

Optimization Model for Face Detection in Video Sequences

Fernández Fabián, Karla Marie^{1,1}, Machaca Arceda, Vicente Enrique², Gutiérrez Cáceres, Juan Carlos³, Rivera Tito, Jorge Julian⁴

¹karla.m.f.f@gmail.com, ²enriquefirst@gmail.com, ³jcgutierrezc@gmail.com, ⁴titex777@gmail.com,
¹Ms, ²Ms, ³EdD; Universidad Nacional de San Agustín

Abstract– This paper aims to provide a contribution to facial recognition systems with video cameras in real time, as the most common problems and their solution methods are identified to minimize errors and improve quality in both final results percentages certainty in detecting and minimizing processing time and computational cost, because thanks to the parallelization process that offers CUDA, will be easier to optimize the results

Keywords-- Resolution enhancement; video; MR; kernel; recognition of key points; face detection; GPU; parallel computing; computational cost..

Digital Object Identifier

(DOI):<http://dx.doi.org/10.18687/LACCEI2016.1.1.125>

ISBN: 978-0-9822896-9-3

ISSN: 2414-6390

Optimization Model for Face Detection in Video Sequences

Fernández Fabián, Karla Mariel¹, Machaca Arceda, Vicente Enrique², Gutiérrez Cáceres, Juan Carlos³, Rivera Tito, Jorge Julian⁴

¹karla.m.f.f@gmail.com, ²enriquefirst@gmail.com, ³jcgutierrezc@gmail.com, ⁴titex777@gmail.com,

¹Ms, ²Ms, ³EdD; Universidad Nacional de San Agustín

Abstract– This paper aims to provide a contribution to facial recognition systems with video cameras in real time, as the most common problems and their solution methods are identified to minimize errors and improve quality in both final results percentages certainty in detecting and minimizing processing time and computational cost, because thanks to the parallelization process that offers CUDA, will be easier to optimize the results.

Keywords– Resolution enhancement; video; MR; kernel; recognition of key points; face detection; GPU; parallel computing; computational cost.

I. INTRODUCTION

In recent years the advancement of technology has increased surprisingly incorporating the identification of persons within the image processing and video as a valuable source of information, and also that associated with the human face have become a form of mass data storage [1], against which it arises the need for detection systems and automated recognition of faces [2], reliable and low margin of error. This project aims to solve the problem standardize and improve the input information systems face recognition with a parallel optimization to ensure better output data and achieve face recognition in real time.

II. DEVELOPMENT OF CONTENTS

A. Problems face detection

Within the field of computer vision and vision problems are most clearly marked as shown in Figure 1, because of its importance as main input information to various detection and recognition algorithms.

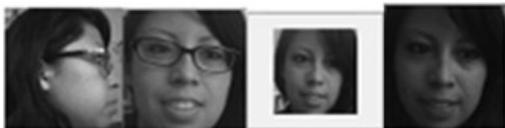


Fig. 1. Various problems in face detection.

Within the detection and face recognition, the large part is due to work with video sequences, as detailed below:

1. Incorrect detection: For the variation in approach and/or rotation may be different from the optical axis of the video camera and position of people [15] also have defects that can be noise during recording, affecting video frames.

Digital Object Identifier (DOI):

<http://dx.doi.org/10.18687/LACCEI2016.1.1.125>

ISBN: 978-0-9822896-9-3

ISSN: 2414-6390

2. Occlusion: Due to the presence of beard, hair, glasses, and other factors that stand in front of the faces obstructing visibility such as: birds, branches, bodies or other objects [16].

3. Lighting: Due to the different lighting conditions that can affect the rate of sharpness of the face, causing confusion or hiding the face or part of the face under shadows that hinder the full face detection or causing confusion algorithms face detection [17].

4. Variation in resolution videos: Currently there are different types of video cameras, which also have different levels of recording resolution [18], which has many classifications of videos that cannot control algorithms and significantly influences the results.

5. Computational cost: recognition processing becomes heavy due to the amount of frames generated for each video.

6. Time: Due to the amount of information and robustness of each of the algorithms used in the systems of recognition of faces, it adds up making the runtime extra time, avoiding abilities to obtain results in real time [10].

