Paper: jc*-**-**-****

Robust Facial Expression Recognition using Near Infrared Cameras

Laszlo A. Jeni[1], Hideki Hashimoto[2] and Takashi Kubota[1]

[1] Dept. of Electrical Engineering, The University of Tokyo
[2] Dept. of Electrical, Electronics and Communication Engineering, Chuo University

[Received 00/00/00; accepted 00/00/00]

Abstract. In human-human communication we use verbal, vocal and non-verbal signals to communicate with others. Facial expressions are a form of non-verbal communication, recognizing them helps to improve the human-machine interaction. This paper proposes a system for pose- and illumination-invariant recognition of facial expressions using near-infrared camera images and precise 3D shape registration. Precise 3D shape information of the human face can be computed by means of constrained local models (CLM), which fits a dense model to an unseen image in an iterative manner. We used a multi-class SVM to classify the acquired 3D shape into different emotion categories. Results surpassed human performance and show pose-invariant performance. Varying lighting conditions can influence the fitting process and reduce the recognition precision. We built a near-infrared and visible light camera array to test the method with different illuminations. Results shows that the near-infrared camera configuration is suitable for robust and reliable facial expression recognition with changing lighting conditions.

Keywords: Emotion Recognition, 3D Face Tracking, Near Infrared Camera, Constrained Local Models

1. Introduction

Nowadays robotic systems are becoming more and more complex and the need for communication between humans and artificial systems is growing. Intelligent systems [18] are able to recognize [20, 26], track [28, 31] and support people [27], however when important information has to be communicated, we feel that personal presence is obligatory, because beyond verbal there are many other communication channels in operation.

Facial expressions are a form of non-verbal communication; it is an outward reflection of a person's emotional condition. Recognizing these expressions helps us to estimate the emotional state of a person.

In the last decade many approaches have been proposed for this problem (see [35] and references therein). Most of the existing methods require frontal faces with minimal head rotations or working with texture information [2, 4, 10, 17, 19, 33]. Recently, high quality facial expression recognition algorithms have been introduced. These algorithms make use of textural information [5, 23]. However, texture information may be prone to changes introduced by the pose, not to mention light conditions.

Markerless three dimensional motion tracking [1, 11, 21] is still difficult as the motion parameters have to be extracted from a 2D image sequence. The rigid (head motion) and non-rigid (expressions) motion of the head are combined in video images, therefore it is necessary to separate these [34].

Optic flow based feature point tracking has been used for object tracking extensively [8, 13, 14]. However, tracking faces while displaying expressions influences the efficiency: during facial expressions structures (and therefore the tracked landmarks) appear/disappear on the face (wrinkles, teeth).

We are experiencing a breakthrough in this field due to the advance of learning algorithms, most notably the advance of Constrained Local Models (CLM) [7, 30].

The main contribution of this paper is a system for pose- and illumination-invariant recognition of facial expressions using near-infrared cameras and 3D shape information only. We built a camera array that can record high quality images in the visible light and near-infrared domains and we compared the perfomance in these two domains. We found a considerable advantage for the near-infrared images both in the head pose estimation task and in the CLM fitting task. The proposed method is pose invariant and works in real-time, making the use of the system for real-life applications suitable.

This paper is organized as follows: in Section 2 we will introduce and talk about the concept of emotion recognition and facial feature extraction. Section 3 describes the experimental setup for the synchronized near-infrared and visible light image capturing. Section 4 introduces the collected dataset and other expert annotated databases used in this research. Section 5 shows experimental results and we conclude in Section 6 with a discussion.

2. Theory

2.1. Facial Expressions

Facial expressions result from one or more motions or positions of the muscles of the face. These movements

convey the emotional state of the individual to observers. Facial expressions are a form of nonverbal communication. They are a primary means of conveying social information among humans, but also occur in other mammals as well as some other animal species. Facial expressions and their significance in the perceiver can, to some extent, vary between cultures.

The focus of our interest is the facial macro-expressions. We display these facial expressions in our daily interactions with other people all the time, when we don't want to conceal our emotions. Usually they last from second to 4 seconds. There are seven universal facial expressions [9], which are present in every culture. These are anger, contempt, disgust, surprise, fear, happiness and sadness.

To describe these emotions Ekman et al. proposed an anatomically oriented coding scheme, the Facial Action Coding System (FACS)[9]. This system is based on the definition of action units (AUs) of a face that cause facial movements. Each action unit may correspond to several muscles that together generate a certain facial action. As some muscles give rise to more than one action unit, correspondence between action units and muscle units is only approximate. 46 AUs were considered responsible for expression control and 12 for gaze direction and orientation.

