a problematic issue. Despite large increase in speed, the generation of visual words dictionary still takes some time. One iteration with 32 simultaneous distance calculations (CPU, 8 threads and 4 values in SIMD instructions) takes from several minutes to few hours for the presented data. Thus, we would like to know if it is worth iterating until k-means converges or it is possible to stop earlier.

Obviously, in each iteration of k-means the total distance between cluster centers and data points decreases. This decrease leads to cluster centers improvements and in result, to better retrieval quality. Achieved results are presented in Tab. 1. Our experiments show that it is sufficient to perform only a few k-means iterations.

Tab. 1. Retrieval quality for various number of k-means iterations.

iteration number	k-means distance	distance change	RootSIFT descriptor [18]		
			$\cos(x, y)$	$BC(x, y)$	$\chi^2(x, y)$
<i>Oxford5K image dataset, 500000 clusters</i>					
0	1,732,112,755	—	0.740	0.761	0.760
1	1,511,378,391	220,734,364	0.751	0.773	0.772
2	1,486,011,493	25,366,898	0.758	0.783	0.783
5	1,465,147,423	20,864,070	0.764	0.786	0.786
10	1,459,041,178	6,106,245	0.763	0.787	0.786

Tab. 2. Retrieval quality for various vocabulary sizes and similarity measures.

vocabulary size	SIFT descriptor [6, 7]			RootSIFT descriptor [18]		
	$\cos(x, y)$	$BC(x, y)$	$\chi^2(x, y)$	$\cos(x, y)$	$BC(x, y)$	$\chi^2(x, y)$
<i>Oxford5K image dataset</i>						
50000	0.673	0.661	0.659	0.708	0.692	0.696
100000	0.688	0.688	0.684	0.736	0.741	0.741
200000	0.720	0.738	0.733	0.760	0.771	0.770
500000	0.726	0.756	0.748	0.762	0.787	0.787
1000000	—	—	—	0.746	0.781	0.779
<i>Oxford5K image dataset – original clusters reference</i>						
1000000	0.636 [10]	—	—	0.683 [18]		
<i>Paris6K image dataset</i>						
200000	—	—	—	0.742	0.753	0.759

3.2. Bag of words similarity measures

A standard framework for large scale image retrieval uses *cosine similarity* of image BoW representation. Our experiments have shown that better results may be obtained with other similarity measures. Table 2 compares BoW retrieval results obtained for: *cosine similarity*, *Bhattacharyya coefficient* and χ^2 *histogram distance*. *Cosine similarity* turns to be the least effective because bag of words is in fact a histogram. Similar conclusions have been drawn by Zisserman [18] regarding SIFT features, when RootSIFT features were designed.

Comparison of retrieval quality is presented in Tab. 2. Precision–recall curves are shown in Fig. 1. Retrieval quality using *Bhattacharyya coefficient* and χ^2 *histogram distance* is higher in 7 out of 10 tested cases. These 7 cases are the important ones, because they have the highest overall quality. Achieved quality values are higher by up to 4%. Yet another interesting result is the comparison with state-of-the-art reference

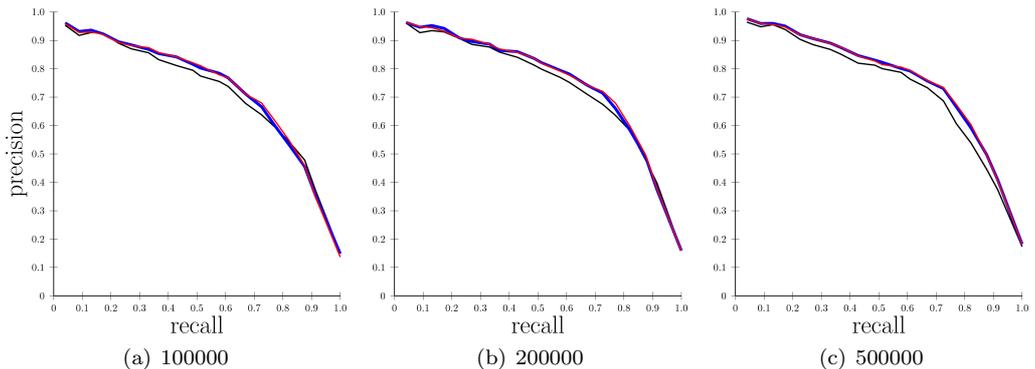


Fig. 1. Precision–recall curves for various vocabulary sizes and similarity measures (cos, BC, χ^2); Oxford5K database, RootSIFT descriptor.

results. Using the proposed approach, the achieved results quality is 8% higher for RootSIFT features and 9% higher for SIFT features.

3.3. Discussion on experimental protocol

In this section we would like to point out a small flaw in the broadly accepted experimental protocol. As pointed in [10] all 16.7M feature vectors are used in visual words dictionary generation. This approach does not seem to be valid in terms of machine learning quality estimation, because it is positively biased. The value of the bias increases with the size of the dictionary, because there are less and less vectors assigned to each group. Given *Oxford5k* database and 1M clusters, there are only about 17 vectors per cluster.

Comparison of clustering with and without query vectors is presented in Tab. 3. A decrease in retrieval quality is clearly visible in all cases.

Tab. 3. Retrieval quality for visual words dictionary generation with and without query vectors.

vocabulary size	grouping with all vectors			grouping without query vectors		
	$\cos(x, y)$	$BC(x, y)$	$\chi^2(x, y)$	$\cos(x, y)$	$BC(x, y)$	$\chi^2(x, y)$
	<i>Oxford5K image dataset</i>					
50000	0.708	0.692	0.696	0.646	0.650	0.652
100000	0.736	0.741	0.741	0.717	0.715	0.712
200000	0.760	0.771	0.770	0.731	0.743	0.741

There are two conclusions. First, more care should be taken when experimental protocols are constructed, because flaws like this one may lead to incorrect conclusions. Second, despite the drop in quality, *Bhattacharyya coefficient* and χ^2 *distance* are still better than *cosine similarity*, especially when the number of clusters increases.

3.4. Proposed spatial verification

Next performed experiment addresses the proposed spatial verification. Retrieval is organized in a hierarchical way [10, 18]. First, BoW similarity is calculated for the entire database. Results are ordered according to similarity (most similar come first). Top n images processed using spatial verification ($n = 1000$, according to the accepted protocol). After the verification, the subset of images is resorted once again.

First, we present baseline results of the spatial verification. The results of the following approaches are shown in Tab. 4:

- plain BoW with BC similarity measure (reference),
- spatial verification without BC integration (no square-root and L2 distance as norm),
- BC_{SV} variant without *tf-idf* weighting and without relative size of neighborhood,
- BC_{SV} variant without relative size of neighborhood,
- BC_{SV} variant without *tf-idf* weighting,
- full BC_{SV} proposed approach.

The proposed BC_{SV} outperforms all the other tested approaches. The following order of contribution arises out of the presented result:

- BC integration plays the key role and contributes most,
- relative key point size weighting is secondary,
- *tf-idf* contribution is least significant, but still permanent.

It is worth emphasizing that simple neighbors counting without the proposed extensions achieves worse results than plain bag of words retrieval (see Tab. 4, first and second rows). Let us now present the experimental setup of the parameter ϵ (see eq. (9))

Tab. 4. Retrieval quality comparison of BoW and spatial verification variants.

<i>retrieval approach</i>	<i>tf-idf neighbors</i>	<i>relative size weight</i>	<i>BC BoW integration</i>	dictionary size		
				100K	200K	500K
BoW, BC		–		0.741	0.771	0.787
partial BC_{SV}	no	no	no	0.726	0.744	0.749
partial BC_{SV}	no	no	yes	0.775	0.795	0.809
partial BC_{SV}	yes	no	yes	0.779	0.797	0.810
partial BC_{SV}	no	yes	yes	0.783	0.804	0.819
BC_{SV}	yes	yes	yes	0.785	0.807	0.820

Tab. 5. Retrieval quality for various ϵ parameter values in spatial verification.

dictionary size	ϵ parameter value				
	0.05	0.10	0.15	0.20	0.25
Oxford5K database, RootSIFT features					
100000	0.777	0.792	0.785	0.784	0.781
200000	0.791	0.809	0.807	0.805	0.803
500000	0.803	0.817	0.820	0.818	0.817

Tab. 6. Retrieval quality comparison for BoW and proposed spatial verification.

dictionary size	method			
	BoW	BC_{SV}	BC_{SV} - BoW	gain(BC_{SV} , BoW)
Oxford5K database, RootSIFT features				
50000	0.692	0.754	0.062	8.9%
100000	0.741	0.785	0.044	5.9%
200000	0.771	0.807	0.036	4.6%
500000	0.787	0.820	0.033	4.2%
1000000	0.781	0.799	0.018	2.3%
reference [18]	0.683	0.720	0.037	5.4%
Oxford5K database, RootSIFT features, no queries in clusters				
50000	0.650	0.713	0.063	9.6%
100000	0.715	0.769	0.054	7.5%
200000	0.743	0.786	0.043	5.7%
Oxford5K database, SIFT features				
50000	0.661	0.719	0.058	8.8%
100000	0.688	0.745	0.057	8.3%
200000	0.738	0.785	0.047	6.4%
500000	0.756	0.787	0.031	4.1%
reference [18]	0.636	0.672	0.036	5.7%
Paris6K database, RootSIFT features				
200000	0.753	0.783	0.030	4.1%

and (10)) specifying the neighborhood size. Tab. 5 presents retrieval quality with various values of the parameter. Best retrieval quality is reached for $\epsilon \in \langle 0.10, 0.20 \rangle$. Suggested value of ϵ is 0.15. This value is used in all the presented experiments. Detailed results of the final BC_{SV} approach are presented in Tab. 6. Precision–recall curves for various vocabulary sizes are presented in Fig. 2. Retrieval with spatial verification clearly dominates bag of words retrieval in all the tested cases.

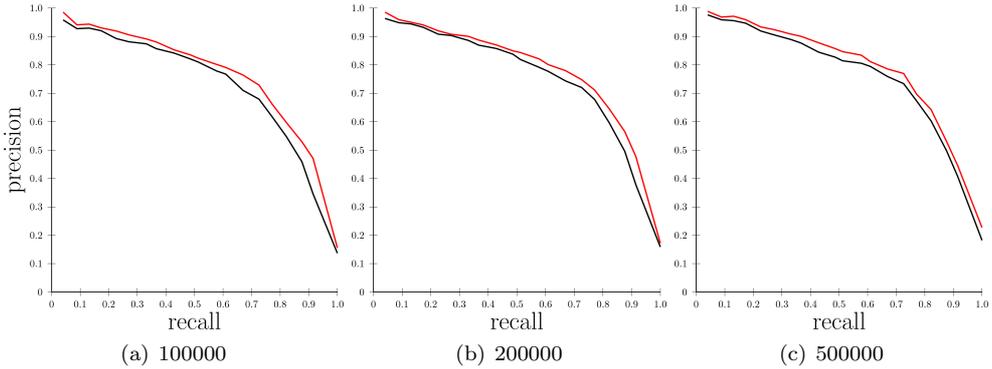


Fig. 2. Precision–recall curves for retrieval with and without spatial verification (BoW, SV), Oxford5K database, RootSIFT descriptor, various vocabulary sizes.

Due to differences in BoW retrieval quality, comparison with reference approaches is done using $gain(x, y)$ measure for a given quality measure q :

$$gain(x, y) = \frac{q(x) - q(y)}{q(y)}. \quad (14)$$

This approach works in favor of the reference approach. Reference BoW quality is lower than that achieved BoW quality. Thus it is more difficult for the proposed method to get a similar quality increase. Presented results show that the proposed retrieval scheme outperforms reference approach both in absolute values and in gain values.

3.5. Cross-database retrieval

The last presented test addresses cross-database retrieval. Visual words dictionary is built on one database but tests are performed on a different database. In the presented tests we use Oxford5K and Paris6K databases. This is a very important test because it minimizes the problem of cluster over-training.

Achieved retrieval quality is much lower comparing to single database tests (see Tab. 7). Visual words dictionary is not fine tuned for the retrieved data. One can observe large differences in cluster quality directly estimated using the distance criterion which k -means minimizes (Tab. 7, first column).

Interestingly, relations between quality of tested approaches are similar to those in single database tests. *Cosine* similarity has worst results, spatial verification on *Bhattacharyya coefficient* achieves highest quality. Gain values for spatial verification routine on cross-database retrieval are larger than those for a single database test. For Oxford5K database it reaches 9.7%, for Paris6K database it is equal to 6.9%.

Tab. 7. Retrieval quality for cross database visual dictionary generation, 200000 clusters.

k-means distance	retrieval approach				gain(BC_{SV} , BoW)
	$\cos(x, y)$	$BC(x, y)$	$\chi^2(x, y)$	BC_{SV}	
<i>Oxford5K image dataset, Paris6K clusters</i>					
1,617,336,879	0.587	0.603	0.607	0.662	9.7%
<i>Oxford5K image dataset, Oxford5K clusters</i>					
1,538,261,883	0.760	0.771	0.770	0.807	4.8%
<i>Paris6K image dataset, Oxford5K clusters</i>					
1,879,399,544	0.608	0.619	0.630	0.662	6.9%
<i>Paris6K image dataset, Paris6K clusters</i>					
1,774,815,215	0.742	0.753	0.759	0.783	4.1%

3.6. Concluding remarks

Taking into account the above experiments we found the following conclusions. Quality of visual words dictionary has a large impact on the retrieval quality. The above statement confirms the results and conclusions reached by other researchers. However, the usage of exact k-means instead of approximate or hierarchical versions seems to be a better choice. Modern hardware (both CPU and GPU) makes usage of exact k-means no longer a blocker, as it was several years ago. We have also found out that only few iterations of k-means is sufficient to get satisfying retrieval quality.