B. Face detection

Now face recognition video was one of the new fields where it is dabbling biometrics [13], where the properties of the video allow some considerations with the movement sequence of images that make up the video, allowing more feasible way the location of moving objects in the image, thanks to frame the difference that exists between the images. For video processing, there are various techniques, including one that measures variations in vertical and horizontal to find the eyes [14].

On the other hand, to find the edges of the face and body, uses a spatial-temporal filters Gaussian.

C. Super-resolution

Reconstruction Super-resolution (SR) is based on a fusion of several low-quality images (LR) which should provide a higher quality output with better optical resolution [20]. Even cases where LR tend to be noisy and/or fuzzy real [21] image are presented.

For this reason the SR can perform a reconstruction of one image and the reconstruction of several images (most used in case of videos).

D. Parallelization process

In recent years the graphics processing unit (GPU) became a programmable processor, multi-core and

parallelizable, is examining and exploiting the computing power that keeps the GPU, since it has been concluded that the time GPU performance is better than the time it takes the CPU as seen in Figure 2, but both get equal power consumption [19].

In 2006 NVIDIA Corporation released CUDA (Compute Unified Device Architecture) with a programming model with its accompanying API, which allow harness the power of GPU computing with minimal learning effort.

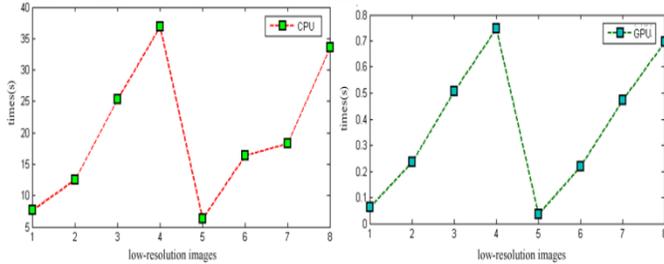


Fig. 2. Time difference with GPU parallelization processes.

III. PROPOSAL

This paper aims to provide a contribution to systems face recognition with a focus on video-security cameras institutions, in order to perform real-time processing, taking into account the current problems of poor and very poor resolution obscures identify the faces of people, is that a method that will provide lower computational cost, improved resolution video and face detection, making each of these processes simultaneously, thanks to the classifiers arises waterfall and the use of GPU for better understanding the flow of each component of the project in figure 3 which will provide greater impact in the field of public safety by providing better recognition results of criminals, for which the following are considered demonstrated steps:

a) Standardization video, with improved lighting and increased resolution

b) Face detection, thanks to the feature extraction [8] and data classification with the classification algorithm "Haar Cascade" Viola and Jones [9], [10] which allows a better filtering detection layers.

c) Reduction of the computational cost in processing through the CUDA architecture enables parallel processing that leverages the great power of the GPU (Graphics Processing Unit) as shown in the paper Marwa [11] to improve processing time both normalization and detection.

Where first tests will be performed with different algorithms and methods, then the best methods and algorithms according to the results will be implemented in processing time and percentage of certainty in detecting faces and then implement improvements and standardization video (lighting and resolution) will proceed to implement them CUDA to maximize their ability to multi-processing.

Then will be detailed in Figure 3 each more detailed diagram step, i.e. first normalization that includes improved resolution and lighting and their individual operation within a process GPU and then the next block corresponding to the detection of faces that will be parallel to the super resolution for each video stream presented.

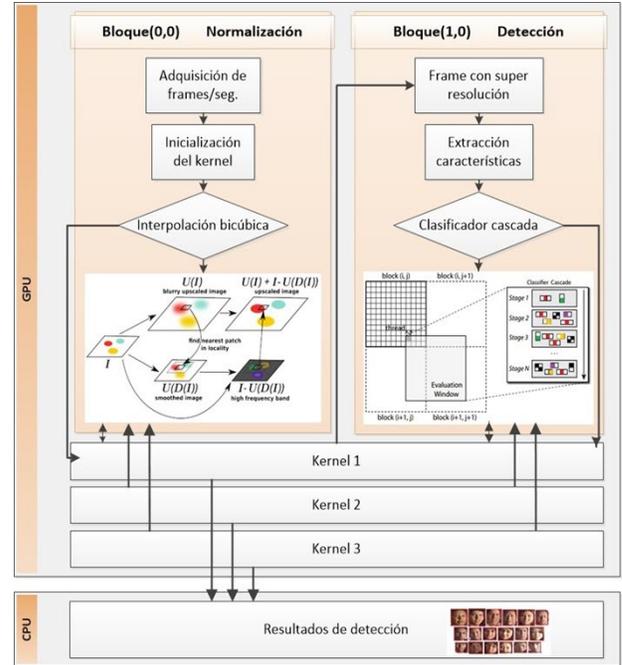


Fig. 3. Communication diagram for each project component.