2.2. Constrained Local Models

Constrained Local Models are generative parametric models for person-independent face alignment. They apply region templates and use fast gradient algorithms in order to optimize them. The shape model of a 3D CLM, for example, is defined by a 3D mesh and in particular the 3D vertex locations of the mesh. Consider shape of a 3D CLM as the coordinates of N 3D vertices that make up the mesh:

$$s = (x_1, y_1, z_1, x_2, y_2, z_2, ..., x_N, y_N, z_N)^T. \quad (1)$$

This model allows linear shape variation: the shape s can be expressed as a base shape s_0 plus a linear combination of m shape vectors $s_i \in \mathbb{R}^{3M}, (i = 1, ..., m)$:

$$s = s_0 + \sum_{i=1}^{m} p_i s_i. \quad (2)$$

where the the coefficients p_i are the shape parameters and we assume that the vectors s_i are orthonormal.

The shape models are normally computed from hand annotated training images. The standard approach is to apply Principal Component Analysis (PCA) to the training meshes [6]. The base shape s_0 is the mean shape and the vectors s_i are the m eigenvectors corresponding to the m largest eigenvalues. Before applying PCA we can normalize the meshes using a Procrustes analysis [16] to remove variation due to a chosen global shape normalising transformation. Therefore the resulting PCA is only concerned with local, non-rigid shape deformation.

In our work, we used the 3D CLM method [30] which fits its model to an unseen image in an iterative manner by generating templates using the current parameter estimates, correlating the templates with the target image to generate response images and optimizing the shape parameters so as to maximize the sum of responses. The interested reader is referred to [7] and [30] for the details of the CLM algorithm. We note that the 3D CLM of [30] is using 6 rigid and 24 non-rigid parameters , where the non-rigid parameters are determined by principal component analysis (PCA) by starting from 66 marker points (also called landmarks), i.e., from 3*66=198 dimensions.

2.3. Multi-class Support Vector Machine for Emotion Classification

Support Vector Machines (SVMs) are very powerful for binary and multi-class classification as well as for regression problems. They are robust against outliers. For two-class separation, SVM estimates the optimal separating hyper-plane between the two classes by maximizing the margin between the hyper-plane and closest points of the classes. The closest points of the classes are called support vectors; they determine the optimal separating hyper-plane, which lies at half distance between them.

We are given sample and label pairs $(x^{(k)}, y^{(k)})$ with $x^{(k)} \in \mathbb{R}^m$, $y^{(k)} \in \{-1, 1\}$, and $k = 1, ..., K$. Here, for class 1 (class 2) $y^{(k)} = 1$ ($y^{(k)} = -1$). Assume further that we have a set of feature vectors $\boldsymbol{\phi}(= [\phi_1; \ldots; \phi_M]) : \mathbb{R}^m \rightarrow \mathbb{R}^M$, where M might be infinite. The support vector classification seeks to minimize the cost function

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{K} \xi_i \quad (3)$$

$$y^{(k)}(w^T \phi(x^{(k)}) + b) \geq 1 - \xi_i, \xi_i \geq 0. \quad (4)$$

In this study we used the LIBSVM software [3] We used multi-class classification, where decision surfaces are computed for all class pairs, i.e., for k classes one has $k(k-1)/2$ decision surfaces and then applies a voting strategy for decisions. Multi-class SVM is considered competitive to other SVM methods [3], but in this case we found very little differences if any when compared it with one-against all procedures. In all cases, we used only linear classifiers.

3. Experimental Setup

First, we provide details about our Near-Infrared Imaging Hardware (Section 3.1). It is followed by the details of the Infrared-Visible Light Camera Array we used for dataset building.

3.1. Near-Infrared Imaging Hardware

One of the main issues with CLMs is that they are sensitive to appearance changes. Illumination conditions and skin color differences from the ones found in the training dataset can influence the performance of the model. To overcome this limitation the proposed system works in

near infrared domain, where we have greater control over these artifacts.

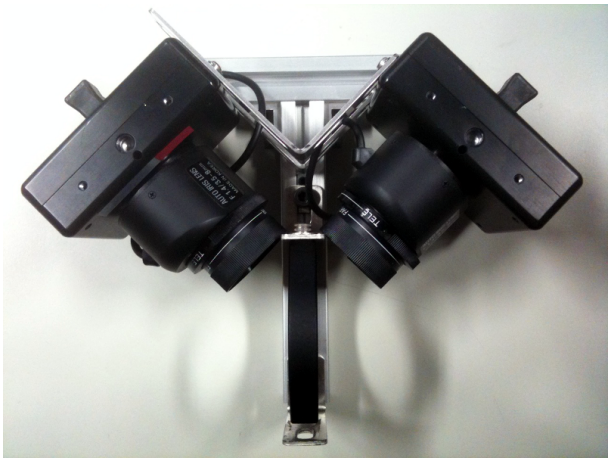


Fig. 1. Infrared-Visible Light Camera Array.