The second conclusion addresses a small flaw in the widely used experimental protocol [10]. We state that clustering all data, together with query vectors, breaks the principles of machine learning. We show that when query vectors are removed from the clustering process, the retrieval quality degrades.

The third conclusion focuses on the retrieval process itself. We suggest that *Bhattacharyya coefficient* and χ^2 distances should replace *cosine similarity*. They give better retrieval quality (see Tabs. 1, 2, 3 and 7). We have also presented an alternative to RANSAC-based spatial verification routine. The proposed approach combines neighborhood consistency with term weighting and bag of words similarity measurement. Measured gain values over standard bag of words approach are highest for cross-database retrieval. In the presented scenario they reach 9.7%.

4. Summary

This paper shows that the potential of simple image retrieval approaches have not been fully explored. Vector space model is a well known, researched and established technique.

However, our approach has outperformed the recently reported results [18] by 10% using the same experimental protocol. The reasons of such an improvement are:

- better visual words dictionary generation using *k-means* instead of *approximate k-means*,
- application of less popular *Bhattacharyya coefficient* or χ^2 distance instead of *cosine similarity*.

The third reason and research contribution of the paper is the spatial verification. It integrates neighborhood consistency with *tf-idf* neighbor weighting and key point size ratio weighting. It is formulated as histogram weighting routine for *Bhattacharyya coefficient*. Measured gain values for the proposed spatial verification are between 2.3% and 9.7%.

Acknowledgment

The work is financed from the European Union structural funds, Innovative Economy Operational Program, POIG.01.01.02-14-013/09.

References

1975

- [1] G. Salton, A. Wong and C. S. Yang. A vector space model for automatic indexing. *Communications of the ACM*, vol. 18(11), 1975, pp. 613–620.

1996

- [2] M. Fischler and R. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *4th European Conference on Computer Vision (ECCV'96)*, Cambridge (UK), 1996, pp. 683–695.
- [3] C. Schmid and R. Mohr. Object recognition using local characterization and semi-local constraints. Technical report, INRIA, 1996.

1997

- [4] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5), 1997, pp. 530–535.

2003

- [5] J. Sivic and A. Zisserman. Video google: a text retrieval approach to object matching in videos. *Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV'03)*, vol. 2, 2003, pp. 1470–1477.

2004

- [6] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *International Journal of Computer Vision*, vol. 60, 2004, pp. 63–86.
- [7] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, vol. 60, no. 2, 2004, pp. 91–110.

2006

- [8] J. Zobel and M. Alistair. Inverted files for text search engines. *ACM computing surveys (CSUR)*, vol. 38(2), 2006.

2007

- [9] S. H. Cha. Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, vol. 1(4), 2007, pp. 300–307.
- [10] J. Philbin, O. Chum, M. Isard, J. Sivic and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

2008

- [11] J. Philbin, O. Chum, M. Isard, J. Sivic and A. Zisserman. Lost in quantization: improving particular object retrieval in large scale image databases. *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [12] O. Chum, J. Philbin and A. Zisserman. Near duplicate image detection: min-hash and tf-idf weighting. *Proceedings of the British Machine Vision Conference*, 2008, pp. 812–815.
- [13] R. Farivar, D. Rebolledo, E. Chan and R. H. Campbell. A parallel implementation of k-means clustering on GPUs. *The 2008 International Conference on Parallel and Distributed Processing Techniques and Applications*, 2008, pp. 340–345.

2009

- [14] O. Chum, M. Perdoch and J. Matas. Geometric min-hashing: finding a (thick) needle in a haystack. *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 17–24.

2010

- [15] H. Jegou, M. Douze and C. Schmid. Improving bag-of-features for large scale image search. *International Journal of Computer Vision*, vol. 87, 2010, pp. 316–336.
- [16] D. D. Yang and A. Sluzek. A low-dimensional local descriptor incorporating TPS warping for image matching. *Image and Vision Computing*, Vol. 28(8), August 2010, pp. 1184–1195.

2011

- [17] M. Paradowski and A. Sluzek. Local keypoints and global affine geometry: triangles and ellipses for image fragment matching. *Innovations in Intelligent Image Analysis* (eds. H. Kwasnicka, L. Jain), Springer Verlag, vol. 339, 2011, pp. 195–224.

2012

- [18] R. Arandjelovic and A. Zisserman. Three things everyone should know to improve object retrieval. *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2911–2918.

TECHNIQUE TO GENERATE FACE AND PALM VEIN-BASED FUZZY VAULT FOR MULTI-BIOMETRIC CRYPTOSYSTEM

N. Lalithamani¹ and Dr. M. Sabrigiriraj²

¹Assistant Professor(SG), Amrita School of Engineering

Amrita Vishwa Vidyapeetham, Amritanagar(PO), Ettimadai, Coimbatore – 641 112.

n.lalitha@cb.amrita.edu

²Head of the Department, Department of Computer Science and Engineering

SVS College of Engineering, Coimbatore – 642 109.

sabari_giriraj@yahoo.com

Abstract. Template security of biometric systems is a vital issue and needs critical focus. The importance lies in the fact that unlike passwords, stolen biometric templates cannot be revoked. Hence, the biometric templates cannot be stored in plain format and needs strong protection against any forgery. In this paper, we present a technique to generate face and palm vein-based fuzzy vault for multi-biometric cryptosystem. Here, initially the input images are pre-processed using various processes to make images fit for further processing. In our proposed method, the features are extracted from the processed face and palm vein images by finding out unique common points. The chaff points are added to the already extracted points to obtain the combined feature vector. The secret key points which are generated based on the user key input (by using proposed method) are added to the combined feature vector to have the fuzzy vault. For decoding, the multi-modal biometric template from palm vein and face image is constructed and is combined with the stored fuzzy vault to generate the final key. Finally, the experimentation is conducted using the palm vein and face database available in the CASIA and JAFFE database. The evaluation metrics employed are FMR (False Match Ratio) and GMR (Genuine Match Ratio). From the metric values obtained for the proposed system, we can infer that the system has performed well.

Key words: multimodal biometric cryptosystems, biometric template security, palm vein feature extraction, face feature extraction, fuzzy vault, secret key.

1. Introduction

Biometric technology offers to identify people through physical measurements of unique human characteristics or behavior. In other words, we all have unique personal attributes that can be used for distinctive identification purposes, including a fingerprint, the pattern of a retina, and voice characteristics. Biometric technology offers the promise of an easy, secure method to make highly accurate verifications of individuals. It guarantees a means of identification that cannot be stolen, lost or forgotten, are being increasingly demanded in security environments and applications like access control and electronic transactions. By replacing Passwords, biometric techniques can potentially

prevent unauthorized access to or fraudulent. However, even the best biometrics to date is still facing numerous problems, some of them inherent to the technology itself. Multibiometrics are a relatively new approach to overcome these problems by having multiple samples of a single biometric trait (called multi sample biometrics) or samples of multiple biometric traits (called multi-source or multimodal biometrics) [5].

The use of biometric data in the context of identity attribute verification poses several non trivial challenges because of the inherent features of the biometric data. In general, two subsequent readings of a given biometrics do not result in exactly the same biometric template. Therefore the matching against the stored template is probabilistic. Storing biometric templates in repositories along with other personally identifiable information introduces security and privacy risks [6]. Those databases can be vulnerable to attacks by insiders or external adversaries and may be searched or used for purposes other than the intended one. If the stored biometric templates of an individual are compromised, there could be severe consequences for the individual because of the lack of revocation mechanisms for biometric templates [26].

In many medical practices, X-ray and ultrasonic scanning are used to form vascular images. This is not acceptable for general purpose biometric applications in the real world. Therefore, obtaining the palm vein pattern images in a fast and non-invasive manner is the key challenge in a vein pattern biometric system [10]. The personal identification using hand and palm vein has gained more and more research attentions these years as it seems a better biometric feature than finger print and other modalities [4,14]. Palm print recognition is getting popular in personal authentication because it provides robust features from a large palm area and the palmprint image can be captured with a cost effective device. In general, a typical palm print acquisition device operates under visible light and can acquire three kinds of features: principle lines (usually the three dominant lines on the palm), wrinkles (weaker and more irregular lines) and ridges (patterns of raised skin similar to fingerprint patterns). A resolution of about 100 dpi (dots per inch) [2, 3] can be used to acquire principal lines and wrinkles while a higher resolution, usually 500 dpi, is required to acquire ridge features. However, such a high resolution will increase significantly the computational cost to extract ridge features because of the large image size of palm, and hence prevents the system from being implemented in real time. Therefore, most of the palmprint base systems capture low resolution palmprint images using CCD (charge coupled device) cameras and many algorithms have been proposed for feature extraction and matching [11, 12, 17, 18, 20]. Face is another biometric modality is used regularly these days and has been researched to have better results.

The main challenge for embedded versions is to provide a secure storage of the reference template. Embedded devices are vulnerable to eavesdropping and attacks. Thus alternative protection mechanisms need to be investigated. Recently, a novel cryptographic technique called the fuzzy commitment scheme has been proposed for biometric

authentication [1]. The scheme integrates well known error control coding methods and cryptographic techniques to construct a novel type of cryptographic system. Instead of an exact, unique decryption key, a reasonable close witness can be accepted to decrypt the commitment. This characteristic makes it possible for protecting the biometric data using traditional cryptographic techniques [25].

In this paper, we have designed and developed a face and palm vein-based fuzzy vault technique that improves the multimodal biometric template security especially. Initially, the preprocessing steps are applied to both palm-vein and face images for enhancement and smoothing. Then, the features of two biometric images are extracted by finding the common unique points in the images and the extracted features are combined along with chaff points for generating multimodal biometric template. Finally, the multimodal biometric template and the input key are used to generate the fuzzy vault. For decoding, the multimodal biometric template from palm vein and face image will be constructed and it is combined with the stored fuzzy vault to generate the final key. Finally, the experimentation will be conducted using the palm vein and face database available in the CASIA and JAFFE database.

The rest of the paper is organized as follows: A brief review of researches related to the proposed technique is presented in section 2. The proposed biometric storage using fuzzy vault technique is presented in Section 3. The detailed experimental results and discussions are given in Section 4. The conclusions are summed up in Section 5.

2. Brief Review of Related Works

Literature presents lot of works related to biometric recognition and template storage. Here, some of works related to these are reviewed in this section. Abhilasha Bhargav et al. [26] have presented biometrics-based identifiers for digital identity management. They presented algorithms to reliably generate biometric identifiers from a user's biometric image which in turn was used for identity verification possibly in conjunction with cryptographic keys. Their algorithm captured generic biometric features that ensured unique and repeatable biometric identifiers. They also ensured security and privacy of the biometric data. K. Nandakumar [21] has presented a fingerprint cryptosystem based on minutiae phase spectrum. He proposed a minutiae representation known as the Binarized Phase Spectrum (BiPS), which was a fixed-length binary string obtained by quantizing the Fourier phase spectrum of a minutia set. He secured the BiPS representation using a fuzzy commitment scheme employing turbo codes. He also proposed a technique for aligning fingerprints based on the focal point of high curvature regions.

Dr. S. Ravi and Mahima S. [27] have presented several significant strides in facial expression recognition and its applications. At first, the facial expression was measured by virtue of its position. The results for this procedure was far from accurate because

static nature and deficient of intensity measurements. The facial EMG has made a radical transformation in the calculation of automatic variation in onset and offset of facial expression. The shortfall owing to static nature has been rectified with the insertion of facial expression recognition by image sequence. The expressive expression where mapping is carried out with ratio images, made it probable to calculate the transformation of one person's expression by calculating expression ratio image to produce more expressive facial expressions of any other person. Ghandi et al. [16] have presented a technique to detect facial emotion by employing a modified Particle Swarm Optimization algorithm, which they named as Guided Particle Swarm Optimization (GPSO). The technique engages tracking the activities of Action Units (AUs) located at suitable points on the face of a subject. Two dimensional rectangular shaped search spaces were defined around each of the action units and particles were then described to contain a constituent in each domain, efficiently building a 10-dimensional search space inside which particles fly in search of a solution. Multiple swarms were employed where each swarm had a target emotion.