A. Super-Resolution and standardization of video sequences

A classic form of super resolution is resizing the kernel, using the non-adaptive interpolation [3], as this approach interpolation treats all pixels in an image alike, where normally the location of the pixel is multiplied with the function kernel and then multiplied by the known discrete sample; the same process is repeated for all pixel locations.

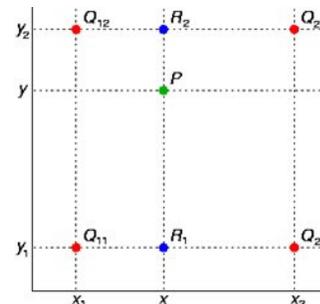


Fig. 4. Red dots: known data, green point: Point to interpolate. (Excerpted from Wikipedia bilinear interpolation).

Then adjacent pixels ranging from 0 to 256 are used for bilinear interpolation which determines the value of the gray

level of the weighted average of the four nearest pixels and assigns that value to the output coordinate. It can be seen in figure 4, this interpolation is performed in one direction and back to the other [4].

The Super Resolution (SR) reconstruction is represented by equation (1).

$$Y = DBM + n \quad (1)$$

Where:

D = sampling operators, B = blur, M = deformation, n = noise (Gaussian white noise commonly additive).

Where each is also correspondingly observed image "HR" high resolution desired.

Then follows the similarity nonlocal descriptor where the auto nonlocal similarity information is useful for improving the quality of the reconstructed images, where [5] proposes a method that builds the nonlocal similarity descriptor to present the structural characteristics of the image and helps to predict global and local information containing movement between continuous frames, in addition to minimizing search time in an image rather than the entire sequence of video images.

The local movement is used to represent the motion information more accurate to better illustrate the details of the descriptor of nonlocal similarity, in Figure 3-3 (a) a red and a green squares are shown, then they are transferred to new destinations with a single motion vector from a time t, to time t+1, and then from time t+1 to t+2, the red, green and blue squares move to nearby places with three different movement vectors, which are given from local movements of different objects in the image. This behavior often occurs irregularly throughout the video sequence for that reason it is called nonlocal similarity, note that is very relevant the estimate of movement, as well as shown in part (b) of Figure 5, the predetermined threshold is a good way to estimate or update the missing pixel value with the value of weighted fusion of similar pixels wanted.

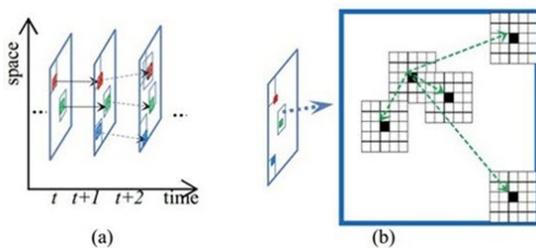


Fig. 5. (a) Global and local movement simultaneously. (b) Motion vectors. Extracted from [5].

B. Face detection

Face detection is given according to the location of interest points, basically it consists of the following steps:

1. Detection and segmenting regions of the skin.
2. Check each component of the face (mouth, nose, eyes, and eyebrows). Where a bounding box of the face detected

depending on the anthropometric form of a human face is marked.

Then a deep cascade evaluation is made with a group of classifiers organized in stages containing filters [6], in order to detect faces in multiple locations and different sizes. Also the kernel not only have to test all pixels, but also all the possible dimensions, as we see in Figure 6. Moreover the data structure cascade is constrained by dependencies between stages, and should be evaluated sequentially for each candidate [7].