We used DragonFly2 cameras from Point Grey Research [29]. This is a high-speed IEEE-1394a (FireWire) digital camera with on-board color processing and auto white balance functions. The imaging sensor is a Sony 1/3" progressive scan CCD with a good quantum efficiency in the near infrared domain.

The color version comes with an infrared cut-off filter. In the newer models this filter is held in place by a metal plate that is screwed to the lens holder, therefore the removal is quite simple. We replaced this filter with a Fuji IR-76 IR-pass gelatin filter, which cuts off the entire visible light domain under 760nm.

In addition, the DragonFly2 has a DC auto-iris output and can be used with CCTV auto-iris lenses to control the amount of light that falls onto the CCD. Therefore with active near infrared illumination we can use this modified camera for near infrared imaging.

3.2. Infrared-Visible Light Camera Array for Dataset Building

The main goal of this study is to compare the performance of CLMs for facial expression tracking in visible light and infrared domain. For collecting synchronized visible light and infrared video sequences we built a dual camera array, which can record both domains at the same time.

The camera array contains a near infrared camera (mentioned in the previous section) and a normal visible light DragonFly2 camera (see Figure 1) mounted 90° angle. The input beam is splitted by a cold mirror from ThorLabs, which transmits 85% of the infrared light (from 750nm to 1200nm) and reflects 90% of the visible light (from 420nm to 630nm).

The cameras are connected to the same IEEE-1394 bus and automatically synchronized to each other. The maximum deviation between them is 125 μs. A sample photo from the two cameras can be seen on Figure 2.

Fig. 2. A sample overlapped photo from the camera array. Left side is from the IR, right side is from the VL photo.

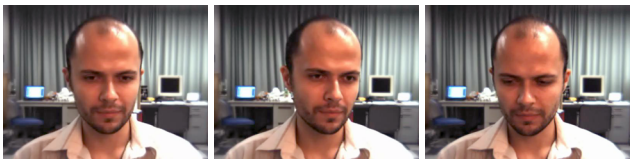
4. Datasets

4.1. IR-VL Face Tracking Dataset

We used the IR-VL camera array (described in the previous Section) to build a near infrared - visible light database for face tracking.

We recorded short video sequences (300 frames each) of five subjects at the resolution of 640×480 pixel and 30 fps. The subjects were asked to perform various head movements, including translation and rotation, without distinctive facial expressions. Ten video sequences were collected from each subject, the dataset contains 50-50 synchronized visible light and near-infrared sequences overall.

A sample IR-VL sequence from the dataset can be seen in Figure 3



(a)

(b)

Fig. 3. An example sequence recorded by the camera array. (a) Visible Light domain (b) Near Infrared domain.

4.2. The Karolinska Directed Emotional Faces

During the simulation we used the The Karolinska Directed Emotional Faces (KDEF) [24]. This database was created in the Karolinska Institute in Sweden to be used for psychological and medical research purposes, particularly suitable for perception, attention, emotion, memory and backward masking experiments.

The set contains 70 individuals, each displaying 7 different emotional expressions (neutral, afraid, angry, disgusted, happy, sad and surprised). Each expression being photographed (twice) from 5 different angles (4900 pictures in total).

The dataset is not annotated with facial landmarks. In the experiments, the landmarks were provided by the CLM-tracker itself.

4.3. The Oulu-Casia Near-Infrared Dataset

To compare the emotion classification in visible light and near-infrared domain we used the Oulu-Casia Near-Infrared Visible Light Dataset [32]. The dataset consists of six facial expressions from 80 people, most of them are either Finnish or Chinese.

The subjects were asked to sit in front of a camera that captures near-infrared and visible light images at the same time. They were asked to perform different facial expressions according to a given example during the recording sessions. The imaging hardware captured images at 320×240 pixels at 25 fps.

All of the expressions are recorded in three different lighting conditions: strong, weak and dark illumintaions. The dataset contains overall 2880 video sequences.

5. Experimental Results

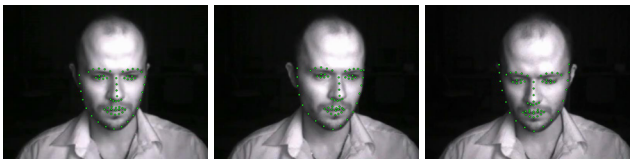
In this section we provide details about out experiments in near infrared and visible light domains. Experiments on comparing the CLM based pose esimation and CLM fitting are described in Section 5.1 and 5.2, respectively. Emotion classification experiments on the Karolinska Dataset are detailed in Section 5.3 and on the Oulu-Casia Dataset in Section 5.4.