Peng Li et al. [23] have presented an alignment-free fingerprint cryptosystem based on fuzzy vault scheme. They introduced an alignment-free fingerprint cryptosystem based on fuzzy vault scheme was developed fusing the local features, known minutia descriptor and minutia local structure, which were invariant to the transformation in fingerprint capturing. The proposed fingerprint cryptosystem avoided the alignment procedure and improved the performance and security of the fuzzy vault scheme at the same time. David Zhang et al. [20] have presented online joint palmprint and palmvein verification method for increasing the anti-spoof capability of the system. Yiding Wang et al [25] have presented hand-dorsa vein recognition based on partition local binary pattern. They introduced hand-dorsa vein recognition method based on Partition Local Binary Pattern (PLBP). The method employed hand-dorsa vein images acquired from a low-cost, near infrared device. After preprocessing, the image was divided into sub-images. LBP uniform pattern features were extracted from all the sub-images, which were combined to form the feature vector for token vein texture features. The method was assessed using a similarity measure obtained by calculating the distance between the feature vectors of the tested sample and the target sample.

Maleika Heenaye et al. [22] have presented feature extraction of dorsal hand vein pattern using a fast modified PCA algorithm based on Cholesky decomposition and Lanczos technique. Principle Component Analysis (PCA) was a successful method which was originally applied on face biometric. They modified PCA using Cholesky decomposition and Lanczos algorithm to extract the dorsal hand vein features. This modified technique decreases the number of computation and hence decreases the processing time. The eigenveins were successfully computed and projected onto the vein space. Zhenhua Guo et al. [19] have presented palmprint based verification method. They proposed a unified distance measure and provided some principles for determining the parameters

of the unified distance. They showed that, using the same feature extraction and coding methods, the unified distance measure got lower equal error rates than the original distance measures. Zhenhua Guo et al. [28] have presented empirical study of light source selection for palmprint recognition. They analyzed that most of the current palmprint recognition systems used an active light to acquire images, and the light source was a key component in the system. Although white light was the most widely used light source, little work had been done on investigating whether it was the best illumination for palmprint recognition. They studied the palmprint recognition performance under seven different illuminations, including the white light. A large database showed that white light was not the optimal illumination, while yellow or magenta light could achieve higher palmprint recognition accuracy than the white light.

Xiangqian Wu et al. [24] have presented a biometric system based on hand vein. They presented a hand vein recognition system, which extracted and combined the dorsal, palm and finger vein for personal recognition. In the proposed system, the whole infrared frontal and back images of a hand were initially captured. Secondly, the Region Of Interest (ROI) of dorsal, palm and finger vein images were cropped. Thirdly, the veins in each ROI were extracted and matched by using multi-scale 2-D Gaussian matched filter. Finally, the matched distances were fused to form the final distance for decision by employing SVM classifier.

3. Proposed Technique to Generate Face and Palm Vein-Based Fuzzy Vault for Multi-Biometric Cryptosystem

Template security is an important concern in biometric systems because unlike passwords, stolen biometric templates cannot be revoked. Due to these reasons, biometric templates should not be stored in plaintext form and fool-proof techniques are required to securely store the templates such that both the security of the application and the users' privacy are not compromised by adversary attacks. However, a multi-biometric system requires storage of multiple templates for the same user corresponding to the different biometric sources. Hence, template security is even more critical in multi-biometric systems where it is essential to secure multiple templates of a user.

In our technique, we make use of biometric modalities of face and palm print of a person. The input biometric image is initially pre-processed and feature extracted. Features from both the biometric modalities are then combined to have the combined biometric feature vector, to which the chaff points and secret key points are added to have the fuzzy vault. In the test phase, the test person's face and palm vein biometric images are matched to the fuzzy vault and if matched, secret key is extracted by the technique. The technique is split into four phases: a) Pre-processing, b) Combined feature vector generation, c) Fuzzy vault generation and d) Test case.

3.A. Pre-Processing Phase

The input face and palm vein images are pre-processed before its feature extracted so as to have better results. Pre-processing makes the image fit for further processing of feature extraction and gets rid of the noise and blur in the input image. The pre-processing methods involved in face and palm vein are almost same except in few steps. Palm vein image is initially done by cropping, which is absent in processing of face image. The other difference is that edge detection is used in face image instead of motion filter which is used in palm vein image. The block diagram of the pre-processing phase is given in Figure 1.

Cropping refers to removal of unwanted areas from the input image so as to obtain the image with area of interest. In our case, the inner part of the input palm vein is only required, hence we remove other parts including fingers from the image. Image Scaling is the process of enhancing the smoothness, sharpness and resizing a digital image. The main objective of the scaling is to rescale the input image size to a standard size and as a result of image scaling different size images are resized to a single standard size. Image scaling makes the further processing easier. The resized images are then type converted to double precision format. Double precision is a computer number format that store in

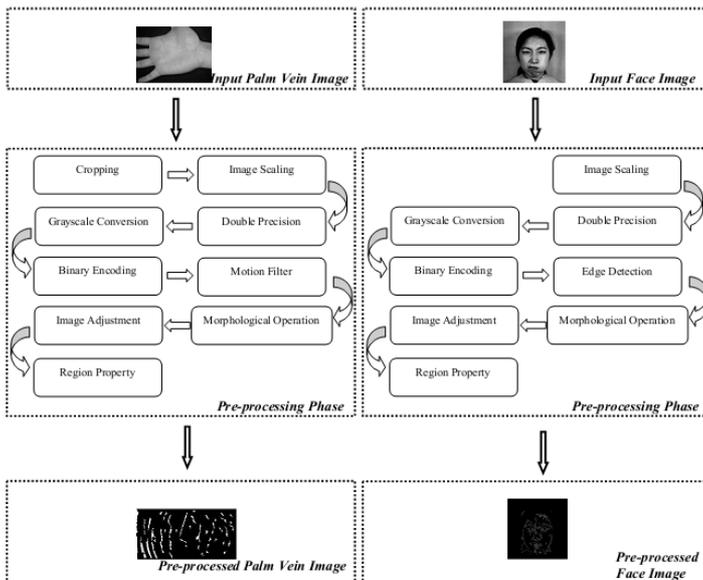


Fig. 1. Block Diagram of the Pre-Processing Phase

two adjacent storage locations. A double-precision number, also called as a double, may be defined to be an integer, fixed point or floating point.

The input images in RGB format are then converted to grey scale images. Greyscale images are composed exclusively of shades of grey, varying from black to white based on intensity value. The grey image is then binary changed where the image is converted to binary format. A binary image has only two possible values for each pixel which is either digits zero or one which are typically represented by the black and white colours. The binary image is motion filtered in case of palm vein image where the linear motions of the images are approximated. Motion filtering is carried with the help of line smoothing process. Smoothing of the image is to build an approximating function that effort to detect vital patterns, while leaving out noise. Here, the image pixels are modified so that the noise pixels are reduced and smoother image is obtained. Here, the smoothed values can be written as a linear transformation of the observed values. In case of face image, the edge detection is carried out to find out the edge points in the input face image. Edge detection for face is vital as features have to be extracted from the face area and not from the outside area. Detection of face edge also results in extraction of points form face edges which is a property of the individual's face as the shape of face varies from person to person.

The image is enriched through image adjustments where various parameters are decided for adjusting the images such as image intensity levels, brightness and colour balance. These parameters improve the quality of the image which in turn yields better feature extraction. Morphological Operation is employed subsequently which is defined by image processing operations that process images based on shapes is a broad set of morphology. The output image is based on a comparison of the corresponding pixel in the input image with its neighbor's and the value of each pixel in a morphological operation. We add pixels to the boundaries of objects in an image and we make use of dilation in morphological process for both palm vein image as well as face image. Set-theoretic processes like union and intersection are employed to describe morphological operations. The two inputs for the morphological operation are binary image and structuring element. Initially the image Im is adjusted for contrast and intensity and is subsequently converted to the binary form Im_b . The enhanced image is obtained through the morphological operation 'imerod' that utilizes the structuring element Sa using the equation:

$$Im_b \ominus Sa = \bigcap_{j \in Sa} A_{-j}$$

The above equation is employed for computing the erode function. The maximum intensity pixels of the image alone are selected by using morphological operation. Hence, adjusted image is additionally improved by exploiting the morphological operation.

The image obtained after the morphological operation will have many unwanted pixels in it, so in order to filter out those we use the region property. The palm veins will

not only have required lines but also some other unwanted small lines and pixels. Hence, these unwanted elements are filtered out from the image for better feature extraction using region property. Similarly in the face image, unwanted elements are removed from the image using region property.

3.B. Combined Feature Vector Generation Phase

In this phase, the features points extracted from the face and palm vein images (Figure 2). The extracted features are added with the randomly generated chaff points for the user to generate the combined feature vector. For any person, respective face images and palm vein images are taken as the input and the respective features are extracted from these images. Feature extraction is carried out based on finding out the unique points from the images of the concerned person. Suppose for a person i , let the input palm images be represented by $P_i = \{p_{i1}, p_{i2}, \dots, p_{iNp}\}$, where Np is the total number of palm vein images available for the person. Similarly, let the input face images of the person i be represented by $F_i = \{f_{i1}, f_{i2}, \dots, f_{iNf}\}$, where Nf is the total number of face images available for the person. From all the input palm images of the user, the feature points are extracted which are the common unique points in all the images. That is, if a point ep denoted by co-ordinates (x, y) is common to all images $p_{i1}, p_{i2}, \dots, p_{iNp}$, then point

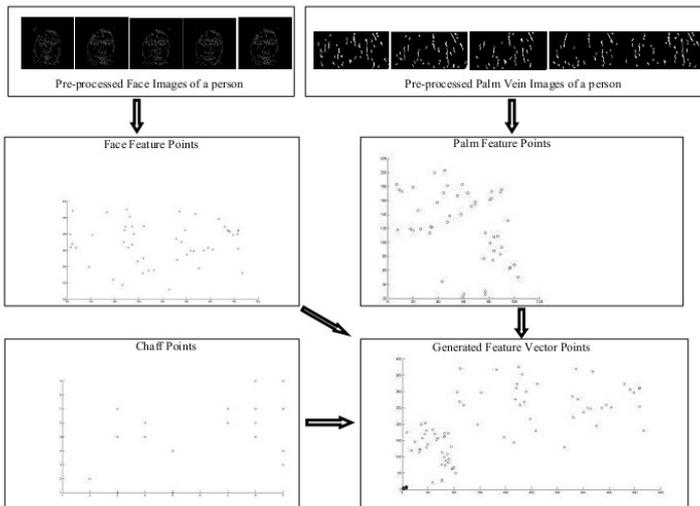


Fig. 2. Block Diagram of Combined Feature Vector Generation Phase. Face Feature Points Range Y axis-100 to 400, X axis-100 to 500, Palm Feature points Y axis- 0 to 220, X axis-0 to 120 Chaff Points Range Y axis-1 to 9, X axis-1 to 9, Generated feature points Y axis- 0 to 400, X axis-0 to 500.

$ep(x, y)$ is taken as a palm image feature point. Similarly, all other common unique points from the input palm images of the person are found out and let the extracted feature points from be represented as $Ep_i = \{ep_{i1}, ep_{i2}, \dots, ep_{iNep}\}$, where Nep is the total number of palm feature points extracted for person i . Similarly, the feature points are extracted from the face images of the person by finding out the common unique points from the input face images of the person. Let the extracted feature points from be represented as $Ef_i = \{ef_{i1}, ef_{i2}, \dots, ef_{iNef}\}$, where Nef is the total number of face feature points extracted for person i .

For any person, apart from the face and palm feature points, some additional random points are added known as chaff points. Chaff points are added to improve the security while forming the combined feature vector. Let the chaff points for person i be represented as $Ec_i = \{ec_{i1}, ec_{i2}, \dots, ec_{iNec}\}$, where Nec is the total number of chaff points added. Chaff points are randomly added for the person. The combined feature vector is formed by combining the feature points from face, palm and the chaff points. Hence the combined feature vector for a user i can be represented as $E_i = \{Ep_i, Ef_i, Ec_i\}$, which can be expanded to $E_i = \{ep_{i1}, ep_{i2}, \dots, ep_{iNep}, ef_{i1}, ef_{i2}, \dots, ef_{iNef}, ec_{i1}, ec_{i2}, \dots, ec_{iNec}\}$ and the total number of extracted points in the combined feature vector is $Nep + Nef + Nec$. In our technique, though we extract all the common points, we make use of only some points in order to reduce the complexity and time of execution. Here we limit the extracted points from face and palm to 50 and number of chaff points to 20 so as to have total of 120 points in the combined feature vector for a person. Selection of these 50 points from each modality is carried out on based on first come first serve basis. That is first 50 unique common points found out from each of face and palm would form the feature vector. Feature vector also includes 20 chaff points to make a total of 120 points.

3.C. Fuzzy Vault Generation Phase

Fuzzy vault improves the security of template security by the addition of secret key concept into the feature vector. Initially, the input key is encoded to have the respective points which are added to the points of the feature vector to have the fuzzy vault points. The number of secret key points generated is directly dependent on the number of digits in the secret key and if secret key is 4 bit long, then 4 points will be added to the feature vector to form the fuzzy vault. Block diagram of the fuzzy vault generation is given in Figure 3.