It is understood that the normalization (super resolution and lighting) will work in a block (0.0), while it in parallel will work in the block (1.0) with face detection. As CUDA consists of functions running simultaneously in several very lightweight threads on the GPU, these threads have a hierarchy: the blocks, which are a set of threads in 1, 2 or 3 dimensions, and then this one grid which is a set of blocks that can be 1 or 2 dimensional [12]

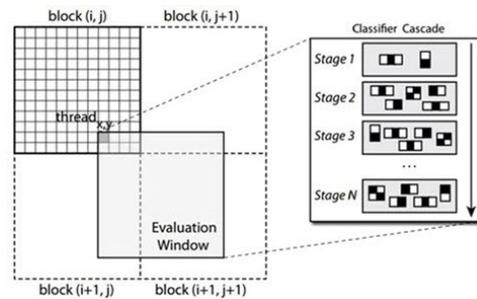


Fig. 6. Evaluation cascade classifier block. Excerpted from [7].

Following an assessment of various techniques, models and face detection algorithms, the table where I compare the results of some studies on face detection is performed is obtained.

TABLE 1
COMPARATIVE ANALYSIS OF METHODS OF VIDEO ON FACE DETECTION.

Method	Data base	Accuracy	Real time
Optimized parallel implementation of face detection based on the GPU component WaldBoost. [11]	45	16 times CPU	X
Classifier cascade mode, Classifier, "Haar" [9]	26,338	97.69 %	
ENCARA2: Real-time detection of multiple faces in different resolutions in video streams, based on skin color, Rowley-Kanade detector and Haar. [10]	26,338	91.92%	
HMME optimized with API VFW (Video for Windows) for video acquisition. [22]	20	95.98%	X
NVIDIA GeForce 8400GS GPU using a real video sequence with 2 frames per second. [23]	100	77%	X

IV. EVALUATION

For assessment and verification of variety of processing time face detection, it was based on a laptop with features: Intel core i7 processor with 2.6GHz processor speed.

A. Database

The database used in the tests are taken from a video surveillance camera of one of the rooms at the Catholic University San Pablo with an average of 30 frames per second which indicates a percentage of accepted resolution for best results.

B. Experiments

The first processed video lasts 8 seconds Figure 7, here the camera recorded a person getting into a laboratory, tested algorithm is Kanade-Lucas-Tomasi (KLT) verified that the detection key points is optimal, since the individual is to face directly to the camera which facilitated the location of points and their respective monitoring and we had good results until we got a face occlusion by hair and lighting.



Fig. 7. Detection and location of points of interest person 1.

The following video lasts 5 seconds Figure 8, which is already cut at the time of action of a person entering the laboratory, tested algorithm is the KLT where detection of points of interest is optimal since the person enters his face directly into the camera which facilitated the location of the key points marked in white and while walking, the detector follow the points and variation of movement for each frame of video.

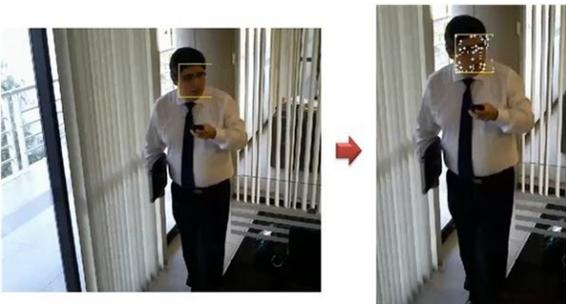


Fig. 8. Detection and location of points of interest person 2.

In Figure 9, the detection of only 2 faces on a video that was processed and improved the resolution to 700 x 497. Although shown improvement, the third faces detected in the image involved, tested algorithm is Kanade-Lucas-Tomasi (KLT).



Fig. 9. Improved image resolution and face detection.

Tests with different video features were made. A low resolution video (535 kbps with 10 frames per second) is shown in Figure 10, here we didn't detect faces but we could see an improvement in video quality.



Fig. 10. Improved video resolution of 10 fps and 535 kbps.

While the test performed on a video detector faces (987 kbps with 20 frames per second) also got confused with the train window frame, marking him as a false positive as shown in the Figure 11.



Fig. 11. Improved video resolution of 20 fps and 987 kbps.

It conducting more experiments with different videos with variation in resolution and different sizes, it concluded that the resolution increases in a low percentage when videos of very low resolution are processed, such as video in particular has very few pixels, and the data flow is also very low, ie, it is a very poor quality video. After making improved video resolution and after entering the detector face, showing detection errors, because the algorithm could not detect all

faces and generate confusion with pixels as shown in the background of the scene in the Figure 9 and 10.