5.1. Head Pose Estimation in IR - VL

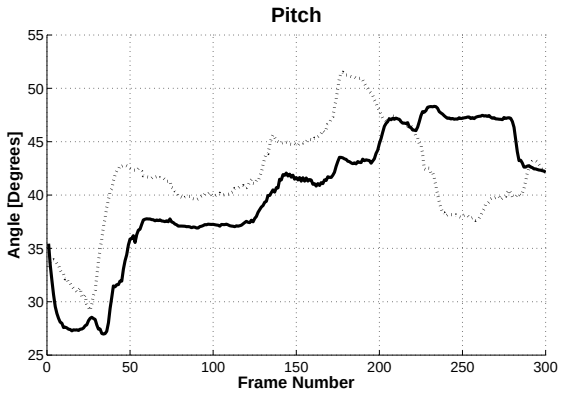
In this experiment we compared the CLM technique performance in visible light and infrared domains as a function of pose estimation.

We used our IR-VL Face Tracking Dataset, calculated the facial landmarks by the CLM method and estimated the head pose from the 3D mesh. Figure 4 shows a considerable difference between the pose estimation in the infrared and visible light domain.

Results show that the tracker performs better in the near-infrared domain, produces less spikes in the tracking.

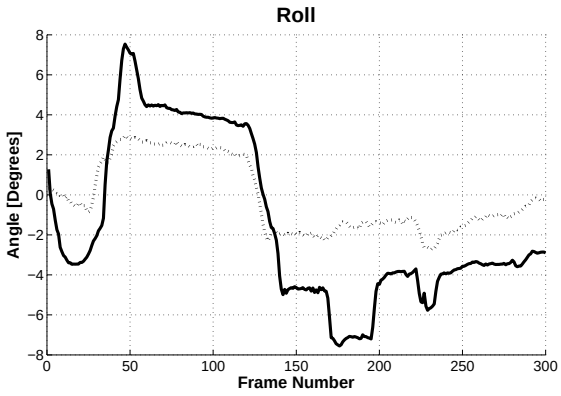


(a)



(b)

(c)



(d)

Fig. 4. An example tracking sequence from the IR-VL database. (a): Annotated frames from a sequence. (b-e): Pitch/Yaw/Roll estimation. In each graph, the solid curve depicts the visible light tracker and the dotted curve depicts the infrared tracker.

5.2. CLM Fitting Comparsion in IR - VL

In this experiment we compared the CLM's performance as a function of landmark position estimation in visible and infrared domains. We calculated the RMSE of the estimated landmark positions of the frontal face $x^{(f)}$ and the tracked one $x^{(t)}$.

$$RMSE(x^{(f)}, x^{(t)}) = \sqrt{\frac{\sum_{i=1}^{M}(x_i^{(f)} - x_i^{(t)})^2 + (y_i^{(f)} - y_i^{(t)}))^2}{2M}} \quad (5)$$

As illustrated in the upper subfigure of Figure 5 the CLM tracker accumulates considerably more error in the visible light domain than in the infrared domain.

We divided the landmarks into five groups based on facial features (profile, eyebrow, eyes, mouth and nose region) and calculated the average RMS error over the whole dataset.

The bottom subfigure of Figure 5 shows the comparison of the RMSE in the infrared and visible light domain by groups.

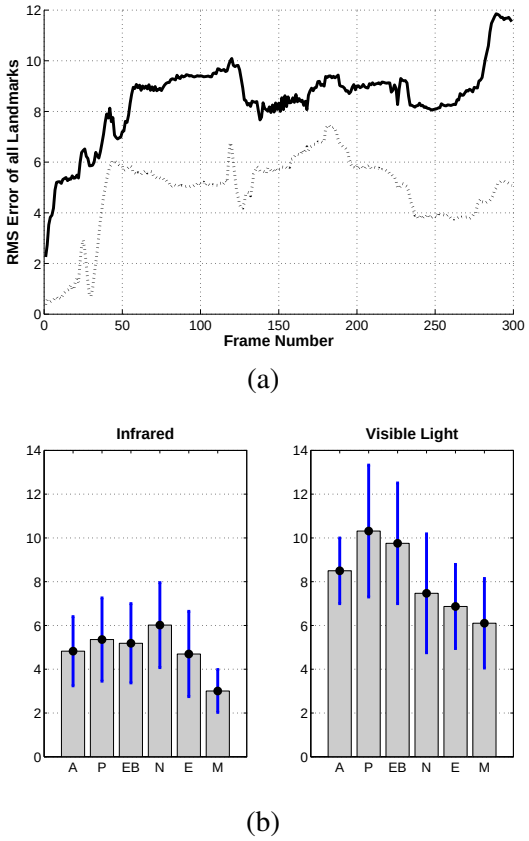


Fig. 5. RMSE of the reconstructed landmark positions in pixels. (a) RMSE of all landmarks from an example tracking sequence. The solid curve depicts the visible light tracker and the dotted curve depicts the infrared tracker. (b) RMSE of different landmark groups: A - All landmarks / P - Profile / EB - Eyebrow / N - Nose / E - Eyes / M - Mouth. The distortion was compared to a frontal face. 1 pixel error for all landmarks corresponds to 1 RMSE unit.