Generation of points for the secret key is based on the below mentioned designed mechanism which provides security to the combined face and palm vein templates. The x-coordinates of the secret key points are the digit itself and the y-coordinate is the next odd number in case if the digit is odd and next even number in case of an even digit. Suppose the input key is of size Nk and the key is represented by $K_i = K_{i1}K_{i2}\dots K_{iNk}$, where K_{ij} is the j^{th} digit of the i^{th} person secret key. Taking a digit K_{ij} in the secret key, corresponding point KL_{ij} is formed by co-ordinates (K_{ij}, L_{ij}) , where L_{ij} is the

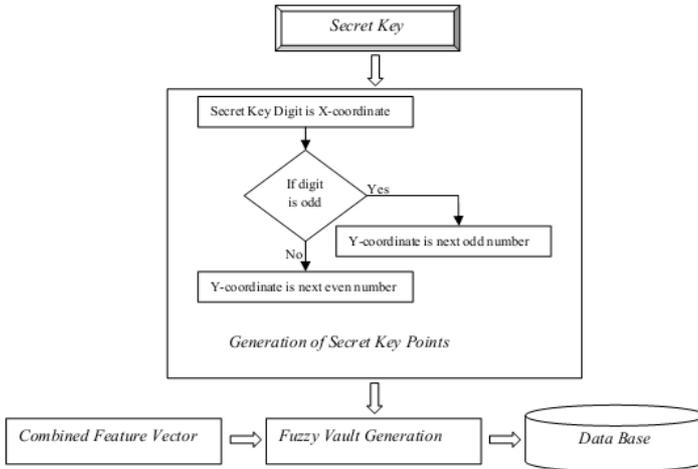


Fig. 3. Block Diagram of the Fuzzy Vault Generation Phase

next odd number after K_{ij} when K_{ij} is odd and L_{ij} is the next even number after K_{ij} when K_{ij} is even. Similarly all the points for each of the secret key digit are formed in the similar manner so as to result in points $EK_i = \{KL_{i1}, KL_{i2}, \dots, KL_{iNk}\} = \{(K_{i1}, L_{i1}) (K_{i2}, L_{i2}) \dots (K_{iNk}, L_{iNk})\}$. These points are added to the concatenated vector to form the fuzzy vault which is given by $FV_i = \{E_i, EK_i\}$, which can be expanded to form $FV_i = \{Ep_i, Ef_i, Ec_i, KL_{i1}, KL_{i2}, \dots, KL_{iNk}\}$. The feature vector can be represented by

$$FV_i = \{ep_{i1}, ep_{i2}, \dots, ep_{iNep}, ef_{i1}, ef_{i2}, \dots, ef_{iNef}, ec_{i1}, ec_{i2}, \dots, ec_{iNec}, KL_{i1}, KL_{i2}, \dots, KL_{iNk}\}.$$

Hence, the total number of points in the fuzzy vault is $Nep + Nef + Nec + Nk$. In our technique, we employ secret key of size 4 so as to generate total of 124 points in the fuzzy vault. Each of the person will have a corresponding fuzzy vault and all the fuzzy vaults formed FV_i for $0 < i < Np$ are stored in the database, where Np is the total number of persons.

3.D. Test Case Phase

In this phase, a test person's face and palm images are given as input which is pre-processed and feature extracted to form the combined feature vector. The input feature vector is compared to the fuzzy vaults in the database and if matched, the secret key is generated to confirm with the person and authentication is provided. Block diagram of test case phase is given in Figure 4.

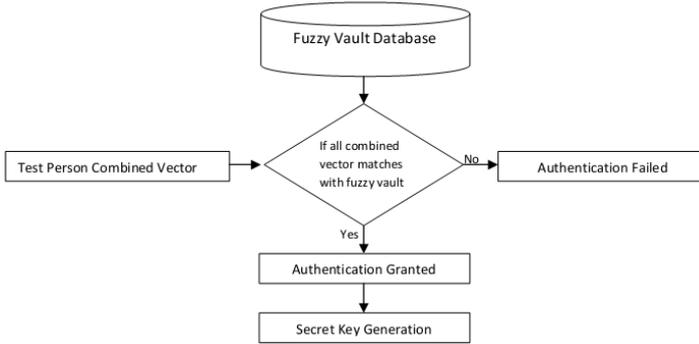


Fig. 4. Block Diagram of the Test Case Phase

Let the input person’s feature vector points be represented by $E_t = \{E_{p_t}, E_{f_t}, E_{c_t}\}$, which is compared to fuzzy vault in the database, FV_i for $0 < i < N_p$. If all the points feature vector of the test person matches into the fuzzy vault, then the person is granted authentication else the authentication is denied. Once all the points in test person feature vector matches with the fuzzy vault form the database, then certain points in the fuzzy vault will be still be left alone. These points are the secret key points and the x-coordinate of these points will give the secret key of the person. Suppose $(K_{i_1}, L_{i_1}) (K_{i_2}, L_{i_2}) \dots (K_{i_{N_k}}, L_{i_{N_k}})$, then the secret key is $KL_{i_1}, KL_{i_2}, \dots, KL_{i_{N_k}}$. The generation of key for the person is a second confirmation of the person and improves the template security.

4. Results and Discussions

The proposed technique to generate face and palm vein-based fuzzy vault for multi-biometric cryptosystem is evaluated and analyzed in this section. Section 4.1 gives the experimental set up and the evaluation metrics employed. In section 4.2, data set description is given and in section 4.3 it gives the data set preparation. The experimental results are discussed in section 4.4 and performance analysis is made in section 4.5. Comparative performance is given in section 4.6

4.1. Experimental set up and evaluation metrics

The proposed technique is implemented in MATLAB on a system having 6 GB RAM and 2.6 GHz Intel i-7 processor. For determining the accuracy and efficiency of the technique, the error rates are measured and analyzed. NGRA (Number of Genuine Recognition Attempts) gives the number of attained matches. Rejection of face and

palm print images F_{ij} may happen due to various reasons and all these rejections are summed up to have REJ_{ENROLL} .

$$GMR(t) = \frac{gms}{NGRA}$$

Here, gms is the genuine matching score. In addition with, each of the face and palm images $K_{1i}, i = 1, 2, \dots, 10$ is matched against with the first set of face and palm print images from database $F_{ik}(i < k \leq 10)$ and the corresponding Impostor Matching Score (ims) is calculated. The number of matches (denoted as NIRA - Number of Impostor Recognition Attempts) is $((50 \times 49) / 2) = 1225$ only if, $REJ_{ENROLL} = 0$.

$$FMR(t) = \frac{ims}{NIRA}$$

Furthermore, the $FMR(t)$ (False Match Rate) and $GMR(t)$ (Genuine Match Rate) are calculated from the above distributions for t ranging from 0 to 1.

4.2. Dataset description

Face image The Japanese Female Facial Expression (JAFFE) Database [30] contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral, Figure 5, bottom) posed by 10 Japanese female models. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects.

Palm print CASIA Palm-print Image Database [29] contains of 5,502 palmprint images captured from 312 subjects. For each subject, images are obtained for both the left and right palms (Figure 5, top). All the images are 8 bit gray-level JPEG files.



Fig. 5. Images of the face and the palm print taken from the respective databases

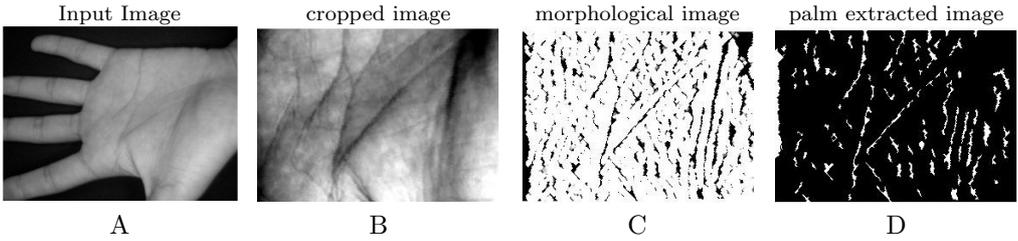


Fig. 6. Palm image outputs at various stages

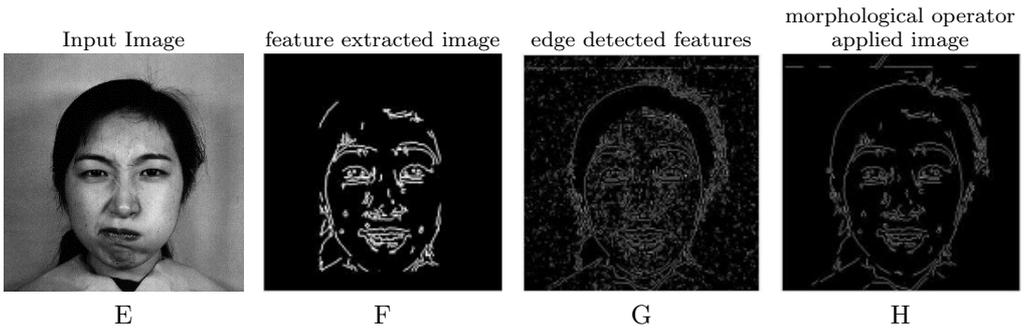


Fig. 7. Face image outputs at various stages

4.3. Dataset preparation

The fact that we are taking the face and palm print from different databases, the face and the palm print will not be of the same person. But as we have to evaluate the proposed technique with both the palm and face image, we approximate the corresponding images from different databases to be of a single person. That is, one image from the hand vein database and another image from the palm print database is combined together to form a single persons biometrics.

4.4. Experimental Results

The section gives the image at different stages of execution. For palm print, the images at six different stages are given. Here Fig. 6 gives the palm images at various stages; A gives the input image which cropped to have B. The morphological image obtained is given by C, and D gives the extracted palm print image. Figure 7 gives the face images at different stages: E gives the input face image, F – the feature extracted image, G shows the edge detected features, and H – the morphological operator applied image.



Fig. 8. Plot of GMR values

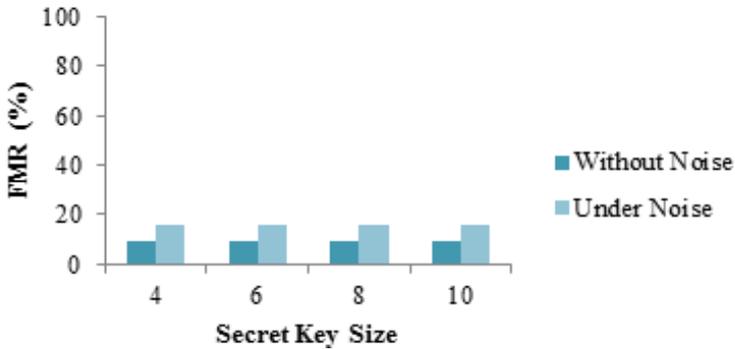


Fig. 9. Plot of FMR values

4.5. Performance analysis

The evaluation metric values of GMR and FMR obtained for the proposed technique is discussed in this section. The study is made in presence and in absence of noise and also by varying the key size of the secret key. Table 1 gives the evaluation metric values obtained without noise and Table 2 gives the values with noise. The noise added is the salt and pepper noise. GMR and FMR values are plotted in Figures 8 and 9, respectively.

Inferences from the Tables 1 and 2, Figures 8 and 9 are as follows.

- Table 1 gives the evaluation metric values of GMR and FMR obtained without noise and Table 2 gives the obtained values with noise.
- Figure 8 gives the GMR plot and Figure 9 gives the FMR values.
- The values are obtained by varying the key word size and values are taken for key word size at 4, 6, 8 and 10.

Tab. 1. Evaluation metrics obtained without noise

		Secret Key Size			
		4	6	8	10
Evaluation	GMR	91%	91%	91%	91%
Metrics	FMR	09%	09%	09%	09%

Tab. 2. Evaluation metrics obtained under noise

		Secret Key Size			
		4	6	8	10
Evaluation	GMR	84%	84%	84%	84%
Metrics	FMR	16%	16%	16%	16%

- All cases irrespective of neither the key size nor noise, obtain a high GMR and low FMR which clearly indicates the effectiveness and stability of the proposed technique.

4.6. Comparative Analysis

Various papers related to biometric authentication and recognition uses different biometric modalities, extraction techniques, authentication techniques, evaluation metrics, image quality and image resolutions. Some of the techniques along with modalities and evaluation metric employed and the respective results are given in Table 3. From the table, we can see that our proposed technique has produced good results.

5. Conclusion

This paper presents a technique to generate face and palm vein-based fuzzy vault for multi-biometric cryptosystem. Input images are pre-processed using various processes to yield images fit for further processing. The technique is split into four phases: a) Pre-processing, b) Combined feature vector generation, c) Fuzzy vault generation and d) Test case. The experimentation is conducted using the palm vein and face database available in the CASIA and JAFFE database. The evaluation metrics used is FMR (False Match Ratio) and GMR (Genuine-Match Ratio) and from the values obtained we have inferred that our proposed technique have performed well.