V. PRELIMINARY RESULTS

To improve the resolution was tested with a database 50 videos, with different duration, the super-resolution was performed on normal processing and optimized GPU as shown in Table II below, where the result shown 4 videos random.

TABLE II.
RESULTS OF PROCESSING WITH SUPERRESOLUTION

<i>T. process</i>	<i>GPU time</i>	<i>Video normal</i>	<i>Super resolution</i>
8 min.	2.39 min.	16 frame/s.	34 frame/s.
6 min.	1.56 min.	14 frame/s.	29 frame/.
5 min.	1.12 min.	14 frame/s.	27 frame/s.
2 min.	13 seg.	7 frame/s.	13 frame/s.

As for the results in terms of face detection and tracking of points of interest are summarized in Table III by comparing the time it took each video to process, taking into account the duration and "frames" per second that contains each video, remaining as implementation, implement the method in GPU to improve processing time.

TABLE III
IMPROVED RESULTS RESOLUTION 3 VIDEO RANDOM
DATABASE UCSP

<i>Resolution enhancement</i>	<i>Time process</i>
Video 5 seconds 30 frames x sec.	157 s.
GPU Acceleration	
Video 8 seconds 29 frames x sec.	43 s.
Video 7 seconds 30 frames x sec.	39 s.

In Table III we have face detection accuracy based on various video features. The measure is based on framerate and the data flow (kbits per second). It is observed that while the data flow is lower, it is more difficult to detect faces. Moreover we saw that the super resolution improve the quality and face detection but when the video have a very poor resolution and data flow the results were bad.

TABLE IV
RESULTS AVERAGE PROCESSING TIME IN SECONDS AND
PERCENTAGE OF SUCCESSES IN THE FACE DETECTION
OPTIMIZATION WITH SUPER-RESOLUTION.

<i>Videos with variation in frames</i>	<i>Average speed(fps)</i>		<i>% of detect faces</i>	<i>Accu racy</i>
	<i>CPU</i>	<i>GPU</i>		
Videos with lower framerate 10 fs/s	170	57	1-	14%
Videos with lower framerate 15 fs/s	295	98	2-	37%
Videos with framerate near to 25 fs/s	320	105	4+-	78%

<i>Videos with variation in frames</i>	<i>Average speed(fps)</i>		<i>% of detect faces</i>	<i>Accu racy</i>
	<i>CPU</i>	<i>GPU</i>		
Videos with framerate equal to 30 fs/s	365	121	5+-	91%
Videos with higher framerate 35 fs/s	390	130	8+	97%

VI. CONCLUSIONS

- It is shown that the implementation of super-resolution improves the quality of videos and also generates as many frames per second as well as higher data flow, as full empty pixels, which contributes to process as many images that together help to get one frame of higher resolution.
- As for videos of very low resolution, the SR does not meet expectations, because having too little of pixels per frame, interpolation is difficult without being able to generate higher resolution video.
- The processing of face detection and tracking of points of interest can be performed faster thanks to GPU processing, verifying the reduction in computational time and cost.

ACKNOWLEDGMENT

- This work, going on to thank God, my parents and my brothers for guiding me and pushing me to go ahead, my advisor and research guide Juan Carlos Gutiérrez Cáceres and CONCYTEC that through CITEC of the National University of Saint Augustine offered me the opportunity to be part of the Master's program and to make possible studies, research and projects of great impact on computer vision and technology.

REFERENCES

- [1] O. Velarde A. y L. Chacón O. P. Rivas P. "Reconocimiento facial en ambientes no cooperativos". 2011.
- [2] Dario Eduardo Villalon de la Vega. "Diseño e implementación de una plataforma de software para reconocimiento facial en video". Master's thesis, Universidad de Chile, Facultad de ciencias físicas y matemáticas. Departamento de ingeniería eléctrica, 2012.
- [3] JD Van Ouwerkerk. "Image super-resolution survey". Image and Vision Computing, 24(10):1039-1052, 2006.
- [4] Indumati Agrawal, Mehul Parikh, PG Student, and GEC Modasa. "Techniques for image super resolution-a survey". International Journal For Technological Research In Engineering, 2: ISSN (Online): 2347 - 4718, 2015.
- [5] Jie Hu, Hailiang Li, and Ying Li."Real time super resolution reconstruction for video stream based on gpu". In Orange Technologies (ICOT), 2014 IEEE International Conference on, pages 9-12. IEEE, 2014.
- [6] Paul Viola and Michael Jones. "Rapid object detection using a boosted cascade of simple features". In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, volume 1, pages I-511. IEEE, 2001.
- [7] David Oro, Carles Fernández, Carlos Segura, Xavier Martorell, and Javier Hernando. "Accelerating boosting-based face detection on gpus. In Parallel Processing (ICPP)", 2012 41st International Conference on, pages 309-318. IEEE, 2012.