We can see that certain groups (for example, eyes and mouth region) are more stable. These landmarks have better error surface than others.

5.3. Emotion Classification on KDEF Dataset

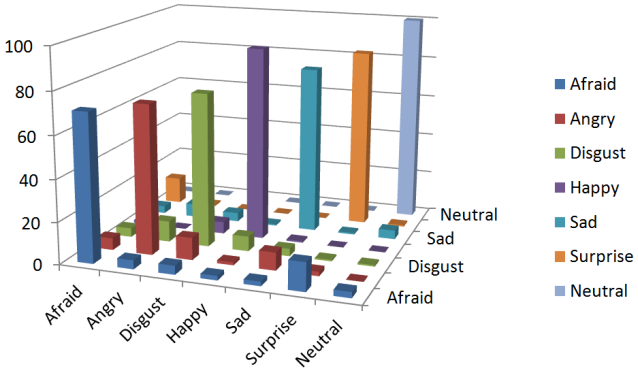
In this experiment we studied the performance of the multi-class SVM using CLM method on the Karolinska database [24] (details of the dataset are provided in Section 4.2).

First we tracked the facial expressions with CLM tracker and extracted the 3D positions of the facial landmarks.

In the next step, we performed a personal mean shape normalization [15]: we calculated an average shape for each subject (the so called personal mean shape) and then we computed the differences between the features of the emotional shape and the features of this personal mean shape. This step is important, because it removes the personal variation of the shape.

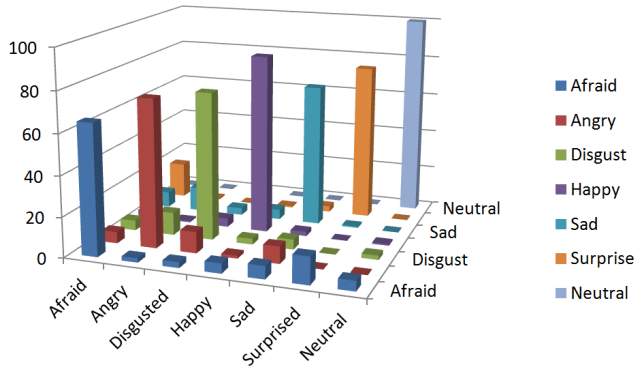
In the last step we trained a multi-class SVM using the leave-one-subject-out cross validation method on the normalized shapes. The result of the classification is shown in Figure 6: emotions with large distortions can still be recognized in about 85-90% of the cases, whereas more subtle emotions are sometimes confused with others.

These recognitions rate higher than a human observer [24] and slightly lower than the method proposed by Lucey et al. in [22], which utilizes both shape and texture information. We note that the method in [22] works just on frontal faces, while our shape based method works on rotated faces also (see Figure 7). A comparison of the different results can be seen in Figure 8.



	Afraid	Angry	Disgusted	Happy	Sad	Surprised	Neutral
Afraid	**70.71**	4.29	4.29	2.14	2.14	13.57	2.86
Angry	5.71	**71.43**	10.71	1.43	8.57	2.14	0
Disgusted	4.29	10	**73.57**	7.14	3.57	0.71	0.71
Happy	1.43	0	5.71	**92.14**	0.71	0	0
Sad	3.57	6.43	4.29	0.71	**80**	0.71	4.29
Surprised	12.86	0.71	0	0	0	**85.71**	0.71
Neutral	0	0	0	0	0	0	**100**

Fig. 6. Confusion matrix for the proposed method on the frontal faces from the Karolinska dataset.



	Afraid	Angry	Disgusted	Happy	Sad	Surprised	Neutral
Afraid	**65**	2.14	2.86	5	6.43	13.57	5
Angry	5.71	**73.57**	10.71	1.43	8.57	0	0
Disgusted	5	11.43	**73.57**	2.86	5	0	2.14
Happy	3.57	0.71	4.29	**88.57**	2.14	0	0.71
Sad	7.86	12.14	3.57	5	**70.71**	0.71	0
Surprised	17.86	0.71	0	0.71	2.86	**77.86**	0
Neutral	0	0	0	0	0	0	**100**

Fig. 7. Confusion matrix for the proposed method on the half-profile faces from the Karolinska dataset.

	Feature	Afraid	Angry	Disgusted	Happy	Sad	Surprised	Neutral	Average
Human observer	–	43.03	78.81	72.17	92.65	76.7	96	62.64	74.57
Lucey et al. '10	Shape + Texture	65.22	75	94.74	100	68	96	100	85.57
This work (frontal faces)	Shape only	70.71	71.43	73.57	92.14	80	85.71	100	81.94
This work (half profile)	Shape only	65	73.57	73.57	88.57	70.71	77.86	100	78.45

Fig. 8. Comparison of the different result on the Karolinska dataset. The human observer values are from [12].