Tab. 3. Comparative analysis for various methods

Paper	Mechanism	Modality	Evaluation Metrics
Fuzzy Vault for Fingerprints [8]	Fuzzy Vault based Authentication	Fingerprint	False Reject Rate (FRR)= 21%
			Genuine Accept Rate (GAR)=79%
Multispectral Palm Image Fusion for Accurate Contact-free Palmprint Recognition [15]	Curvelet Transform based Recognition	Palm Print	Discriminating Index (d)= 5.9608
			Equal Error Rate (EER)= 0.58%
Fingerprint-Based Fuzzy Vault:Implementation and Performance [13]	Fuzzy Vault based Authentication	Fingerprint	False Accept Rate (FAR)=0.08%
			Genuine Accept Rate (GAR)=85%
A Study of Hand Vein Recognition Method [9]	Feature Point Extraction based Recognition	Hand Vein	Pass Ratio (PR) = 99.1%
			Rejection Rate(RR)=0.9%
Our proposed technique	Fuzzy Vault based Authentication	Face and palm vein	GMR=91%
			False Match Rate(FMR)= 09%

References

1999

- [1] A. Juels and M. Wattenberg, A fuzzy commitment scheme, 6th ACM Conference on Computer and Communications Security New York, pp.28-36, 1999.

2003

- [2] C. Han, H. Cheng, C. Lin and K. Fan, Personal authentication using palm-print features, Pattern Recognition, 36, pp. 371-381, 2003.
- [3] D. Zhang, W. Kong, J. You and M. Wong, Online palmprint identification, IEEE Transactions on Pattern Analysis and Machine Intelligence, 25, pp. 1041-1050, 2003.

2004

- [4] C. L. Lin and K. C. Fan, Biometric Verification Using Thermal Images of Palm-Dorsa Vein Patterns, IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 2, pp. 199-213, 2004
- [5] U. Uludag, S. Pankanti, S. Prabhakar, and A. K. Jain, Biometric cryptosystems: issues and challenges, Proc. of the IEEE, vol. 92, no. 6, pp. 948-960, 2004.

2005

- [6] R. Dhamija and J. D. Tygar, The battle against phishing: Dynamic security skins, In Proceedings of the Symposium on Usable Privacy and Security ACM Press, pp. 77-88, 2005.
- [7] Shenglin Yang and Ingrid Verbauwhede, Automatic Secure Fingerprint Verification System Based on Fuzzy Vault Scheme, In Proceeding of IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 5, pp. 609-612, 2005.
- [8] Umut Uludag, Sharath Pankanti, and Anil K. Jain, Fuzzy Vault for Fingerprints, Lecture Notes in Computer Science, Vol. 3546, pp. 310-319, 2005.

- [9] Yuhang Ding, Dayan Zhuang and Kejun Wang, A Study of Hand Vein Recognition Method, In Proceedings of the IEEE International Conference on Mechatronics & Automation, pp. 2106-2110, 2005.
- 2006**
- [10] Lingyu Wang and Graham Leedham, Near- and Far- Infrared Imaging for Vein Pattern Biometrics, In Proceeding of IEEE International Conference on Video and Signal Based Surveillance, pp. 52, 2006.
- 2007**
- [11] P. H. Hennings-Yeomans, B. V. K. Kumar, and M. Savvides, Palmprint classification using multiple advanced correlation filters and palm-specific segmentation, IEEE Transactions on Information Forensics and Security, 2, pp. 613-622, 2007.
- [12]] D. Hu, G. Feng, and Z. Zhou, Two-dimensional locality preserving projections (2DLPP) with its application to palmprint recognition, Pattern Recognition, 40, pp. 339-342, 2007.
- [13] Karthik Nandakumar, Anil K. Jain and Sharath Pankanti, Fingerprint-Based Fuzzy Vault: Implementation and Performance, IEEE Transactions on Information Forensics and Security, Vol. 2, No. 4, pp. 744-757, 2007.
- [14] Yi-Bo Zhang, Qin Li, Jane You and Prabir Bhattacharya, Palm Vein Extraction and Matching for Personal Authentication, Advances in Visual Information Systems, vol. 4781, pp. 154-164, 2007.
- 2008**
- [15] Ying Hao, Zhenan Sun, Tieniu Tan and Chao Ren, Multispectral Palm Image Fusion For Accurate Contact-Free Palmprint Recognition, In proceedings of IEEE International Conference on Image Processing, pp. 281-284, 2008.
- 2009**
- [16] Bashir Mohammed Ghandi, Ramachandran Nagarajan and Hazry Desa, Classification of Facial Emotions using Guided Particle Swarm Optimization I, International Journal Computer and Communication Technology, Vol. 1, No. 1, pp. 36-46, 2009.
- [17] Su, C. (a), Palm extraction and identification, Expert Systems with Applications, 36, pp. 1082-1091, 2009.
- [18] Su, C. (b), Palm-print recognition by matrix discriminator, Expert Systems with Applications, 36, pp. 10259-10265, 2009.
- 2010**
- [19] Zhenhua Guo, Wangmeng Zuo, Lei Zhang and David Zhang, palmprint verification, Neurocomputing, vol. 73, no. 4-6, pp. 944-950, 2010.
- [20] David Zhang, Zhenhua Guo, Guangming Lu, Lei Zhang, Yahui Liu and Wangmeng Zuo, Online joint palmprint and palmvein verification, Expert Systems with Applications, 2010.
- [21] K. Nandakumar, A fingerprint cryptosystem based on minutiae phase spectrum, IEEE International Workshop on Information Forensics and Security, pp. 1-6, 2010.
- [22] Maleika Heenaye-Mamode Khan, Naushad Mamode Khan and Raja K. Subramanian, Feature Extraction of Dorsal Hand Vein Pattern using a fast modified PCA algorithm based on Cholesky decomposition and Lanczos technique, World Academy of Science, Engineering and Technology, vol. 61, 2010.
- [23] Peng Li, Xin Yang, Kai Cao, Xunqiang Tao, Ruifang Wang and Jie Tian, An alignment-free fingerprint cryptosystem based on fuzzyvaultscheme, Journal of Network and Computer Applications, vol. 33, no. 3, pp. 207-220, 2010.
- [24] Xiangqian Wu, Enying Gao, Youbao Tang and Kuanquan Wang, A Novel Biometric System Based on Hand Vein, In proceedings of 5th International Conference on Frontier of Computer Science and Technology (FCST), pp. 522-526, 2010.
- [25] Yiding Wang, Kefeng Li and Jiali Cui, Hand-dorsa vein recognition based on partition Local Binary Pattern, In Proceeding of IEEE 10th International Conference on Signal Processing, pp. 1671-1674, 2010.

- [26] Abhilasha Bhargav, Anna Squicciarini, Elisa Bertino, Xiangwei Kong and Weike Zhang, Biometrics-Based Identifiers for Digital Identity Management, In Proceedings of the 9th Symposium on Identity and Trust, pp. 84-96, 2010.

2011

- [27] Christopher Alvino, Christian Kohler, Frederick Barrett, Raquel E. Gurb, Ruben C. Gurb and Ragini Verma, Study of the Changing Trends in Facial Expression Recognition, International Journal of Computer Applications, Vol 21, No. 5, pp. 0975-8887, May 2011.
- [28] Zhenhua Guo, David Zhang, Lei Zhang, Wangmeng Zuo and Guangming Lu, Empirical study of light source selection for palmprint recognition, Pattern Recognition Letters, vol. 32, no. 2, pp. 120-126, 2011.
- [29] Palm print database. Online: <http://www.idealtest.org>.
- [30] Face database. Online: <http://www.kasrl.org/jaffe.html>.

NOVEL APPROACH BASED ON TOPOLOGICAL SIMPLIFICATION ALGORITHM OPTIMIZED WITH PARTICLE SWARM OPTIMIZATION

ZuKuan WEI, ZhaoXin WANG, HongYeon KIM,
YoungKyun KIM, and JaeHong KIM

ZuKuan WEI, Cloud Computing, Research Department, ETRI, Daejeon 305-700, Korea

Email: anlexwee@etri.re.kr

Abstract. The movement of people can be considered as the flow of liquid, so we can use the methods employed for the flow of liquid to understand the motion of a crowd. Based on this, we present a novel framework for abnormal behavior detection in crowded scenes. We extract a topological structure from the crowd with the topology simplification algorithm. However, a conventional topology simplification algorithm can not work well if we apply it to the crowd directly because there is too much noises produced by the random motion of the people in the original image. To overcome this, we make a step forward by optimizing this model using Particle Swarm Optimization (PSO) [5] to perform the advection of particle population spread randomly over the image frames. Then we propose two new methods for analyzing the boundary point structure and extraction of a critical point from the particle motion field; both methods can be used to describe the global topological structure of the crowd motion. The advantage of our approach is that each kind of abnormal event can be described as a specific change in the topological structure, so we do not need construct a complex classifier, but can classify the crowd anomalies dynamically and directly. Moreover, the approach monitors the crowd motion macroscopically, making it insensitive to the motion of an individual, disregarding the global movement. The result of an experiment conducted on a common dataset shows that our method is both precise and stable.

Key words: topological structure, particle swarm optimization, abnormal behavior, crowd behavior modeling

1. Introduction

Recently, there has been increasing interest in video surveillance of crowded scenes within the computer vision community. This brings many new challenges and problems, like pedestrian detection, tracking in the crowd and crowd behavior modeling. Among these applications, the central task is to automatically analyze and detect abnormal events in a crowd video.

In the field of intelligent monitoring, crowd behavior detection is quite different from individual behavior analysis. In the crowd, the complete trajectory of each individual can not be captured easily by a camera. Besides, handling every state of every person is a tough task for the device. Because of the interactions among a large number of

people in a crowd, simple analysis of individual behavior will pose more difficulties and lower the accuracy of the task. As further noted, modeling the motion of individual is neither sufficient nor efficient. We need to spend more time analyzing the behavior of the crowd globally. For instance, the trajectory-based method proposed in [14, 15] will be useless if we mainly want to detect global behaviors, such as formation/dispersion and splitting/merging in a crowd.

According to a well-established analysis of the crowd behavior model, we can divide the models into three main approaches. (1) Microscopic approach, defining pedestrians' motivation for the movement and treating the crowd behavior as the result of a self-organization process [4] (2) Macroscopic approach, which focuses mainly on goal-oriented crowds. This method does not pay attention to the motion of each person, and group habits are determined by global goals and destinations. All people in the crowd are partitioned into different groups to follow the predetermined habits. Then a macroscopic model is established [11]. For example, in [18], a dynamic texture model is established to jointly model the crowd's dynamics and appearance. This method explicitly detects both temporal and spatial anomalies in a crowded scene. (3) Hybrid approaches combining microscopic and macroscopic methods to analyze both individual behavior and the overall crowd status simultaneously. This hybrid way corrects every individual's behavior to optimize the features of global behavior. However, none of these three approaches can detect anomalies directly, and a complex classifier is necessary to obtain the final results. Unfortunately, an additional classifier brings extra difficulties to the task.

In various works, different kinds of local motion patterns are captured from the crowd as features. Following this, classifiers are trained by utilizing those features. The global crowd behavior is modeled implicitly by the classifiers. The final performance of anomaly detection is closely related to the choice of the classifier and of the training data. If we want to obtain the results directly, without constructing and training a classifier, we should try analyzing the crowd in another way. The new idea proposed to this end in the present paper is inspired by the method known as topological simplification.

We will extract the topological structure from the particle motion field with topological simplification algorithm, which can represent the global behavior of the crowd motion explicitly. Then we can monitor changes in the topological structure to detect the abnormal crowd behavior directly. Therefore, we do not need to build a classifier to train it and classify the anomalies. We also advance this hypothesis by optimizing the model using Particle Swarm Optimization (PSO) to perform the advection of a particle population spread randomly over the image frames. The population of particles is drifted towards the areas of the main image motion driven by the PSO fitness function aimed at minimizing the interaction force, so as to model the most diffused and normal behavior of the crowd. In this way, particles converge naturally towards the significant moving areas in the scene, and in particular towards the parts that are likely show a

high interaction force. So we can find that in a certain area, all of the particles we have extracted have encountered fewer interference factors. There is no need to eliminate the interferential points with velocity close to zero. Therefore, the altered topology simplification algorithm can be used directly in the area where the particles have been updated by PSO. After the PSO optimization, we can use more accurate critical points to form a topology which is closer to the natural conditions in the crowd. Then we can detect anomalies by tracking this topology because the crowd motion varies together with the topology change. The advantage of our method is that we do not need to create a complex classifier; it can detect anomalies directly by monitoring the changes in the topological structure of the crowd. Moreover, the model proposed in this paper has good robustness and it is quite insensitive to the motion of an individual that does not affect the global motion.

The remainder of paper is organized as follows. In section 2, we give a brief introduction to the particle swarm optimization algorithm and the topological representation of a 2D dense vector field. In section 3, our approach to constructing a topological structure optimized by PSO and detecting abnormal behavior using this topology are given. Section 4 provides implementation details and experiment results.

2. Related Work

2.1. Particle Swarm Optimization (PSO)

Over the recent years, the PSO (similar to Evolutionary approach) has been developing very rapidly. In computer science, PSO is a computational method that optimizes a problem by iteratively trying to improve a candidate solution with regard to a given quality measure. PSO optimizes a problem by having a population of candidate solutions, here dubbed particles, and moving these particles around in the search-space according to simple mathematical formulae over the particle position and velocity. Each particle's movement is influenced by its local best known position, and is also guided towards the best known positions in the search-space, which are updated as better positions are found by other particles. This is expected to move the swarm toward the best solutions. PSO is applied to monitor abnormal crowd behavior, mainly because the crowd is a biological system as well as a social system. More precisely; there are complex interactions between the communities and the environment, especially interpersonal interactions. Besides, PSO has originated from the simulation of a simple social system, and it can simulate unpredictable group behavior using local information.