- [8] Javier Eslava Ríos. "Reconocimiento facial en tiempo real". Master's thesis, Universidad Autónoma de Madrid, Área de Tratamiento de Voz y Señales. Dpto. de Ingeniería Informática, 2013.
- [9] Paul Viola and Michael J Jones. "Robust real-time face detection. *International journal of computer vision*", 57(2):137-154, 2004.
- [10] M Castrillón, Oscar Déniz, Cayetano Guerra, and Mario Hernández. "Encara2: Real-time detection of multiple faces at different resolutions in video streams". *Journal of Visual Communication and Image Representation*, 18(2):130-140, 2007.
- [11] Marwa Chouchene, Fatma Ezahra Sayadi, Haythem Bahri, Julien Dubois, Johel Miteran, and Mohamed Atri. "Optimized parallel implementation of face detection based on gpu component". *Microprocessors and Microsystems*, 2015.
- [12] Santiago Cortijo Pablo Croveto, Daniel Palomino. "Reconocimiento de patrones faciales en tiempo real mediante transformada de wavelet y computación paralela". Universidad Nacional de ingeniería. Centro de tecnologías de información y comunicaciones, 2010.
- [13] Dario Eduardo Villalon de la Vega. "Diseño e implementación de una plataforma de software para reconocimiento facial en video". Master's thesis, Universidad de Chile, Facultad de ciencias físicas y matemáticas. Departamento de ingeniería eléctrica, 2012.
- [14] Marcelo J Armengot Iborra. "Análisis comparativo de métodos basados en subespacios aplicados al reconocimiento de caras". Universidad de Valencia. 2006.
- [15] Federico Matta and Jean-Luc Dugelay. "Person recognition using facial video information: A state of the art". *Journal of Visual Languages & Computing*, 20(3):180-187, 2009.
- [16] Edy Winarno, Agus Harjoko, Aniat Murni Arymurthy, and Edi Winarko. "Improved real-time face recognition based on three level wavelet decomposition-principal component analysis and mahalnobis distance". *Journal of Computer Science*, 10(5):844-851, 2014.
- [17] Enrique G Ortiz and Brian C Becker. "Face recognition for web-scale datasets. *Computer Vision and Image Understanding*", 118:153-170, 2014.
- [18] Shervin Rahimzadeh Arashloo and Josef Kittler. "Fast pose invariant face recognition using super coupled multiresolution markov random fields on a gpu". *Pattern Recognition Letters*, 48:49-59, 2014.
- [19] Jung, C., Ke, P., Sun, Z., & Gu, A. "A fast deconvolution-based approach for single-image super-resolution with GPU acceleration". *Journal of Real-Time Image Processing*, 1-12. 2015.
- [20] Wang, K., Wang, L., Lu, J., Sun, Y., & Zhao, S. (2012, September). "A GPU-based implementation on super-resolution reconstruction". In *Image Processing (ICIP), 2012 19th IEEE International Conference on* (pp. 849-852). IEEE. 2012.
- [21] Chen, C., Liang, H., Zhao, S., Lyu, Z., & Sarem, M. (2014). "A novel multi-image super-resolution reconstruction method using anisotropic fractional order adaptive norm". *The Visual Computer*, 1-15. 2014.
- [22] José Antonio Huamán Layme. Localización, seguimiento y reconocimiento de rostros empleando métodos estadísticos pca y hmme. Universidad Nacional de Ingeniería. Programa Cybertesis PERÚ, 2007.
- [23] Yousri Ouerhani, Maher Jridi, Ayman Alfalou, and C Brosseau. Optimized pre-processing input plane GPU implementation of an optical face recognition technique using a segmented phase only composite filter. *Optics Communications*, 289:33-44, 2013.