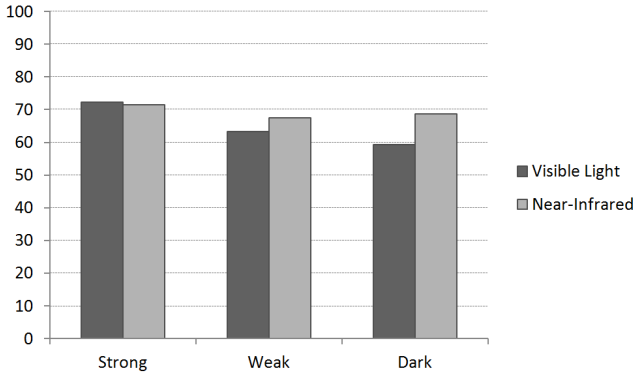
5.4. Emotion Recognition in Near-Infrared Domain

As we saw in Section 5.1 and 5.2, the CLM based head pose estimation and facial landmark registration is more precise in the near infrared domain. Also, in the previous section we showed on the Karolinska dataset that shape based emotion classification works well in the visible light domain. To investigate further the difference between the visible light and near infrared domain, we used the Oulu-Casia dataset for emotion recognition [32].

We characterized the dataset using the CLM tracker and extracted the 3D landmark positions. We performed a personal mean-shape normalization [15] on the extracted faces and trained a multi-class SVM using the leave-one-subject-out cross validation method on the normalized shapes.

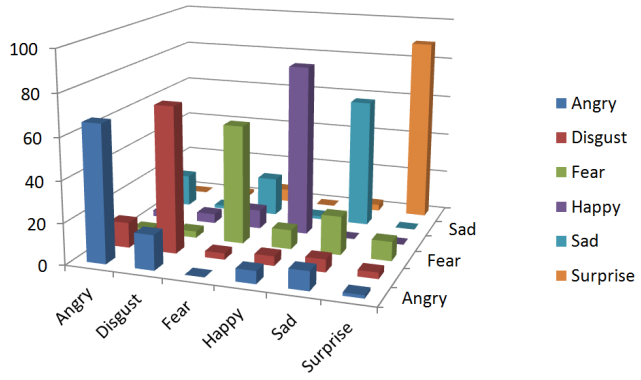
We performed this process on the visible light and near-infrared images using the three illumination conditions available in the dataset (strong, weak and dark illumination). The classification results are shown in Figure 9. The classification rate in the near-infrared domain is more consistent across the different illuminations than the visible light domain results.

The complete confusion matrices for the near-infrared and visible light images from the strong illumination subset are shown in Figure 10 and 11 respectively. Emotions with large distortions can be recognized in about 70-90% of the cases in both domains.



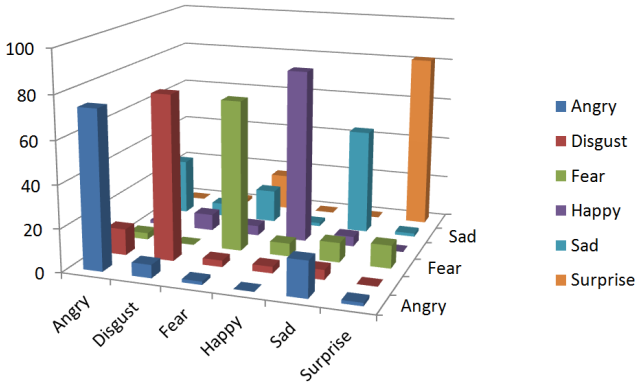
	Angry	Disgust	Fear	Happy	Sad	Surprise	Average
NI Strong	66.15	70.77	57.81	83.08	62.5	89.23	71.59
VL Strong	73.85	76.92	70.77	81.54	49.23	81.54	72.31
NI Weak	55.38	53.85	64.62	78.46	70.77	81.54	67.44
VL Weak	47.69	61.54	46.15	73.85	67.69	83.08	63.33
NI Dark	50.77	67.69	55.38	81.54	73.85	83.08	68.72
VL Dark	43.08	53.85	56.92	67.69	52.31	81.54	59.23

Fig. 9. Recognition rates of the visible light and near infrared images from the Oulu-Casia dataset.



	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	**66.15**	16.92	0	6.15	9.23	1.54
Disgust	12.31	**70.77**	3.08	4.62	6.15	3.08
Fear	1.56	3.13	**57.81**	9.38	18.75	9.38
Happy	3.08	4.62	9.23	**83.08**	0	0
Sad	15.63	1.56	18.75	1.56	**62.5**	0
Surprise	0	1.54	6.15	0	3.08	**89.23**

Fig. 10. Confusion matrix for the near-infrared images from the strong illumination subset of the Oulu-Casia dataset.