PSO is initialized with a population of N -dimensional particles distributed randomly, and then pursues the final optimal solution by iteration. At each iteration, particles track two "Extreme Values" to update themselves. The first one, obtained by the particle itself, is called *pbest*. The other one is obtained by the whole group, and represents

the current global best value, denoted by $gbest$. Furthermore, it does not need the whole group; part of the group near the particle is enough to obtain the local extreme value. The $pbest$ value represents the position associated with the $best$ (i.e., minimum or maximum) fitness value of the particle obtained at each iteration. The $gbest$ value represents the best position among all the particles in the swarm, i.e., the position of the particle assuming the minimum or maximum value when evaluated by the fitness function. When the particle velocity changes, the particle will be updated according to the following equations [5]

$$v_i^{new} = w \cdot v_i^{old} + C_1 \cdot rand() \cdot pbest_i - present_i^{old} + C_2 \cdot rand() \cdot (gbest_i - present_i^{old}) \quad (1)$$

$$present_i^{new} = present_i^{old} + present_i^{new} \quad (2)$$

where $v_i^{new}(v_i^{old})$ is the particle velocity after(before) updating, w is the inertia weight used for balancing the local and global search in the PSO, C_1 , C_2 are the learning factors or acceleration parameters that drive every particle closer to the $pbest$ and $gbest$ values, $rand()$ is a random number between 0 and 1, and $present_i^{new}(present_i^{old})$ is the particle's updated(current) position, respectively.

2.2. Topological simplification of a sparse vector field

Topological simplification is widely used for data simplification and visualization of 2-D and 3-D velocity vector fields in fluid mechanics computations (CFD). The basic idea is to describe the structure of a dense vector field by certain special points, called critical points, and curves connecting these points. These points and curves can be used to determine qualitative behavior of the velocity field. This means that although we cannot reconstruct the original velocity field from this structure, it can be estimated up to topological equivalence, which is good enough for the analysis of the vector field behavior, especially for the particle motion field. Besides, there is a topological classification theory on the classification of these critical points based on the local structure around them. Each category corresponds to one type of anomalies that we want to model in the training phase.

According to theory of calculus, every person in the picture can be represented by a velocity vector field function $v(x, t)$, where, x are the spatial coordinates in the image, and t is the time. However, there are so many pixels in a given image that the velocity vector field will be too complex to calculate. Accordingly, we should simplify the vector field to reduce the number of its dimensions. The method based on topological simplification proposed by Helman and Hesselink [3] can be well applied here. This approach extracts a topological structure by detecting and classifying the critical points. The

flow field topology can be used to express the characteristics of the flow field structure. Subsequently, the topological analysis of the flow field is expanded from planar flow field to wide surface flow field, 3D flow field and unsteady flow field [10]. However, the initial topology analysis is not exhaustive. Some important features, such as open boundaries or an extremely isolated limiting ring, may be lost or omitted. This is why, based on the boundaries of the flow detection, Ken Wright proposed another two methods for analyzing the phase plane and the parallel vector [6, 7]. The topological structure can represent the vector field features very well, while also simplifying the vector field very well. The Topological Simplification Algorithm divides the pixels in the image into two parts: normal points, whose velocity exists, and critical points, whose velocity has vanished. The core idea of the method is to extract the topological structure of the vector field, and use this structure to describe the qualitative behavior of the vector field.

The topological structure consists of certain critical points and certain curves. A critical point is a domain in the image where the magnitude of the corresponding vector field vanishes. It is also known as a singular point, or singularity. The curves join a critical point to the next one, and divide the field domain into regions. In each region, the vector field has a different behavior. Qualitative behaviors of the velocity field can be completely determined by these points and curves. The use of such techniques significantly reduces the amount of data we need to process. In a 2-D dense field, there are six kinds of critical points (Fig. 1): repelling node, attracting node, repelling focus, attracting focus, saddle point and center point. Among them, the most special point is the saddle point, which has been proposed by Scheuermann [8]. Since the crowd cannot perform as well as the saddle, we will not discuss the saddle point any more. For each point x in the whole domain of an image $I \subset R^2$, the velocity field in the neighborhood of the critical points can be linearly approximated by the linear flow equation

$$v(x, t) = \frac{d_x}{d_t} \approx A(t)x + b(t) \quad (3)$$

where, x is the position of the critical point we are interested in, and t is the time. For each explicit vector field, we can define a flow $\phi_t : I \rightarrow I$ that can make the flow as smooth as possible, where $\phi_t(x) := \phi(x, t)$. In the above Equation, $b(t)$ is a 2D vector representing the position of the critical point in the velocity field, and $A(t)$ is the $2 * 2$ Jacobian matrix [3] proposed by Helman and Hesselink, defined as

$$A \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix} \quad (4)$$

Here u and v represent the projections of the velocity vector V on the x -axis and the y -axis, respectively. If A is invertible, then in the neighborhood of the critical point the local shape of $v(x, t)$ can be determined by the two eigenvalues of the matrix A . In

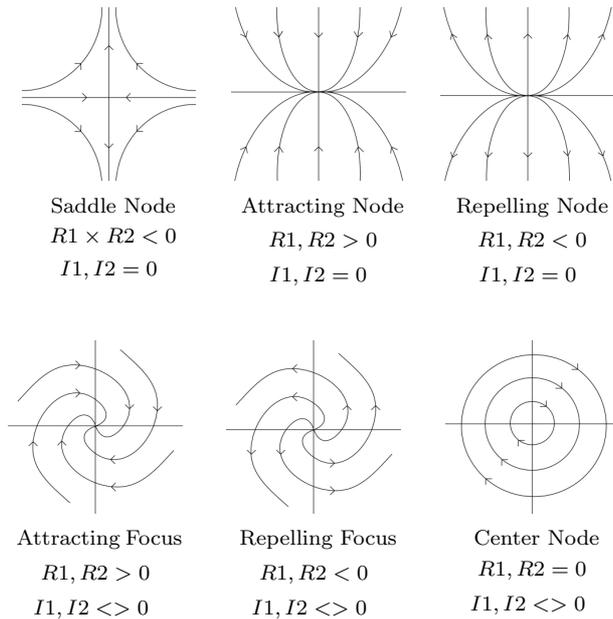


Fig. 1. Classification of critical points. $R1$ and $R2$ denote the real parts of the eigenvalues of the Jacobian, while $I1$ and $I2$ denote their imaginary parts.

Fig. 1, if both eigenvalues are real-valued and have opposite signs, this is a saddle point; if the two eigenvalues are both positive real numbers, the point is an attracting node; if the two eigenvalues are both negative real numbers, the point is a repelling node; if the two eigenvalues are conjugate complex numbers and their real parts are positive, the point is an attracting focus; if the two eigenvalues are conjugate complex numbers and their real parts are negative, the point is a repelling focus; and finally if the two eigenvalues are conjugate imaginary numbers, then the critical point is a center point.

According to the invariance of the topological structure, if a vector field is transformed by applying a continuous map to the original vector field in the neighborhood of critical points, then the type of the corresponding critical points does not change. In other words, although the velocity fields may seem to be quite different in different scenarios, the behavior of the velocity field can be characterized by the same type of critical points. The situation is like in Fig. 2, where the three critical points are of the same kind.

Thus, we can track a critical point to analyze the topology of the crowd using the method of extracting the crowd velocity field. For example, we can consider a repelling

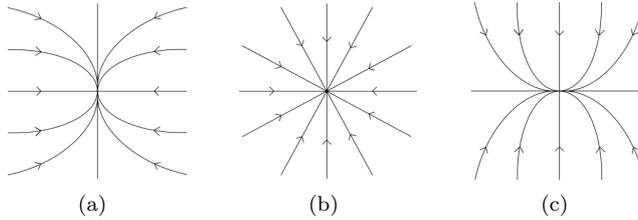


Fig. 2. Three different motion fields belonging to the same type of critical node: an attracting node.

point, which represents the situation where all the persons near the critical point are moving away from that point to all directions.

3. Our Approach

In a word, traditional topological simplification methods cannot be directly applied to monitor a video; we must find a new way to solve the problems given by the original image. In the original image, there are many points whose velocity fields are zero, such as background points. The points we obtain from the image are affected by a lot of noise caused by human arm and leg swings, and the velocity fields obtained from image pixels are noncontiguous or piecewise continuous. On the other hand, a drawback of using traditional topological simplification is only that it assumes that the crowd follows a fluid-dynamical model, which is too restrictive when modeling masses of people. Elements of the crowd may also move along unpredictable trajectories, which will result in an unstructured flow. To overcome these drawbacks, we propose a novel particle advection using PSO that can update the position of each particle to get rid of the noise directly, and form a better topological structure to be monitored. Then we can obtain the parameters of a linear stream model in strong interference with the RANSAC algorithm to determine the crowd type.

3.1. Fitness functions and particle position updates

According to previous studies, because of the fact that a pedestrian in a video will be sheltered from other people, or due to random variation in the population density or differences in the image resolution, we cannot track the trajectory of each single person in a dense crowd. On the opposite, we use the same method as proposed in [13, 16] and apply particle convection to analyze the crowd behavior in the video. We need to calculate the optical flow (OF) information to obtain the particle motion field. The OF represents the velocity distribution of the brightness mode in the image. It can

represent not only the spatial arrangement of a moving object, but also the change rate of its distribution.

Every particle extracted from every frame of the video can be considered as the clue of the driving force. Then the particle will be revalued by the fitness function just because of the driving force. This important factor was neglected by most of the previous studies. Even if the driving force was calculated just like in [13, 16], particle convection was calculated on a coarse level. The researchers imposed a grid on the original image, and used every node on the grid to represent a particle. To solve the problem, we propose a novel approach. It amounts to using PSO to improve the particle position, which can make the particle come closer to its real position. Then we can locate the anomalies based on the new result we have obtained. First of all, we define random initialization of the particles in the first frame of the video as the first input of the PSO algorithm. In order to simplify our simulation, we always choose about 1/3 evenly distributed pixels in the first frame as the first original particles. The same way of choosing a random input is applied in the subsequent experiments in this paper. From such an initial stage, we obtain the first estimate of $pbest$, and the global $gbest$ for each particle. The particles are defined by their 2D value corresponding to the pixel coordinates in the frames. At each iteration, the $pbest$ value is updated only if the present position of the particle is better than the previous position according to the fitness function. We consider the result obtained as the input for the next frame, then apply the same procedure until the video ends. Actually, the fitness function can capture the most wanted interactions among the crowd which drive the pedestrians' movement. For each particle, the fitness function is the factor which can revalue every particle using the OF. We first define the intensity of the optical flow at a given position in the image for a particle i as

$$W_i = O_{avg}(x_i^{new}) \quad (5)$$

where O_{avg} represents the average optical flow at the particle coordinates x_i^{new} . Thus, the average on can be obtained by all of the previous frames. Then we define the velocity field W_i^p for the most wanted particle

$$W_i^p = O(x_i^{new}) \quad (6)$$

where $O(x_i^{new})$ is the current OF of particle i updated by PSO. In fact, this OF value is the average value computed in a small spatial neighborhood to avoid numerical instabilities in the optical flow. Like in [16], we consider the velocity derivative $\frac{dW_i}{dt}$ as the force driving the particle, where t is the time between the current video frame f and the previous frame $f - 1$. This process is in some way mimicked by the particles which are driven by the optical flow towards the areas with larger motion in the image. In this way, the more regular the pedestrians' motion, the smaller the interaction force, since the flow of people movement varies in a smooth way. Accordingly, we define the fitness

function as

$$FitPos = \min_{i=1,2,\dots,K} \left\{ \frac{1}{\tau} W_i^p - W_i - \frac{dW_i}{dt} \right\} \quad (7)$$

where K is the number of particles we have initialized in the image. Particle position can be updated every time to bring it closer to reality.

3.2. Analysis of crowd behavior based on topological structure

The approach proposed in Section 2.2 cannot be applied to detecting anomalies in the crowd, because the motion field of the particle is sparse. In order to capture the crowd topology, we propose an improved method for determining the type of a critical point.

3.2.1. Virtual critical points

In the conventional method, the boundary points do not change over time. However, when analyzing the crowd, the situation is opposite; the boundary points might work as virtual critical points. Boundary points are the points whose velocity tends to zero, and their behavior can be described by the sets $\alpha - limit$ and $\beta - limit$, where $\alpha - limit$ ($\beta - limit$) are defined through subsets $\alpha(x)$ ($\beta(x)$) of the image domain I consisting of points $y \subset I$ such that $\phi(x) \rightarrow y$ when the time $t \rightarrow \infty (-\infty)$. In other words, the particle at position x will reach $\alpha(x)$ ($\beta(x)$) after an infinitely long time. In a sparse velocity field, the limit set can be determined by numerical integration. In order to extract virtual critical points, we need to integrate the boundary points. For this purpose, Tricoche's cluster algorithm [9] can be used. Since points in the same cluster will exhibit the same action, a virtual point can be defined at the center of each cluster. The type of each virtual point can be determined by the type of the limit set: $\alpha - limit$ and $\beta - limit$.

The topological structure consists of points and relationships between points represented by curves. We define those relationships as below. If there is a trajectory linking points in the limit sets corresponding to a sink N and a source S , we connect the sink and the source by a curve. The process can be described as follows: For every moving point x , we first find the $\alpha - limit$ set and the $\beta - limit$ set, and then the corresponding critical points $Sink_i$ and $Source_j$; following this, we add the connection relationship count C_{ij} . When we obtain the complete count set, we search for an element c which is bigger than the threshold $C_{threshold}$ in the count set; then this c is added to the topological structure. As a result, the procedure of extracting the topological structure can be defined as in Fig. 3.

3.2.2. Critical points

The boundary points have been treated above, so we will not consider them again in this section. The points that could possibly be critical are determined by the PSO at first. In the area under consideration, the particles are clustered, and definitely have non-zero-velocity. So we do not need to define a threshold to remove the points whose

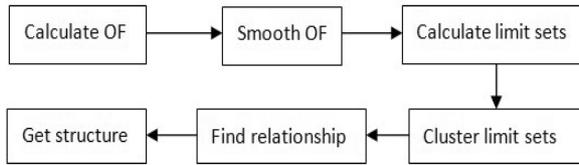


Fig. 3. Overall procedure of topological structure extraction.

velocity tends to zero. Like in [10], the original image is cut into a few grids, and the revised RANSAC [13] algorithm is applied to each area where the particles have been updated by the PSO. Then the linear stream parameters A and b can be determined. After the estimation of critical points, the critical point type can be determined by calculating the eigenvalue of the Jacobian matrix A . If a critical type which looks like the anomaly type we are interested in has just occurred, we will calculate the probability of the anomaly to reduce error accumulation. Here, we just simply define the probability as the percentage of points moving abnormally, i.e. $P(abnormal) = N_{total}/N_{abnormal}$. Whether a point is an abnormal point or not can be determined based on the differences between the calculated results and the estimated results for the point's velocity vector. If the difference is bigger than the threshold, the point can be considered an abnormal one. The overall estimation procedure is listed in Fig. 4.

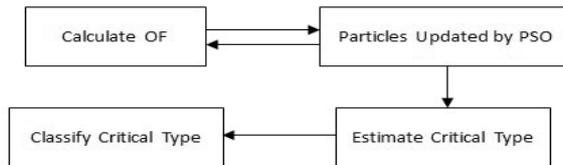


Fig. 4. Overall procedure of critical points estimation.

3.2.3. Detection of anomalies

Based on the theory above, the topological structure of the crowd can be described by particle motion field. Accordingly, to monitor the crowd status, it suffices to track the topology. If the topology changes, there must be something happening in the crowd. For the holistic behavior case, the gathering of the crowd can be approximated by a sink in the particle motion field, and the dispersion — by a source. If these structures are detected frequently during some time interval, the corresponding event occurs. For example, as shown in Fig. 5, when the crowd moves as a single entity at first, and then splits into two flows at some moment, then the sink set of the extracted structure will

split into two sinks. Finally, the original source-sink will become two separate source-sinks.

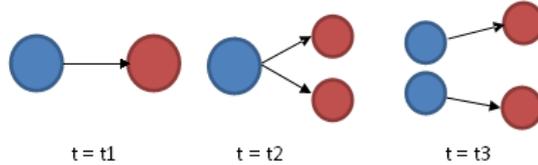


Fig. 5. Example of a crowd as one entity splitting in two. The blue nodes represent sources, the red ones — sinks. The structure will change when the crowd splits.

If the proposed approach is used here, topological invariance can be inherited, and it is not sensitive to noise, while it is loosely related to the individual's motion status. Macroscopical monitoring of the crowd motion appears to be both precise and stable.

4. Experimental Results and Comparisons

The purpose of the approach proposed in this paper is that anomalies in the crowd can be extracted and classified in a precise way. In order to evaluate the method we have proposed, we will consider a few existing methods and compare their results on standard common datasets used for detection of abnormal behavior.

We partition the image into 88 grids and use Black's optical flow algorithm to calculate the velocity of the crowd motion, keeping 25% pixels of the original image to serve as the random input to the PSO. Our experiment is conducted on the abnormal behavior dataset, UMN or PETS. There are two different University of Minnesota scenarios in the UMN dataset, and the total of eleven videos for testing. Pedestrians wander all over the scene for a while, then escape. Abnormal activity can be defined as the escaping motion. As Fig. 6 shows, the left frame is normal, and the right one — abnormal. In order to test the validity of our approach, based on extracting the topological structure, we conduct an experiment based on PETS. Fig. 7 below shows the structure we extract first, and the structure after PSO reconstructed smoothly using RANSAC, which is similar to the original one. When the people in the crowd escape, we can easily find there is high probability of change in the extracted structure.

Fig. 8 is the result we obtain when detecting anomalies in the crowd based on Fig. 6. In Fig. 6, people first wander, and then escape to every direction at an uncertain time. Here, we use the way we have proposed above to detect the time when the abnormal motion began. In Fig. 8, the horizontal coordinate denotes the time, the vertical coordinate represents the probability of anomalies, and the dotted line denotes the threshold determined by experience. To obtain a better result, we build a filter whose average

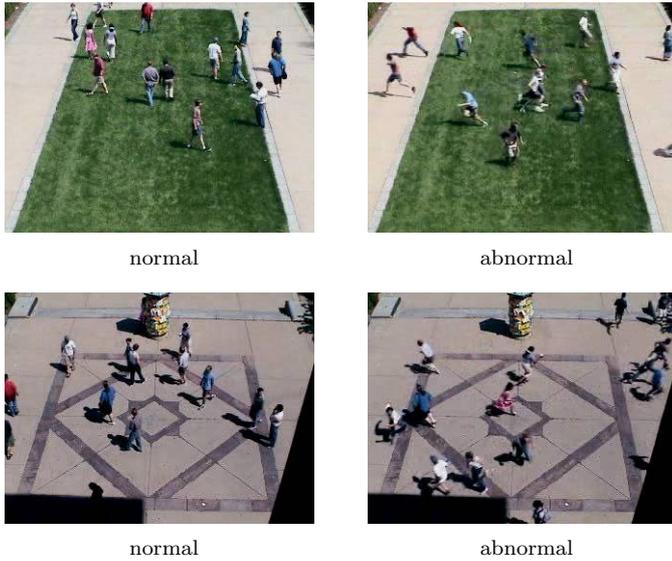


Fig. 6. Sample frames in two different scenes from the UMN dataset: Normal (left) and abnormal (right).

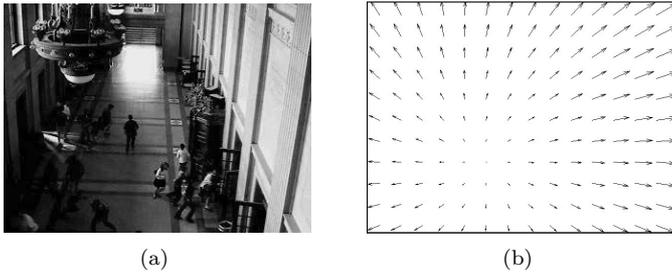
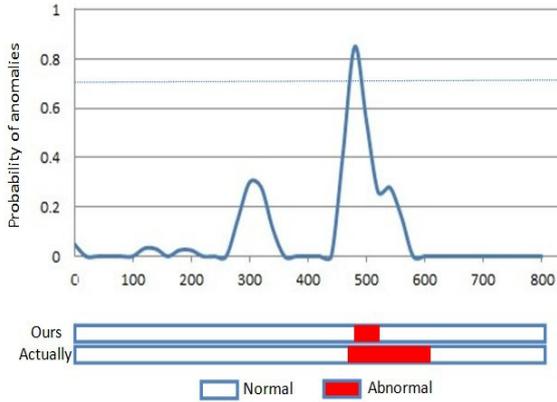
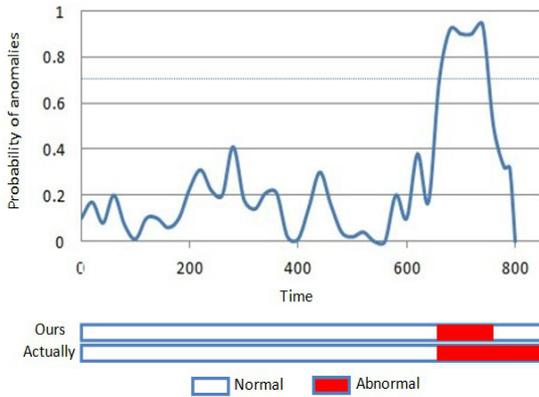


Fig. 7. (a) is the original, and (b) is the structure we extracted from (a) after PSO and RANSAC revised.

window size is 25. Then the result we obtain is handled by a filter. If the value of the result is beyond the threshold, the given frame is most likely an abnormal one. The white part of the strip-chart at the bottom of Fig. 8 denotes the normal frames of the test sequence, and the red part represents the abnormal frames. Hence we can see that the approach we have proposed works well in classifying the abnormal frames in the whole video sequence. Also, the time when attack/repelling anomalies happened can be easily found in this way.



(a) UMN sample 1



(b) UMN sample 2

Fig. 8. Two examples of detecting dispersing motion of the crowd based on the common dataset UMN. The dotted line determined by experience denotes the threshold.

We extract the topological structure from the sample frames just as shown in Fig. 9. If the pedestrians split into two independent smaller groups, topological structure (a) changes to (b). Actually, we can find that the 95th frame is just as (d) shows, and there are only two groups in this frame. This proves that topological structure can represent the real situation in the crowd very well. To verify our approach effectively, we reemploy the classical methods: optical flow [2], social force and MDT(mixture of

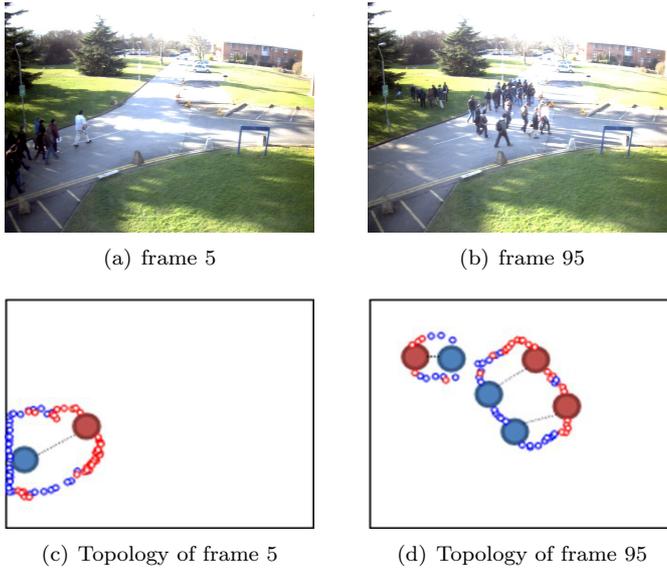


Fig. 9. The structure extracted from an image in the PETS dataset. The blue nodes in (c),(d) are sources, and the red ones — sinks. The smaller blue and red circles denote the α - limit and β - limit sets. The dashed lines denote the relationship between the source and the sink.

dynamic texture) [18] on the PETS dataset. Each of the other three methods has been proved effective in detecting the splitting/emerging anomalies of the crowd. Moreover, they analyze the crowd's behavior at the macroscopic level, just as our approach.

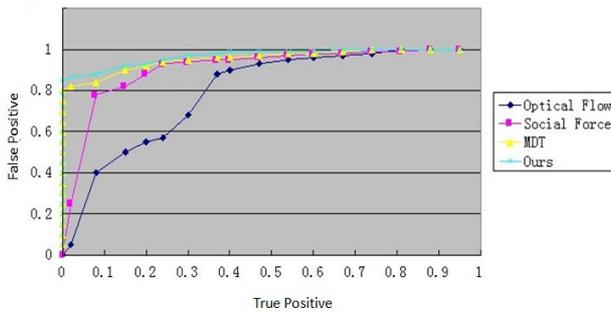


Fig. 10. ROC curves of the three existing methods and of our method.

Tab. 1. The area under ROC of the four approaches.

Approach	Optical Flow	Social Force	MDT	Ours
AUC	0.832	0.955	0.983	0.992

Fig. 10 shows the ROC curve of our method for the frame level anomaly detection on the PETS dataset, while Table.1 shows the quantitative results of the method compared to all three other methods.

Experimental results show that our approach outperforms the optical flow method, the social force method and the MDT method. Moreover, we discover that we do not need to analyze every frame of image data from the video like in the other approaches - for only half of the frames are enough for detecting an anomaly. If we do not use the PSO algorithm, using 2 frames we can also determine the crowd type with some loss of accuracy. On the other hand, there is no need to construct a complex classifier to classify the anomalies in our method, so our approach to finding the time when splitting/emerging anomalies (or other anomalies) occurred is much faster and simpler.