	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	**73.85**	6.15	1.54	0	16.92	1.54
Disgust	12.31	**76.92**	3.08	3.08	4.62	0
Fear	3.08	0	**70.77**	6.15	9.23	10.77
Happy	1.54	7.69	4.62	**81.54**	4.62	0
Sad	26.15	6.15	15.38	1.54	**49.23**	1.54
Surprise	0	1.54	16.92	0	0	**81.54**

Fig. 11. Confusion matrix for the visible light images from the strong illumination subset of the Oulu-Casia dataset.

6. Conclusion

In this paper we used a number of methods to study the performance of shape based facial expression recognition in near-infrared and visible light domains. In the emotion recognition studies, we applied multi-class SVM classification [3]. We used expert annotated face databases [24, 32] as well as video sequences recorded by an IR-VL camera array (Section 4.1). We used CLM method to extract shape data in near-infrared and visible light domains, since it is more precise and may preserve more information than Active Appearance models [25]. We received high recognition rate on the Karolinska dataset [24] using only shape information. This can be of great importance, since shape information is robust against different head rotations. We demonstrated this on the half-profile faces in the Karolinska dataset.

To study the behaviour of the proposed method, we built a camera array that can record high quality images in the visible light and near-infrared domains. We recorded a synchronized head tracking dataset and compared the performance on the visible light and near-infrared sequences. We found a considerable advantage for the near-infrared images both in the head pose estimation task and in the CLM fitting task.

To investigate the performance in different illumination conditions, we used the Oulu-Casia dataset [32] which consists of near-infrared and visible light image sequences recorded in various lighting conditions. We found that the classification rate in the near-infrared domain is more consistent across the different illuminations and outperforms the visible light domain results.

From the point of human-robot interaction, angle and illumination dependence of facial expression recognition is of great importance. Our experiments show that the proposed shape based recognition technique and the used near-infrared camera configuration is suitable for robust and reliable facial expression recognition. Our proposed method works in real-time, making the use of the system for real-life applications available.

Concerning future work, further improvements can be expected by including texture [22] and temporal information [32] in the recognition process, however pose invariant texture extraction with minimal distortions is difficult.

Acknowledgements

We are grateful to Jason Saragih for providing his CLM code for our work.

References:

[1] P. Azad, T. Asfour, R. Dillmann, "Robust real-time stereo-based markerless human motion capture," Humanoid Robots, 2008. Humanoids 2008. 8th IEEE-RAS International Conference on , vol., no., pp.700-707, 1-3 Dec. 2008

[2] M. S. Bartlett, G. Littlewort, M. G. Frank, C. Lainscsek, I. R. Fasel, and J. R. Movellan. Automatic recognition of facial actions in spontaneous expressions. Journal of Multimedia, 1(6), 22–35. (2006)

[3] C.-C. Chang and C.-J. Lin, LIBSVM: a library for support vector machines, http://www.csie.ntu.edu.tw/ cjlin/libsvm, 2001.

[4] Y. Chang, C. Hu, R. Feris, and M. Turk. Manifold based analysis of facial expression. Image and Vision Computing, 24(6), 605–614. (2006)

[5] S. W. Chew, P. J. Lucey, S. Lucey, J. Saragih, J. F. Cohn and S. Sridharan, Person-independent facial expression detection using constrained local models. In Proceedings of FG 2011 Facial Expression Recognition and Analysis Challenge, Santa Barbara, CA, 2011.

[6] T. Cootes, C. Taylor, (1992). Active shape modelssmart snakes. In British machine vision conference (BMVC92) (pp. 266275).

[7] D. Cristinacce and T. Cootes, Automatic feature localisation with constrained local models, Pattern Recognition, 41: 3054-3067, 2008.

[8] D. DeCarlo and D. Metaxas, The integration of optical flow and deformable models with applications to human face shape and motion estimation, in Proc. IEEE Conf. CVPR, San Francisco, CA, 1996, pp. 231238.

[9] Ekman, P., Rosenberg, E. L. (Eds.). What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (2nd ed.). New York: Oxford University Press. (2005)

[10] B. Fasel, F. Monay, and D. Gatica-Perez. Latent semantic analysis of facial action codes for automatic facial expression recognition. In Proceedings of the ACM SIGMM international workshop on multimedia information retrieval (pp. 181–188). (2004)

[11] Y. Furukawa, J. Ponce, "Dense 3D motion capture from synchronized video streams," Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on , vol., no., pp.1-8, 23-28 June 2008

[12] E. Goeleven, R. D. Raedt, L. Leyman, and B. Verschuere. The Karolinska Directed Emotional Faces: A validation study. Cognition and Emotion, 22(6):10941118. 7

[13] J. Hoey, J. J. Little, "Bayesian clustering of optical flow fields," Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on , vol., no., pp.1086-1093 vol.2, 13-16 Oct. 2003

[14] Yun-Shu Hou; Yan-Ning Zhang; Rong-Chun Zhao; , "Robust tracking of nonrigid objects using techniques of inverse component uncertainty factorization subspace constraints optical flow," Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on , vol.9, no., pp.5458-5466 Vol. 9, 18-21 Aug. 2005

[15] L. Jeni, D. Takcs and A. Lorincz. High Quality Facial Expression Recognition in Video Streams using Shape Related Information only. In.: Benchmarking Facial Image Analysis Technologies at ICCV 2011 (accepted).