Actually, the method described in this article can be easily extended to detecting some other anomalies of the crowd, such as abnormal speed. Since the average velocity of the crowd is proportional to the Euclidean norm of the matrix A of the linear flow function, we have $v_{crowd} \propto \|A\|_2 = \sqrt{\lambda_{max}(A^T A)}$, where λ_{max} is the biggest eigenvalue of matrix A . Here we have used another common dataset, PETS, to testify the effectiveness of the algorithm. In this video sequence, the pedestrians walked into the screen from the right hand side, and then began to run away at a certain frame-time. And then they walked in from the left hand side, and began to run away again at another certain frame-time. What can we define about the abnormal activity is that the crowd is moving at a faster rate. Fig. 11 is a shortcut of this experiment. In order to show the performance of our approach, a comparison test with HOG algorithm [17] will be carried on.

Fig. 12 is the result of abnormal speed detection, where the solid line is the Euclidean norm of the matrix A of the linear flow function. The dotted line is a preset threshold, which is configured according to the experience and convenience. If the norm is beyond the threshold, we believe that the crowd may be moving at a faster rate. The strip-chart in Fig. 12 represents the comparison of results obtained with our method and hand-labeled ones. We can see that the Euclidean norm can basically reflect the average speed of the crowd. Our method can be a reliable way of monitoring anomalies in the crowd movement speed.

Table 2 shows the comparison of the performance of the HOG algorithm and our method in velocity anomaly detection, where: TP is the true positive, which represents the number of correctly detected frames among the abnormal frames obtained in the experiment; FP is the false positive, denoting the number of wrongly detected frames

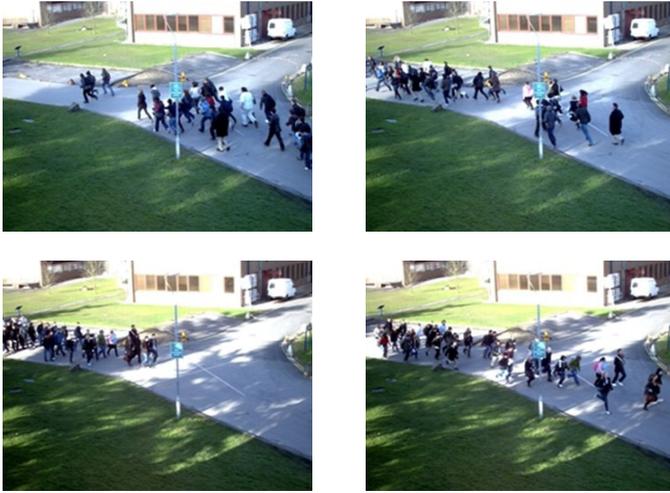


Fig. 11. The structure extracted from an image in the PETS dataset; the blue nodes in (c),(d) represent sources and the red ones — sinks. The smaller blue and red circles represent the α - limit and β - limit sets. The dashed lines denote the relationship between source and sink.

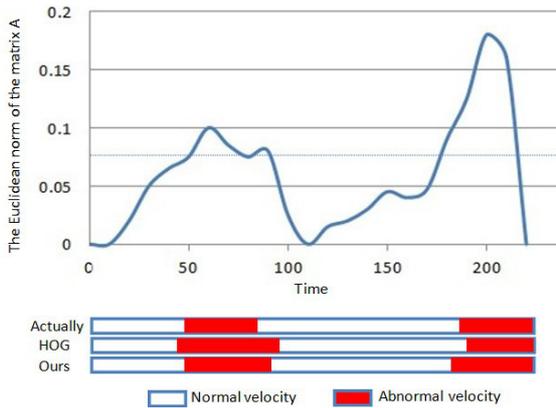


Fig. 12. Result of abnormal velocity detection.

among the abnormal frames obtained in the experiment. The sensitivity is defined as $TP/(TP + FP)$. Compared to HOG, our approach is more accurate. Accordingly, we can employ this improved method to detect the velocity anomalies in a crowd very conveniently and easily.

However, it inevitably also has some drawbacks. In order to avoid building a complex

Tab. 2. Comparison of the two methods with regard to detecting abnormal crowd velocity.

Methods	TP	FP	Sensitivity
HOG	198	22	0.90
Ours	202	18	0.92

classifier, we often use a threshold defined by experience to determine directly whether the crowd is abnormal. Moreover, the saddle point type is not considered in this paper which is a blind spot. This is a very arbitrary behavior. However, the goal of this paper is not to assign blame but to identify important areas for further projects. The experiments have shown that the results are often very close to the real situation, so this method is feasible.

5. Conclusion

In this paper, we have proposed a new approach, based on a topological simplification algorithm, to detecting anomalies in a crowd. To make the conventional topology simplification algorithm applicable directly, we use the PSO algorithm to improve the position of the particle at the next frame. Moreover, use of the PSO can make the particles cluster in a certain calculation area, where the particles meet in a dense state. After process adjustment with the PSO, we obtain a typical topological structure, which consists of critical points and a relationship described by curves linking them. In consequence, each type of crowd anomaly can be represented by a special change in the structure. We can monitor the crowd motion macroscopically instead of analyzing each person. Unlike other known methods, this method does not require constructing a complex classifier, which means that we can skip the learning phase and classify the anomalies of the crowd dynamically and directly. The results are demonstrated for four cases of the crowd motion: formation/dispersion and splitting/merging, and we believe that this approach can be applied to other types of variations, including more general motion.

References

1981

- [1] M. A. Fischler and R. C. Bolles: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6), 381-395.
- [2] B. K. Horn and B. G. Schunck: Determining optical flow. *Artificial Intelligence*, vol. 17, Issues 1-3, 185-203.

1991

- [3] J. Helman and L. Hesselink: Visualizing vector field topology in fluid flows. *IEEE Computer Graphics and Applications*, 11(3), 36-46.

1995

- [4] D. Helbing and P. Molnar: Social force model for pedestrian dynamics. *Physical Review E*, 51: 4282.
- [5] James Kennedy and Eberhart Russell: Particle swarm optimization. *Neural Networks. Proceedings. IEEE International Conference on*. Vol. 4.

1998

- [6] D. N. Kenwright: Automatic Detection of open and closed separation and attachment lines. *Proc. IEEE Visualization'98*, 151-158.

1999

- [7] D. N. Kenwright, C. Henze and C. Levit: Feature extraction of separation and attachment lines. *IEEE Transactions on Visualization and Computer Graphics*, 5(2), 135-144.

2000

- [8] G. Scheuermann, B. Hamann, K. I. Joy, et al: Visualizing local vector field topology. *Journal of Electronic Imaging*, 9(4), 356-367.
- [9] X. Tricoche, G. Scheuermann and H. Hagen: A topology simplification method for 2d vector fields. In *Visualization 2000. Proceedings*, 359-366.

2001

- [10] T. Wischgoll and G. Scheuermann: Detection and visualization of closed stream lines in planar flows. *IEEE Transactions on Visualization and Computer Graphics*, 7(2): 165-172.

2002

- [11] R. L. Hughes: A continuum theory for the flow of pedestrians. *Transportation Research Part B: Methodologica*, 36(6), 507-535.

2006

- [12] Xiaoshan Pan, Charles S. Han, Ken Dauber and Kincho H. Law: Human and social behavior in computational modeling and analysis of egress. *Automation in Construction*, 15(4), 448-461.

2007

- [13] S. Ali and M. Shah: A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. *Computer Vision and Pattern Recognition, CVPR'07, IEEE Conference on*, 1-6.

2008

- [14] A. Basharat, A. Gritai and M. Shah: Learning object motion patterns for anomaly detection and improved object detection. In *Computer Vision and Pattern Recognition. CVPR 2008. IEEE Conference on*, 1-8.

2009

- [15] J. Kim and K. Grauman: Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2921-2928.
- [16] R. Mehran, A. Oyama and M. Shah: Abnormal crowd behavior detection using social force model. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 935-942.
- [17] C. Garate, P. Bilinsky. and F. Brernond: Crowd event recognition using HOG tracker. In *Proceeding of the 12th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance(PETS-Winter)*. Snowbird, USA, 1-6.

2010

- [18] V. Mahadevan, W. Li, V. Bhalodia and N. Vasconcelos: Anomaly detection in crowded scenes. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1975-1981.

Instruction for Authors

Submission of a paper for publication in MG&V is taken to imply that the paper was not previously published (with possible exception of selected conference papers), and neither has been or is being submitted for publication elsewhere. In addition, it is also taken to imply that permission for publication, if needed, has been granted by the appropriate sources. If the manuscript contains any "copyrighted material", the task of obtaining the necessary permissions from the copyright owner(s) is the author(s)'s responsibility.

No short communications are accepted, as MG&V publishes full papers only.

The author actually submitting the manuscript must enclose a note (cover letter) saying that he is submitting a paper with the listed co-authors. Such note is taken to imply that he has obtained their consent for submitting the paper. One of the authors (e.g., the submitting one) should be indicated as the contact person. Otherwise, all correspondence will be sent to the author listed as the first one.

The language of the journal is English.

The papers should be submitted to be considered for publishing in electronic form in one of commonly used formats (.pdf, .doc, .docx, .odt), directly to the Editorial Office e-mail address: MGV@sggw.pl.

Preparation of manuscripts

The layout of the text should contain sufficient spacing and margins. Figures and photos should be included in the text. Colour figures and photos are admissible, provided that colour is essential for the presentation,

The number of pages is limited to 25. The Editorial Office may relax the limit in justified cases.

The paper's title, the authors' names, affiliations and addresses, as well as possible thanks and information about funding, should be placed on a separate page, and the text proper should not contain any such elements in page headers or footers. All the necessary additional elements will be added to the paper's final version following acceptance.

Refereeing process

All papers will be evaluated by one of the Editors, and refereed by three specialists in the given area. The refereeing process ensures anonymity of both referees and authors. Suitability for publication will be judged on the basis of the paper's relevance to the overall aims and scope of MG&V, as well as its quality: originality, technical quality, importance and readability.

Acceptance for publication and preparation of the final version

The authors are informed about the evaluation results and the Editors' decision by e-mail.

The acceptance is conditional on fulfilling further requirements concerning preparation of the final version.

The authors informed about the conditional acceptance are asked to submit the final version of the paper prepared with LaTeX according to the instructions available in the MG&V website: <http://mgv.wzim.sggw.pl>. The bibliography style should also be compliant with the MG&V standard described there. If the text requires nonstandard LaTeX styles or TeX macro definitions, they should be provided too. Nonstandard fonts should not be used without prior arrangement with the Editor. The figures should be sent in separate files, in the .EPS format. It is important that black and white, grey-level and colour figures are prepared in their respective formats, so that no ambiguity arises whether a given figure should be printed in colour or not. In addition, the authors should include a hardcopy-equivalent version of the paper in the .ps or .pdf format for reference purposes.

Together with the final version, the authors should also send their short scientific CVs and photographs (300x400 pixels at a minimum) – or declare explicitly that they do not wish to publish the above.

All the files are to be sent by e-mail to MGV@sggw.pl. In addition, a filled-in and signed copyright transfer form supplied by MG&V (together with confirmation that the paper has not been published elsewhere) should be sent back by ordinary mail.

Each paper undergoes a final correction of English in a limited extent. However, it is the authors' responsibility to make sure that the manuscript is written in clear, correct English. In case of a very low standard of English, the paper will be sent back to the authors for correction, which will considerably delay its publication.

The Editorial Office reserves the right to introduce certain changes aimed at preserving a uniform style of the journal's contents. The authors do not receive any galley proof for acceptance.

Published MG&V issue

After the issue containing the paper is published, each author obtains the author's copy. For this purpose, all authors are requested to provide the Editorial Office separately with the mail address for sending the copy, otherwise all authors' copies will be sent to the first author. The Editorial Board does not send any preprints or electronic versions.

The list of contents of the issue is placed in the MG&V website. Selected papers can also be placed in this website, according to the Editor's policy.

M a c h i n e G R A P H I C S & V I S I O N

CONTENTS

Vol. 23, No. 1/2, 2014

- 3 **Hast A., Marchetti A.** : Improved Illumination Correction that Preserves Medium-Sized Objects
- 21 **Stepień C.** : A method of constructing phyllotaxically arranged modular models by partitioning the interior of a cylinder or a cone
- 37 **Baataoui A., El Akkad N., Saaidi A., Satori K., Masrar Med.** : Robust Method for Camera Self-Calibration from Unkown Planar Scene
- 59 **Torkhani F., Wang K., Chassery J.-M.** : A Curvature-Tensor-Based Perceptual Quality Metric for 3D Triangular Meshes
- 83 **Paradowski M., Durak M., Broda B.** : Bag of Words – quality issues of near-duplicate image retrieval
- 97 **Lalithamani N., Sabrigiriraj M.** : Technique to Generate Face and Palm Vein-Based Fuzzy Vault for Multi-Biometric Cryptosystem
- 115 **Wei Z., Wang Z., Kim H., Kim Y., Kim J.** : Novel Approach Based on Topological Simplification Algorithm Optimized with Particle Swarm Optimization

Printed in Poland

ISSN 1230-0535

Abstracted/indexed in:

ACM Computing Reviews,
ACM Guide to Computing Literature,
BazTech,
COMPENDEX (Engineering Information),
EBSCO Computers & Applied Sciences Complete,
INSPEC Information Services,
PASCAL,
SCOPUS (Elsevier),
VINITI Abstracts Journal (Referativnyi Zhurnal),
International Directory of Design (Penrose-Press).