[16] M. Kamandar, S. A. Seyedin, "Procrustes based shape prior for parametric active contours," Machine Vision, 2007. ICMV 2007. International Conference on , vol., no., pp.135-140, 28-29 Dec. 2007.

[17] S. Koelstra, and M. Pantic. Non-rigid registration using freeform deformations for recognition of facial actions and their temporal dynamics. In Proceedings of the IEEE international conference on automatic face and gesture recognition. (2008)

[18] P. Korondi, H. Hashimoto, Intelligent Space, as an Integrated Intelligent System. Keynote paper of International Conference on Electrical Drives and Power Electronics, Proceedings, 2003, pp. 24-31.

[19] I. Kotsia, and I. Pitas. Facial expression recognition in image sequences using geometric deformation features and support vector machines. IEEE Transactions on Image Processing, 16(1), 172-187. (2007)

[20] J. Lee, K. Morioka, N. Ando, H. Hashimoto, Cooperation of Distributed Intelligent Sensors in Intelligent Environment. IEEE/ASME Transactions on Mechatronics, Vol.9, No.3, 2004, pp.535-543, ISSN 1083-4435

[21] Yifan Lu; , "Markerless human motion capture: An application of simulated annealing and Fast Marching Method," Pattern Recognition, 2008. ICPR 2008. 19th International Conference on , vol., no., pp.1-4, 8-11 Dec. 2008

[22] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2010.

[23] P. Lucey, J. F. Cohn, K. M. Prkachin, P. Solomon, and I. Matthews, Painful data: The UNBC-McMaster Shoulder Pain Expression Archive Database. 9th IEEE Int. Conf. on Automatic Face and Gesture Recogn. (FG2011).

[24] D. Lundqvist, A. Flykt and A. hman. The Karolinska Directed Emotional Faces - KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9. (1998)

[25] I. Matthews, S. Baker, (2004). Active appearance models revisited. International Journal of Computer Vision, 60, 135164.

[26] K. Morioka, H. Hashimoto, Color Appearance Based Object Identification in Intelligent Space. In: Proc. of the 8th IEEE International Workshop on Advanced Motion Control, Kawasaki, Japan, 2004, pp. 505-510

[27] M. Niitsuma, H. Hashimoto, Spatial Memory as an Aid System for Human Activity in the Intelligent Space. IEEE Transactions on Industrial Electronics, Vol. 54, Issue 2, 2007, pp. 1122-1131, ISSN: 0278-0046.

[28] Z. Petres, P. Baranyi, P. Korondi, H. Hashimoto, Trajectory Tracking by TP Model Transformation: Case Study of a Benchmark Problem, IEEE Trans. on Industrial Electronics, vol. 54, no. 3, pp. 1654-1663, June 2007

[29] Point Grey Research, Inc., http://www.ptgrey.com

[30] J. M. Saragih, S. Lucey, and J. F. Cohn, Deformable model fitting by regularized landmark mean-shift. Int. J. Comp. Vision, 91(2): 200-215, 2011.

[31] T. Sasaki, D. Brscic, H. Hashimoto, Human Observation Based Extraction of Path Patterns for Mobile Robot Navigation. IEEE Transactions on Industrial Electronics, Vol. 56, (accepted for publication July 28, 2009)

[32] M. Taini, G. Zhao, S. Z. Li, M. Pietikinen. Facial Expression Recognition from Near-Infrared Video Sequences. In.: Proc. 19th Conference on Pattern Recognition (ICPR 2008), Tampa, FL.

[33] Y. L. Tian, T. Kanade, and J. F. Cohn. Recognizing action units for facial expression analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(2), 97–115. (2001)

[34] P. Vadakkepat, P. Lim, L.C. De Silva, Liu Jing, Li Li Ling, "Multimodal Approach to Human-Face Detection and Tracking," IEEE Trans. on Industrial Electronics, vol. 55, no. 3, pp. 1385-1393, March 2008.

[35] Z. Zeng, M. Pantic, G. Roisman, and T. Huang. A Survey of Affect Recognition Methods: Audio, Visual and Spontaneous Expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(1):3958, 2009.